

The Canadian Institute for  
Advanced Research  
L'Institut canadien de  
recherches avancées

# PROCEEDINGS

**11th Honda Foundation  
Discoveries Symposium**

***Prosperity, Health  
and Well-Being***

October 16 - 18, 1993  
Toronto, Canada



The Honda Foundation

# PROCEEDINGS

**11th Honda Foundation  
Discoveries Symposium**

***Prosperity, Health  
and Well-Being***

October 16 - 18, 1993  
Toronto, Canada



The Honda Foundation

The proceedings of a Symposium "Prosperity, Health and Well-Being" held in Toronto, Canada on October 16-18, 1993, sponsored by the Honda Foundation of Japan in Collaboration with The Canadian Institute for Advanced Research with support from external Affairs and International Trade Canada.

### **Organizing Committee**

#### **Co-Chairs:**

*Toshiyuki Furukawa*, Osaka National Hospital, Japan

*J. Fraser Mustard*, Canadian Institute for Advanced Research, Canada

#### **Members:**

*Kazuhiko Atsumi*, Science Council of Japan

*Robert Evans*, University of British Columbia, Canada

*John Frank*, University of Toronto, Canada

*Toshio Sasaki*, Honda Foundation, Japan

*Kei Takeuchi*, University of Tokyo, Japan

### **Scientific Committee**

#### **Chair:**

*Robert Evans*, University of British Columbia, Canada

#### **Members:**

*John Frank*, University of Toronto, Canada

*Clyde Hertzman*, University of British Columbia, Canada

*Michael Marmot*, University of London, England

*Kanehisa Morimoto*, Osaka University, Japan

*Marc Renaud*, University of Montreal, Canada

*Greg Stoddart*, McMaster University, Canada

Published by The Canadian Institute for Advanced Research, Toronto.

ISBN 1-895872-30-8

## **TABLE OF CONTENTS**

### **Page**

- 1      **Preface – Professor Robert G. Evans (Director, Population Health Program,  
The Canadian Institute for Advanced Research)**
- 3      **Honda Conference Participants – Background Information**

### **Session One - Variations in Health Status**

- 14      **Toshiyuki Furukawa**  
          **“A Transactional Comparison of Average Life Expectancy”**
- 35      **Michael G. Marmot**  
          **“Social Differentials in Health Within and Between Populations”**
- Discussant**
- 50      **George A. Kaplan**
- 55      **John Frank**

### **Session Two - Social Environment and Health**

- 58      **S. Leonard Syme**  
          **“The Social Environment and Health”**
- 65      **Clyde Hertzman**  
          **“The Lifelong Impact of Childhood Experiences:  
A Population Health Perspective”**
- Discussant**
- 76      **Ellen Corin**
- 78      **Akira Okamoto**
- 80      **Chris Power**

### **Session Three - Biological Pathways that Contribute to Variations in Health Status**

- 82      **Max Cynader**  
          **“Biological Pathways that Contribute to Variations in Health Status:  
the Role of Early Experience”**
- 97      **Kanehisa Morimoto**  
          **“Chromosome Alterations as Indices of Lifestyle”**
- 105     **Stephen J. Suomi**  
          **“Social and Biological Pathways that Contribute to Variations in Health Status:  
Evidence from Primate Studies”**
- Discussant**
- 113     **Dan Offord**

### **Session Four - Health Care as a Determinant of Health**

- 115     **John W. Frank (Chair)**  
          **Introductory Comments**
- 118     **Johan P. Mackenbach**  
          **“The Contribution of Medical Care to Mortality Decline: McKeown Revisited”**
- 132     **Noralou Roos**  
          **“Assessing the Relationship between Health Care Use and the Health of a  
Population: Seeking Levers for Policy Makers”**
- Discussant**
- 160     **Patricia Baird**

**Session Five - Health Information : A System of Health Statistics  
Based on the Determinants of Health**

- 162      **Michael Wolfson**  
          “Toward a System of Health Statistics”

**Session Six - Economic Growth, Prosperity, and Health and Well-Being**

- 173      **Nathan Rosenberg**  
          “How the Developed Countries Became Rich”
- 183      **J. Fraser Mustard, John W. Frank**  
          “Historical Perspective on How Prosperity and the Social and Physical  
          Environments Have Affected Health and Well-Being”
- Discussant
- 193      **Richard G. Wilkinson**

**Session Seven - Economic Growth in Today's Global Economy - Implications**

- 208      **Paul M. Romer**  
          “Beyond Classical and Keynesian Macroeconomic Policy”
- 223      **Kei Takeuchi**  
          “Social Stability and Economic Growth”
- Discussant
- 231      **Peter Nicholson**  
235      **Ken Delaney**

**Session Eight - Health Care and the Economy**

- 238      **Naoki Ikegami**  
          “Efficiency and Effectiveness in Health Care”
- 248      **Gregory Stoddart, John Lavis**  
          “Can we have too much Health Care?”
- 262      **Robert G. Evans**  
          “Health Care As a Threat to Health: Defence, Opulence, and the Social  
          Environment”
- Discussant
- 272      **Jonathan Lomas**  
274      **André-Pierre Contandriopoulos**  
278      **John Creighton Campbell**

**Session Nine - Priority-Setting for Investments in Health and Well-Being**

- 280      **Shuzo Nishimura**  
          “Reconsideration on the Concept of Investment and Consumption in Health and  
          Well-Being”
- Discussant
- 286      **John Evans**  
288      **Kazuhiko Atsumi**

# The Eleventh Honda Foundation Discoveries Symposium

---

*Prosperity, Health and Well-Being*

October 16-18, 1993

Toronto, Ontario

Canada

**H**ealth and wealth go together. On average, people in wealthy countries live longer and healthier lives than people in poorer countries. And within countries, wealthy people live longer and are healthier than poorer people. But the connection is not simple; the effects of poverty and material deprivation, for example, do not adequately account for this correlation. We do know, however, that the health and well-being of a country's population are closely linked to the quality of the social, as well as the physical, environments in which they live and work.

The profound economic changes now taking place threaten those environments, and place at risk the health of the whole population, not just the poorest. Understanding the relationships among economic growth, prosperity, and the health of populations has become a critical issue. A more comprehensive view of the factors determining health raises the possibility that devoting excessive spending on health care, by diverting resources from economic growth, may threaten both the prosperity and the health of a population. But only the possibility.

In fact (and surprisingly) the sources of our present prosperity and the factors underlying economic growth have, like the determinants of population health, been poorly understood. The participants at the eleventh Honda Discovery Symposium discuss these subjects, linking together emerging new interpretations of the determinants of both economic growth and health, and of the biological pathways through which the social environment influences individual health. Given this new understanding, what should be our priorities for investments in health?

The sponsor of this symposium, the Honda Foundation of Japan, was established in 1977 by the late Mr. Soichiro Honda to promote understanding of the role of technological innovation in improving the life of populations throughout the world. The Honda Corporation, now one of the world's leading automobile manufacturers, was founded by Mr. Honda after the Second World War. He believed that the wealth from his successful enterprise should contribute to a better understanding of how to build and maintain sustainable, prosperous societies. In addition to sponsoring the Discoveries Symposium, the Honda Foundation also sponsors other meetings and awards the annual Honda prize for distinguished achievement in the field of "eco-technology" (technology that assists people in maintaining harmony with the overall environment surrounding human activities rather than simply improved efficiency in a narrow sense, and greater profit).

Previous symposia have been held in Japan, Italy, France, Sweden, the United States, England, Australia, Belgium, Austria, and West Germany and covered topics such as *Communication in Human Activity*, *The Social Impact of Advanced Technology*, *The Social and Cultural Challenge of Modern Thought*, *The Use of Resources and Technology in the Interest of Mankind*, *Law of Nature and Human Conduct*, and *Basic Life Sciences and Human Society*.

The host for this year's **Honda Discoveries Symposium** is the Canadian Institute for Advanced Research. Founded in 1982, the Institute is a unique institution whose research programs have attracted international attention. It is an "institute without walls". Its mission is to mobilize and focus the knowledge resources and brainpower of a thinly-populated country, linking Canadian researchers, along with key people in other countries, into networks of highly talented individuals working in intellectually demanding fields. This strategy allows the Institute's programs to overcome institutional and regional barriers to the study of complex problems.

The ten programs developed by the Canadian Institute for Advanced Research under the guidance of the Institute's Research Council, a peer group of international experts, embrace three categories:

**Science with major technological potential**

- Artificial Intelligence and Robotics
- Soft Surfaces and Interfaces
- Superconductivity
- Evolutionary Biology

**Science relevant to the physical and biological origins and preservation of our planet:**

- Cosmology (The Origins of the Universe)
- Earth System Evolution (The interaction of the geosphere and Biosphere throughout the history of our planet)
- Evolutionary Biology

**Individual and Societal Health and Well-Being**

- Population Health (The Determinants of Health)
- Economic Growth and Policy (The Determinants of Economic Growth)
- Human Development
- Law and the Determinants of Social Order

(The Program Evolutionary Biology is cited twice in the above categories)

The programs in this section, dealing with the determinants of individual and societal health and well-being, have contributed to the concepts and understanding that will be under discussion at the **Honda Foundation Discoveries Symposium**.

The Canadian Institute for Advanced Research welcomes this opportunity to participate with the Honda Foundation in its efforts to create a better world. We were honoured by the presence at this year's Discoveries Symposium of Mrs. Sachi Honda – like her late husband, a most remarkable person.

**NOTE:** Due to circumstances beyond the control of the symposium sponsors, a small number of presentations by discussants were not available at the time of publication of these proceedings.

# Honda Conference Participants

## Background Information

---

### **Atsumi, Kazuhiko**

Graduating from the University of Tokyo School of Medicine, Dr. Kazuhiko Atsumi has exceeded the boundaries of his career to become President of Nippon Engineering College and Member of the Science Council of Japan. Dr. Atsumi has pursued research in the areas of artificial organs, biomaterial, thoracic surgery, application of computers, medical informatics, medical thermography, medical lasers, and biomagnetism. Evidenced by his research interests he occupies several other leading positions as President of the Japan Society for Artificial Heart and Assisted Circulation, President of the Japan Society of Thermography, President of the Japan Association for Medical Informatics, Vice President of the International Society for Artificial Organs, President of the International College of Thermology.

### **Baird, Patricia**

Dr. Baird is a human geneticist who received her medical degree from McGill University, and served as Head of Medical Genetics at the University of British Columbia for over a decade. Dr. Baird is an Associate of the CIAR Program in Population Health. She is or has been a member of numerous national committees, including the National Advisory Board on Science and Technology, chaired by the Prime Minister; the Medical Research Council of Canada; and the Science Council of Canada Genetic Predisposition Study. She was Co-Chair of the National Forum of Science & Technology Councils in 1991, and is Chairing the federal Royal Commission on New Reproductive Technologies. Among her most current projects, Dr. Baird has focused on the possible relationship of particular paternal occupations to birth defects and on the influence of socioeconomic status on birth defects. She has contributed to knowledge about the incidence and clinical course of Downs Syndrome and neural tube defects for example. Her work has also focused on bioethical questions including the issues of allocation of resources and delivery of health services.

### **Bienenstock, John**

Dr. Bienenstock, physician and scientist, is currently Dean and Vice-President of Health Sciences at McMaster University. After receiving a degree in internal medicine from King's College London, he worked as a Fellow in the Arthritis Unit of the Massachusetts General Hospital and Harvard University. He then went to the State University of New York, Buffalo to work on mucosal immunology. He is author and co-author of more than 300 scientific publications and has published several books. His research focus is in mucosal immunology and allergy and more recently in psychoneuroimmunology.

### **Campbell, John**

Dr. Campbell received the Certificate of the East Asian Institute and his Ph.D. in political science from Columbia University. He is the Director of the East Asia Business Program and Co-director of the Japan Technology Management Program at the University of Michigan. The latter seeks to learn and convey lessons from Japan about management of technology through research, continuing education, and support for students in engineering, science, and

management. His specialization is decision making and public policy. He was awarded the Masayoshi Ohira Memorial Prize in 1993 for his recent book, The Japanese Government and the Elderly, published by Princeton University Press, 1992. His other publications cover topics such as political culture, policy conflict, electoral behaviour, American old-age politics, the auto industry, bureaucratic organization, and U.S.-Japan relations.

#### **Corin, Ellen**

Dr. Corin is Director and Associate Professor in the Department of Psychiatry as well as being cross-appointed in the department of Anthropology at the University of McGill, and an Associate of the CIAR Program in Population Health. Recently, she was nominated on the board of the Academy of Arts and Sciences of the Royal Society of Canada. Dr. Corin's various research project interests are mainly oriented towards studying mental health and integrating those cultural and psychosocial perspectives into contemporary health research.

#### **Contandriopoulos, André-Pierre**

After Dr. Contandriopoulos joined the Health Administration Department at the University of Montreal his research interests in the health care system were reinforced. Showing dedicated interest in his field, he actively participated in establishing the "Groupe de Recherche Interdisciplinaire en Santé" (GRIS), an interdisciplinary research group, and Ph.D programs in community health. He is an Associate of the CIAR Program in Population Health. His research activities are closely linked with those of GRIS and include: the behaviour of doctors as providers of services and medical staff planning; the study of hospital organizations dealing with the analysis of the roles and responsibilities of various hospitals in a given area; the cost of health and the performance of the health care system, and the evaluation of interventions and the methodological aspects of such research.

#### **Crosson, Pierre**

Dr. Crosson is an economist at Resources for the Future, where he has worked for over 25 years. For the last decade or so his research has focused on the impacts of agriculture on the natural resource base and environment, emphasizing the consequences for the sustainability of the agricultural system, in the U.S. and globally. With colleagues at RFF he currently is engaged in developing a program of research on sustainable development. Dr. Crosson received his undergraduate degree from the University of Texas and his Ph.D. in economics from Columbia University.

#### **Culyer, Anthony**

Pro-Vice-Chancellor, University of York, England and Professor of Economics and Head of the Department of Economics and Related Studies, Professor Culyer is mainly concerned with definitions of equity in health, their consistency, and the extent of their conflict with other principles such as those of efficiency and freedom. He has been a member or chair of many U.K. advisory bodies, including the Research and Development Committee of the National Health Service. He is a member of the Research Advisory Group of the Ontario Workers' Compensation Institute and of the Methodological Advisory Group on Noneconomic Loss (Ontario Workers Compensation Board), and a member of the CIAR Population Health Program Advisory Committee.

#### **Cynader, Max**

Trained in Neuroscience at MIT, Dr. Cynader is Professor and Director of Research, Department of Ophthalmology, University of British Columbia, and a Fellow of the CIAR Programs in Artificial Intelligence and Robotics and Human Development. His main

research thrust involves the study of the mechanism by which the visual system processes information, and in normal and abnormal development of these mechanisms. One major area of study is to examine the mechanism underlying cortical plasticity. In the visual cortex it is now clear that what an organism sees in the first few months of postnatal life (the critical period) determines the wiring of the cortex. The nature and mechanisms of this critical period are key in human development. His research is at the forefront in understanding the basic molecular biology of how the nerve cells influence the body's host defence system (immune system).

#### **Delaney, Ken**

Research Director for the United Steelworkers of America in Canada, he is responsible for technical support for collective bargaining and work reorganization initiatives, the development of the Union's policy on economic and political matters and for providing the Union's locals with financial expertise when their employers are being restructured. His work on economic policy includes the Steelworker's paper, "Empowering Workers in the Global Economy". He is a founding member of the Sectoral Skills Council and a member of the Trade and Human Resources committees of the Canadian Steel Trade and Employment Congress. He has a graduate degree in industrial relations. Mr. Delaney is currently the Labour Advisor in residence at Queen's University.

#### **Evans, John**

Dr. Evans is a Canadian physician who received his undergraduate medical training at the University of Toronto and engaged in specialty training in internal medicine and cardiology in London, England, Boston and Toronto. Following five years as a member of the Department of Medicine at the University of Toronto and consulting physician at the Toronto General Hospital, he was appointed as the Founding Dean of the McMaster University Faculty of Medicine in Hamilton, Ontario. From 1972 to 1978, he was President of the University of Toronto. From 1972 to 1983, Dr. Evans served as Director of the Population, Health and Nutrition Department of the World Bank in Washington. In 1983, he returned to Toronto as Chair and CEO of Allelix Inc., a biotechnical research and development company. He is currently Chairman of the derivative company, Allelix Biopharmaceuticals Inc. In 1982, Dr. Evans became a member of the Rockefeller Foundation and is presently Chairman. Dr. Evans is a member of the CIAR Program in Population Health Advisory Committee.

#### **Evans, Robert**

Professor of Economics at the University of British Columbia, Vancouver and a member of the British Columbia Royal Commission on Health Care and Costs. He is involved in health policy issues in the province of British Columbia, Canada as a whole and internationally. Director and Manulife Fellow of the Population Health Program of the Canadian Institute for Advanced Research, he is considered one of the world's leading figures in health policy and health care. His book Strained Mercy is regarded as a classic.

#### **Frank, John**

Physician-epidemiologist interested in disease prevention, the efficient use of health care resources, and the major determinants of population health. Dr. Frank is an Associate Professor at the Faculty of Medicine, University of Toronto as well as the Director of Research/Senior Scientist at the Ontario Workers' Compensation Institute (OWCI), and a Fellow of the CIAR Program in Population Health. His past research projects include the examination of cost-effectiveness of screening tests used in medical care to detect asymptomatic disease and the simulation analysis of all-cause mortality by cholesterol levels for

Japanese American men. In addition, he has recently been developing at OWCI a research program utilizing multidisciplinary methods to shed light on the relative impact of both biophysical and psychosocial risk factors on work-related disability, particularly due to occupational back pain.

**Furukawa, Toshiyuki**

President of the Osaka National Hospital, Japan, and Emeritus Professor of the University of Tokyo, Dr. Furukawa has worked in simulation studies of circulation physiology and multivariate statistical analysis in medical diagnosis. At the Institute of Medical Electronics, Faculty of Medicine, University of Tokyo, he studied information analysis of clinical diagnosis, applications of artificial intelligence to medical consultation, and computer graphics of body surface potential maps and inverse estimation of epicardial potential maps.

**Glaeser, Edward**

Edward Glaeser received his Ph.D. at the University of Chicago in 1992 and is now an Assistant Professor at Harvard University. He teaches Urban Economics to Ph.D. students and Incentive Theory to undergraduates. Dr. Glaeser's interests range from urban issues to economic growth to labour markets. His published work includes papers on cities, economic growth and the economics of fairy tales.

**Goldstone, Jack**

Professor of Sociology at the University of California, Davis. He was previously professor of political science and sociology at Northwestern University, and Director of the Center for Comparative Research in History, Society and Culture at UC-Davis. His research explores political conflict, including the effects of population growth and resource distributions on political stability. His book Revolution and Rebellion in the Early Modern World, which demonstrates the role of population growth in the origins of the major revolutions and rebellions of Europe and China, won the 1993 Distinguished Publication Award of the American Sociological Association.

**Hertzman, Clyde**

Physician/epidemiologist and Director of the Division of Occupational and Environmental Health at the Faculty of Medicine at the University of British Columbia, Vancouver. He has recently completed a study for the World Bank on the subject of the declining health of East European populations. He is a Fellow of the CIAR Programs in Population Health and Human Development. Dr. Hertzman maintains a very broad research program which can be divided into three areas: population health, occupation and environmental health, and health policy in aging.

**Homer-Dixon, Thomas (Tad)**

Professor Homer-Dixon is Coordinator of the Peace and Conflict Studies Program at the University of Toronto, and Assistant Professor in the Department of Political Science. He is also Coordinator of an international research project on Environmental Change and Acute Conflict jointly organized by his Program and the American Academy of Arts and Sciences in Cambridge, Massachusetts. He received his doctorate in Political Science from MIT in 1989 and has written on a range of topics – from arms control to environmental policy, to the philosophy of the social sciences. Recent publications include "On the Threshold: Environmental changes as Causes of Acute Conflict," International Security (Fall 1991); and, with Jeffrey Boutwell and George Rathjens, "Environmental Change and Violent Conflict," Scientific American, February 1993.

**Ikegami, Naoki**

Professor of Health and Public Service Management at the Faculty of Policy Management, Keio University and Professor of Hospital and Medical Administration in Keio's School of Medicine. His interests are in comparative health care systems and he is currently undertaking a longitudinal study of about 1,000 subjects to evaluate the quality of long-term care facilities and to test the Resident Assessment Protocols (RAPs). He is also co-director of a US-Japan research group, funded by the Center for Global Partnership, which is studying why Japan's health care costs are low compared to the US. He continues his clinical activities as a consulting psychiatrist for a nursing home and a home for the mentally retarded.

**Kaplan, George**

Chief, Human Population Laboratory, California Department of Health Services, Berkeley. Research area of interest: low socioeconomic status and health risks. Dr. Kaplan heads the Alameda County Study in California which is one of the richest continuous studies of how the social environment influences the health status of individuals.

**Kuttner, Robert**

Robert Kuttner is co-editor of *The American Prospect*, a new liberal journal of politics and policy. He is author of four books, *The End of Laissez-Faire* (1991); *The life of the Party* (1987); *The Economic Illusion* (1984), and *Revolt of the Haves* (1980). Mr. Kuttner is one of four contributing columnists to *Business Week's* "Economic Viewpoint". Mr. Kuttner has taught at Brandeis, Boston University, the University of Massachusetts, and Harvard University's Institute of Politics. He has been a John Kennedy Fellow. His editorial column was the 1988 winner of the John Hancock Award for excellence in business and financial journalism. Previously, he has been a U.S. Senate investigator for the Committee on Banking, Housing and Urban Affairs, a national staff writer on the Washington Post, and executive director of President Carter's National Commission on Neighbourhoods.

**Lavis, John**

John Lavis is a Researcher in the Department of Clinical Epidemiology and Biostatistics at McMaster University. He studied Medicine at Queen's University, and Health Planning & Financing at the London School of Economics and Political Science. He has worked as a general practitioner, HIV primary care physician, and temporary advisor to the World Health Organization. Current research includes the economic evaluation of HIV prevention strategies in developing countries, methods of changing physician behaviour, development of needs based capitation formulae for Comprehensive Health Organizations, and planning human resources for health.

**Légaré, Jacques**

Professor of demography at the University of Montreal. He was trained in actuarial mathematics and in demography. Author of many books and scientific articles mainly in the field of historical demography of Canada before 1850 and of population aging. He was awarded, with his colleague Hubert Charbonneau, the Tyrrell Medal of the Royal Society of Canada in 1990. Presently, he is a member of the Council of the International Union for the Scientific Study of Population and Foreign Secretary of the Royal Society of Canada.

**Levine, Sol**

Dr. Levine is a Harvard School of Public Health Professor in the departments of Health Policy and Management, and a senior scientist at the Health Institute of New England Medical Center (NEMC) and former Vice-President of The Kaiser Foundation. He is a mem-

ber of the CIAR Program in Population Health Advisory Committee. His research focuses on the relationships between socioeconomic status and health status and he is developing research projects to produce data that will help leaders in government, business, labour, academia and the general public view social conditions and public policy as important – “even vital” – determinants of the entire population’s health.

#### **Lipsey, Richard**

Professor of Economics at Simon Fraser University, British Columbia and Director and Fellow of the CIAR Program in Economic Growth and Policy. Dr. Lipsey has been described as one of the one hundred greatest living economists since Keynes. During his career Dr. Lipsey has held numerous posts in professional organizations, public bodies and learned societies. His current research covers a) economic growth, theory and public policy; b) Canadian trade policy, and c) spatial economic and oligopoly theory. His most well-known text, An Introduction to Positive Economics has been translated into fourteen languages. He is currently researching and writing about Canada, Mexico and the U.S. which will be included in a book on the Case for North American Economic Integration (tentative title). He is the Director and Alcan Fellow of the Economic Growth Policy Program of the Canadian Institute for Advanced Research.

#### **Lomas, Jonathan**

Trained in experimental psychology at Oxford University, he is currently an Associate Professor of Health Policy Analysis in the Department of Clinical Epidemiology and Biostatistics, and the Coordinator of the Centre for Health Economics and Policy Analysis at McMaster University. He is an Associate of the CIAR Program in Population Health. His main interest is in the impact and transfer of research evidence for clinical and governmental decision-making. He has done research on methods for manpower planning, alternative funding and organizational arrangements for the delivery of care, quality assurance, technology assessment, consensus methodology, and appropriateness issues in obstetrical care (particularly cesarean section) and cardiology services.

#### **Mackenbach, Johan**

Professor, Department of Public Health and Social Medicine, Erasmus University, Rotterdam, Netherlands. Of key concern in Professor Mackenbach’s research is the effect of health care on avoidable mortality. Research studies have centred on mortality trends and medical care gains in life expectancy due to declines in mortality from conditions amenable to medical intervention in the Netherlands.

#### **Marmot, Michael**

Michael Marmot is Professor of Epidemiology and Public Health, University College, University of London, England and an Associate of the CIAR Program in Population Health. He is one of the world’s best population epidemiologists whose study of the Whitehall Civil Service in London gives excellent insight into how the environment in which individuals live and work determines their health. He has shown how the social environment creates risks for health and how the social environment creates risks for all major causes of death. Dr. Marmot heads the new Centre for Health and Society at University College, University of London.

#### **Morimoto, Kanehisa**

Professor of Hygiene and Preventative Medicine, Osaka University, School of Medicine, Japan. His research has focused on the mechanism underlying the information of chromoso-

mal alterations in the human body exposed to environmental genotoxics. He directs numerous research projects on disease prevention and health promotion. He has been carrying out experimental surveys on the assessment of comprehensive health potential, or "Health Age", considering somatic, psycho-mental, and genetic health status as a function of total lifestyle such as smoking and alcohol-drinking behaviour, sleeping and working patterns and perceived stress levels. He is also interested in understanding the dynamics among the environmental and health-related factors involved in the formation of quality of life.

#### **Mustard, James Fraser**

Dr. J. Fraser Mustard, President, Canadian Institute for Advanced Research holds a professional history embracing four careers - medicine, medical research, policy advisor to governments, and the creation of the CIAR. After earning his M.D. from the University of Toronto, and Ph.D. from the University of Cambridge, Dr. Mustard moved from the medical faculty of the University of Toronto to help establish the new school of Medicine and Health Sciences at McMaster University. In 1982, he took on the task of creating and establishing a unique Canadian institution, the Canadian Institute for Advanced Research. For his contributions to medical research, furthering the understanding of science and technology, and his pioneering leadership, he has been widely recognized and honoured.

#### **Nicholson, Peter**

Peter Nicholson, Advisory Committee Member for the CIAR Economic Growth and Policy Program is Senior Vice President and Executive Assistant to the Chairman, Bank of Nova Scotia. Nicholson studied Physics at Dalhousie and earned a Ph. D. from Stanford. His professional career has encompassed diverse activities: university educator, federal government policy advisor, MPP, Nova Scotia Legislature, and federal task forces member. He is a member of the National Advisory Board on Science and Technology, the Executive Committee of the Arms Control Centre, Director of the Couchiching Institute, the Chairman of the Fields Institute for Mathematics as well as serving on the Campaign Cabinet of the Metro Toronto United Way.

#### **Nishimura, Shuzo**

Dr. Nishimura is Professor of Hygiene and Insurance Economics, Faculty of Economics, Kyoto University. He is one of the few leading health economists in Japan. He is also a member of several committees of the Ministry of Health and Welfare for the formation of health policy in Japan.

#### **Offord, David (Dan)**

Dr. Offord is a child psychiatrist with major interests in epidemiology and prevention. He is a Professor of Psychiatry at McMaster University, Head of the Division of Child Psychiatry, Research Director of the Chedoke Child and Family Centre, and has been Director of the Child Epidemiology Unit since its inception in 1980. Dr. Offord is an Associate of the CIAR Program in Human Development. He is also Director of the State-of-the-Child Research Unit funded by the Laidlaw Foundation, where the major area is to determine what data should be collected on a regular basis on children and youth in Ontario to inform and evaluate policy. Dr. Offord played a leading role in the Ontario Child Health Study. The major area of the initial cross-sectional study carried out in 1983 was to obtain unbiased principle estimates of the prevalence of emotional and behavioral problems, physical health problems and substance use among Ontario children 4 to 16 years of age. The follow-up study carried out in 1987, focused on issues of outcome, prognosis and risk.

**Okamoto, Akira**

Chief Physician, Department of Pediatrics, AIKU Hospital, Japan. He was involved in the WHO/Healthy Cities Project to Japanese local authorities and public health professionals. He Heads a survey team on Community Based Maternal and Child Health Project in Indonesia and a survey team on Information Networks between Japan and Thailand in the field of child welfare. He is a Member of the steering committee for the Asian Maternal and Child Health Workshop for health officials in Asian countries.

**Power, Christine**

Dr. Power is Senior Lecturer in the Division of Public Health: Epidemiology and Biostatistics Institute of Child Health, London, England and a Scholar of the CIAR Program in Population Health. In her research, Dr. Power uses longitudinal data from the British 1958 birth cohort study to investigate influences on health and development. A major part of her work involves the documentation and explanation of social differences in health in early adulthood. She examines the many dimensions of "health" (including physical development and mental health) in conjunction with various potential health determinants (such as parents and family, education, behaviour and socio-economic circumstances).

**Renaud, Marc**

Professor of Sociology at the University of Montreal, Director of "le Groupe de recherche sur les aspects de la prévention" (GRASP), Chair of the Quebec Council on Social Research, Vice-President of the Canadian Institute for Advanced Research and a Fellow of the CIAR Program in Population Health. Dr. Renaud's research probes into the political implications of the determinants of health. He questions why most resources go to doctors and hospitals and demonstrates in his works that health is above all determined by one's employment status, work climate, supportive relationships with others, and the capacity to cope in case of difficulty.

**Romer, Paul**

Professor of Economics at the University of California at Berkeley and the Royal Bank Fellow of the Canadian Institute for Advanced Research. Economic growth has been the focus of Dr. Romer's research since graduate school and his new ideas have given him a reputation as one of the young leaders in the concept of ideas and economic growth. He has authored numerous articles on the dynamics of evolving economies, including "Increasing Returns and new Developments in the Theory of Growth" and "Human Capital and Growth: Theory and Evidence:." As The Economist (January 1992) reported, "his new concepts of economic growth will profoundly change our understanding of how nations create wealth and the social environment.

**Roos, Leslie**

Leslie L. Roos, Ph.D., Director of the Manitoba Research Data Base, is a Professor in the Department of Community Health Sciences, Faculty of Medicine, University of Manitoba, a National Health Scientist, and an Associate of the CIAR Program in Population Health. Dr. Roos graduated from Stanford University and received his doctoral degree from the Massachusetts Institute of Technology. He has contributed to a wide variety of books and journals on topics such as the use of administrative data bases, comparisons of health care outcomes between the United States and Canada, population-based health trends, and outcomes of care. He has particular interests in record linkage and maximizing the research uses of large data bases. A major project involving the Manitoba Health data bank and Statistics Canada data will make feasible study of a wide range of research and policy questions. This

work will improve our understanding of such topics as: the relationship between the burdens of illness and the health care system and the feasibility of identifying particular groups of mothers at higher risk of low birth weight or sickly infants.

#### **Roos, Noralou**

Trained in political science at Stanford University and at the Massachusetts Institute of Technology. Dr. Roos is currently a Professor at the University of Manitoba and Director of the Manitoba Centre for Health Policy and Evaluation and an Associate of the CIAR Program in Population Health. Her research focuses on the effectiveness of medical practice. Current work deals with the issues facing the health care system such as: how do you make informed decisions pertinent to managing the health care system; how do you ensure the delivery of high quality of care as funds are directed out of the health care system, and how do you evaluate the effectiveness of new technology?

#### **Rosenberg, Nathan**

Nathan Rosenberg is the Fairleigh S. Dickinson, Jr. Professor of Public Policy at the Department of Economics, Stanford University, and is a former chairman of the department. He is an Associate of the CIAR Program in Economic Growth and Policy. Before assuming his current position, Dr. Rosenberg served as the chairman of the Stanford Program on Values, Technology and Society and as director of the Stanford Program on Public Policy. Dr. Rosenberg has written extensively on the economics of technological change. He served as editor of the "Journal of Economic History" between 1972 and 1974, and in 1981 Dr. Rosenberg became a fellow of the American Academy of Arts and Sciences. He was the Pitt Professor of American History and Institutions at Cambridge University in 1990-1991. Professor Rosenberg has published many books including, most recently, Inside the Black Box, Cambridge University Press, 1982, How the West Grew Rich, with L.E. Birdzell, Jr., Basic Books, 1986, and Technology and the Pursuit of Economic Growth, with David Mowery, Cambridge University Press, 1989.

#### **Smil, Vaclav**

Dr. Smil is a Professor at the University of Manitoba in Winnipeg. He was educated at the Carolinum University in Prague (RNDr 1965) and at the Pennsylvania State University (PhD 1972). His interdisciplinary research concentrates on interactions of environment, energy, food, economy, population and public policy. He is the author of more than 150 papers and books. His latest books are General Energetics (John Wiley 1991), China's Environmental Crisis (M.E. Sharpe 1993) and Global Ecology (Routledge 1993).

#### **Stoddart, Gregory**

Trained in economics, Dr. Stoddart is a Professor in the Department of Epidemiology and Biostatistics, Faculty of Health Sciences at McMaster University and a Fellow of the CIAR Program in Population Health. Past research focused on the development and application of methods for the economic evaluation of specific health services and procedures – cost effectiveness and cost-benefit studies of health care interventions. In more recent years, his work has become increasingly conceptual relative to the determinants of health. Issues include socio-economic gradients in health status, psychoneuroimmunological responses, and genetic predisposition, to cite just a few examples. The paper by R. Evans and G. Stoddart "Producing Health, Consuming Health Care", has had a major impact on thinking in North America.

**Suomi, Stephen**

Trained as a psychologist, Dr. Suomi currently is Chief of the Laboratory of Comparative Ethology, National Institute of Child Health and Human Development (NIH) and an Associate of the CIAR Program in Human Development. He has received international recognition for his extensive research on biobehavioural development in rhesus monkeys and other nonhuman primate species. His research at the University of Wisconsin led to his election as a Fellow of the American Association for the Advancement of Science "for major contributions to understanding of nonhuman primates." Since joining NIH in 1983, he has described interactions between genetic and environmental factors that shape individual biobehavioural development, characterized both behavioral and physiological features of distinctive rhesus monkey phenotypes, and demonstrated the adaptive significance of these different phenotypes in naturalistic settings.

**Syme, Leonard**

Trained in medical sociology at Yale, Dr. Syme is Professor of Epidemiology at the School of Public Health, University of California at Berkeley. He is a member of the CIAR Program in Population Health Advisory Committee. He has been a major leader in the United States in areas concerning the social environment, education and health. He has held numerous positions on advisory committees and was a member of several Expert Advisory Panels for the World Health Organization on issues such as Drug Dependence, Mental Health, Non-communicable Diseases and Occupational Health.

**Takane, Yoshio**

Dr. Takane is a Professor of Psychology at McGill University specializing in behavioral statistics and mathematical models. Aspects of his research include multidimensional scaling, analysis of nonmetric data, constrained principal component analysis, structural equation models and other multivariate data analysis techniques, and applications for these techniques to problems in psychology and other social and behavioral sciences. He is a Past President of the Psychometric Society, and a recipient of an award from the Behaviourmetric Society of Japan.

**Takeuchi, Kei**

Professor, Research Center for Advanced Science and Technology, Faculty of Economics, University of Tokyo. His main areas of research are in mathematical statistics, econometrics, and science and technology. His publications include Asymptotic Theory of Statistical Estimation. A joint work with M. Akahira, 1983, Springer Publ., and Considering the Earth (Jpn), Editor, 1992, Mita-Press.

**Tarlov, Alvin**

Director of the Division of Health Improvement at the Health Institute at New England Medical Center. He received his BS from Dartmouth and his medical degree from the University of Chicago. He spent five years in haematologic research before becoming Professor and Chairman of the Department of Medicine at the University of Chicago in 1968 – a post he held for 13 years. In 1984 he became President of the Henry J. Kaiser Family Foundation. Currently he is involved in the development and dissemination of outcomes measures in medical care.

**Wilk, Martin**

Dr. Wilk received his Ph.D. in statistics at Iowa State University. During his career Dr. Wilk held numerous academic posts. From 1955-57 he was Research Associate and Assistant

Director of the Statistics Techniques Group, Princeton University, and from 1959 to 1963, Professor and Director of Research in Statistics at Rutgers University. After serving with Bell Telephone Labs, he joined American Telephone and Telegraph Co. in 1970 as Director, Corporate Modelling Research and from 1976 to 1980 held the position of Assistant Vice President and Director of Corporate Planning. In 1980 he joined the Government of Canada as Chief Statistician of Canada until 1985. Dr. Wilk has served on numerous advisory bodies and task forces. Most recently, he chaired the National Task Force on Health Information (1990-91). He is a member of the CIAR Research Council and is a member of the Advisory Committee for the CIAR Program in Population Health and the Program in Human Development.

**Wilkinson, Richard**

Dr. Wilkinson is a senior research fellow at the Trafford centre for Medical research, University of Sussex, England. His research interests focus on income distribution and life expectancy. For example, he is interested in the cross-sectional relation between GNP per head and to assess whether changes in income distribution over time are related to changes in mortality in developing countries.

**Wolfson, Michael**

Michael Wolfson is a Fellow of the CIAR Program in Population Health and an Associate of the CIAR Program in Economic Growth and Policy. After obtaining his Ph.D in Economics from Cambridge University, he joined the Canadian Public Service in 1974 where he was involved in program review and evaluation, tax policy and pension policy. Since 1985, he has been working at Statistics Canada and is currently Director General of the Institutions and Social Statistics Branch. His recent research interests include income distribution, tax/transfer and pension policy simulation analysis, corporate structure, demographic microsimulation, and health information and determinants.

**Acknowledgements:**

The CIAR and Honda Foundation wish to thank Lori Dunn for her timeless and skilled efforts in compiling this volume of proceedings.

# A Transectional Comparison of Average Life Expectancy

---

*Toshiyuki Furukawa*

## 1. INTRODUCTION

In this information-inflated era, how to "analyze" this information is an important subject of the current period. Although it seems people have become wiser with the increase in available information, in reality people are being twisted around the excessive knowledge and information that surround them. Overflow of false information is also apparent in the relationship between diet-nutrition and health care. Phobia against excessive fat intake, over-expectation on natural foods, are some of the curious behaviors observed. Some say that the prosperity of mankind was a result of accumulation of extra energy consumption as body fat. Under conditions in which men had to hunt in order to earn their daily food, the development of the cerebral cortex, as well as the creation of civilization and knowledge, were unexpected.

What contributed to historical gains in longevity? The answer to this question could be the mass production of soap and plate glass following the sudden rise in modern industry. Soap contributed to the eradication of pathogens and parasites on the body and clothes. Glass gathered in the blessings from the sun and helped eliminate pathogenic microorganisms.

In this report, the fact that the abolition of ignorance and poverty was a requisite in the advanced nations. At the same time, the primary factors in longevity will be discussed, using the multivariate analytic techniques, especially cluster analysis.

## 2. SIMPLE CORRELATIONS OF SOCIAL INDEXES AND AVERAGE LIFE EXPECTANCY

Average life expectancy is still used as one of the concrete indexes of health level for comparison of health and hygiene standards between eras as well as between different nations.

Whether medicine played a role in the prolongation of the expectation of life span is a subject medical scientists have been long questioning. Because medicine prior to the modern age was not easily available to the general public, it was not powerful as to influence average life expectancy. Although there were techniques that contributed to the prevention of epidemics such as smallpox, regardless of disparity in wealth, a physician with sound judgment knows it was prosperity of the society and not medical technology that contributed to the prolongation of expectancy of lifespan. What decreased infant mortality was not pharmaceuticals; it was sufficient nutrition.

The factors involved in average life expectancy cannot be found simply by investigating their correlation. This "correlation" tends to be improperly used in the development of these discussions. For instance, there is a close relationship between the prolongation in average life expectancy of the Japanese and the increase in the production of antibiotics. Does anyone recommend taking a large amount of antibiotics to enjoy a long life? Although nutritionists claim that obesity leads to adult diseases, including diabetes, arteriosclerosis, and so on. But the fact is people are getting obese and at the same time living longer.

There are still numerous more social indexes which correlate with average life expectancy. There is no need to dwell upon so many relationships. A glance on the correlograms is better. The most notable correlations were as follows;

- National income per capita (Fig.1)
- Population per number of physician (Fig.2)
- Circulation of daily newspaper (Fig.3)
- Calorie intake per capita per day (Fig.4)

Another indicator of population health is death rate. The next two examples show the relation between medical facilities and death rate.

- Population per number of physician (Fig.5)
- Population per hospital beds (Fig.6)

As the other indicator of health status, the birth rate shows fairly high correlation to the illiteracy rate and the inverse correlation to the calorie intake per capita per day.

- Illiteracy rate (Fig.7)
- Calorie intake per capita per day (Fig.8)

Although these figures and average life expectancy are related, their interpretation is not straightforward. Social phenomena made up of various complex factors entangle and influence each other. If we go so far as to say there is nothing on earth that is not related to anything, it is almost impossible to find something that is not correlated. Mathematically, this concept of the world can be expressed in terms of multivariate analysis. There is a danger of making errors by marshalling simple correlational coefficients.

### **3. CLUSTER ANALYSIS OF SOCIAL INDEXES - AN ANALYSIS INCLUDING ADVANCED AND DEVELOPING COUNTRIES**

When the correlation between average life expectancy and various social indexes are studied, correlation is seen between a variety of indexes. It is harder to find something that is not correlated. The number of physicians, medical cost, number of hospital beds, etc., are all related with average life expectancy. In such a case, it is difficult to see through the essence of the relationship with an innocent "correlation supremacy principle". In order to objectively analyze the relationship between average life expectancy and social indexes, a multivariate analysis must be performed. Cluster analysis is a part of this method.

Cluster analysis is often presented as a dendrogram, where the shorter the distance of each items, the closer the relationship of these items. So, for example, we can find the expectation of life spans of male and female stand close side by side.

The data analyzed are mainly from "The International Data Book" edited by the Japanese Prime Minister's Office, which includes many quotations from international authentic statistics including WHO's reports. The author had already carried out a cluster analysis of international social indicators of 1975 in 1979. At that time, the two kinds of apparently dominant interrelations were found. One was the relation between average life expectancy and generalization of the wealth and the information in a nation. The second one was the relation between death rates (general, infant, and tuberculosis death rate) and medical abilities (population per number of physician, and hospital bed), and in addition, more closer relation between birth rate and illiteracy rate.

In this report, in order to confirm the validity of the findings on 1975's statistics, the same cluster analysis was done on the 1970's, 1980's and 1990's statistics.

### 3-1 Life-Span-Related Social Indicators

Looking through the results of cluster analysis on the data of 1970's, 1980's, and 1990's, there exist peculiar subclusters of characteristic social indicators. In general, a subcluster consists from the indicators of the wealth and the amount of information in each nation, including the expectancy of life span. Another subcluster consists from the indicators which correspond to a part of death rates and illiteracy rate.

#### (1) 1970's analysis:

In the case of 1970's, the following indicators showed a close relation to the expectancy of life span (Fig.9).

- Calorie intake and protein intake per capita per day
- Cause specific death rates by cerebrovascular diseases
- Cause specific death rates by heart diseases
- Expenditure on social security per GNP
- Enrolled students in higher education rate male and female
- Number of TV sets per capita
- Number of radio receivers per capita
- Circulation of daily newspapers per capita
- Electric energy production per capita

At the same time, the following social indicators tend to flock together in the near cluster.

- Birth rate
- Illiteracy rate
- Population rate of agriculture and fishing
- Infant mortality
- Unemployment rate
- Cause specific death by tuberculosis

On the cluster of 1970's social indicators, there are few relations between population per number of physicians or hospital beds and the death rate.

#### (2) 1980's analysis:

In the case of 1980's, the following indicators show a close relation to the expectancy of life span (Fig.10).

- Enrolled students in higher education rate male and female
- Energy consumption per capita
- Number of radio receivers per capita
- Number of TV sets per capita
- Calorie intake and protein intake per capita per day
- Cause specific death rates by cerebrovascular diseases
- Cause specific death rates by heart diseases
- Expenditure on social security per GNP
- Circulation of daily newspapers per capita
- National income per capita
- Percent of dwellings with tap water
- Percent of dwellings with flush toilet
- Expenditure on education per GNP

There is a closely related group of social indicators in 1980's cluster, as indicated below.

- Birth rate
- Infant mortality
- Illiteracy rate
- Population per number of hospital beds
- Population per number of physicians
- Death rate
- Agriculture production per GDP
- Unemployment rate
- Cause-specific death by tuberculosis

In this case, i.e., cluster structure of social indicators on 1980's statistics is very much close to the results of 1975's. There was a clear evidence emphasizing the efforts of medicine to the health service of population.

### **(3) 1990's Analysis:**

In the case of 1990's, the following indicators show a close relation to the expectancy of life span (Fig.11).

- Enrolled students in higher education rate male and female
- Number of radio receivers per capita
- Number of TV sets per capita
- Expenditure on education per GNP
- Circulation of daily newspapers per capita
- National income per capita
- Percent of dwellings with tap water
- Percent of dwellings with flush toilet
- Energy consumption per capita
- Electric energy production per capita
- Calorie intake and protein intake per capita per day
- Cause specific death rates by cerebrovascular diseases
- Cause specific death rates by heart diseases
- Expenditure on social security per GNP

There is a closely related group of social indicators in 1990's cluster, as follows:

- Birth rate
- Infant mortality
- Illiteracy rate
- Population rate of agriculture and fishing
- Agriculture production per GDP
- Cause specific death rate by tuberculosis
- Cause specific death by tuberculosis

In this 1990's case, the structure of cluster of social indicators relate to the expectancy of life span, and to the birth rate, infant mortality and illiteracy rate, does exist. However, the population per number of physicians and of hospital beds slipped out of place where they were in 1975's and 1980's.

### 3-2 Interpretation Of Cluster Analysis

As a result of these analyses, it is clear that prosperity and the amount of information are the social indicators most closely related to expectancy of life span. For example, what indicators such as the enrolled students in higher education rate, the number of radio receivers and TV sets per capita, circulation of daily newspaper, as well as education and culture, have as a common term, is the amount of information available to the society. On the other hand, these indicators cannot be realized unless the society is prosperous. National income, income tax, and social security are direct indexes of prosperity. The amount of electric power generation, energy consumption, calorie intake, and protein intake also reflect prosperity. Tap water supply and prevalence rate of flushing toilets are indicators of prosperity more than of hygiene. Although the number of crimes is also related to the expectancy of life span in the analysis on 1975 statistics, ironically, this is an indicator of modernization of the society.

Although one can say that the expectancy of life span itself is an index of prosperity, it more or less reflects health standard, since mortality by heart diseases and cerebrovascular diseases are the two factors most closely related to the expectancy of life span. Death by neoplasms, which terrorizes modern people, is not related with expectancy of life span and has been already estimated in other mathematical analysis. Even if cancer in people over 65 years of age is subjugated, prolongation of the expectancy of life span of approximately only one year is expected. Death by cancer has increased mostly as a result of the increase in the number of people who survive up to an age in which cancer can be caused.

Is medicine and medical care then powerless? In the cluster analysis on social indicators, a cluster consisting of health indexes has been found on a branch distant from the expectancy of life span. Mortality rate, number of physicians per capita, number of hospital beds per capita, infant mortality rate, tuberculosis mortality rate, birth rate, and illiteracy rate belong to this group - as far as the data in 1975 and 1980 are concerned. For the first time point, it is assumed that the power of medicine has helped lower mortality rate. Physicians run to the aid of unfortunate people who have been possessed by the God of death. This is rather a favorable view of the actions of the physicians, but (although their competency and capabilities may actually be the result of making deals with the God of death to cover up of their inferiority in strength) medicine is definitely a method to save lives. It is also interesting that birth rate and illiteracy rate are closely related. A saying in Japan "the poor bear many children" should actually be "an illiterate bears many children". When we think of the harmonious future between mankind and nature, family planning on the global scale has now reached a point in which not a moment of delay is permitted.

It was an unexpected result that the level of medical services did not have a substantial effect on several kinds of death rates, in the 1990's analysis. In 1970's analysis, the health and medical programs in the developing countries were relatively poor and not systematic. So, it is reasonable to infer that the power of medicine to assure lower mortality rate is limited in the case of advanced countries. However in the 1990's analysis, a convincing relationship between medical facilities and death rates disappeared. Between 1975 to 1980, the role of medical facilities was clearly characterized as the strong helping hand for the patient, but, the 1990's data showed more a chaotic structure as far as medicine is concerned. Possibly a reason of this incredible metamorphoses may be the medical service restructuring in advanced countries - for instance, restriction of the number of hospital beds, or diversion of hospital system to institutional care for the aged, and so on. At the same time, it must be pointed out that there has been an exodus of a large number of trained physicians from developing countries. Anyway, there remain unsolved problems concerning the best policies to uplift the national level of health care and medical services not only in developing countries, but also advanced countries.

It is the prerogative of the researchers to interpret the result of mathematical analysis, especially of multivariate statistical analysis. If one is allowed to interpret in one's own way, this author is certain that information and prosperity are the two major factors that protect human health. The advanced nations have healthily enjoyed their long expectancy of life span by abolishing ignorance and poverty.

#### **4 ANOTHER CLUSTER ANALYSIS OF SOCIAL INDICATORS - AN ANALYSIS OF THE SELECTED SOCIAL INDICATORS**

From the above analysis, it seems that prosperity and information do not influence the expectancy of life span in advanced nations. Take the data of the EC nations for example: even though they are all equally prosperous and abundant in information, there is quite a difference in the expectancy of life span. Thus, another cluster analysis was performed strictly on social indicators which belong to the "expectancy of life span - prosperity - information" cluster.

In order to make clear the macroscopic trend occurring from 1970 to 1990, another cluster analysis was done, using the items related to the prosperity and the amount of information available to the society. As an attempt to simplify the analysis, the closely related items, (for instance, the expectation of life span) are limited to values for males. As the indicator of nutritional level, protein uptake is preserved and calorie uptake is omitted.

##### **4-1 Analysis Of Social Indicators In Advanced Nations**

Looking at the data from 1970 to 1990, in the case of advanced countries, no obvious trend is seen in this 20 years. That means the relationship of the expectation of life span to the prosperity and the socially available amount of information has been unalterable. But at the same time, it is remarkable that the strength of the relationship between each pair of social indicators has been becoming looser decade by decade (Fig.12-14).

##### **4-2 Analysis Of Social Indicators In Developing Countries**

The same trend also is seen in the analysis of developing countries. These findings can be interpreted as indicating that not only the prosperity but also the amount of information has been saturated in the advanced society, and the current of the times is followed by the developing countries. That means, in the developing countries, there remains relatively closer relationships between the expectation of life span and prosperity, and the amount of information, but the relation is beginning to loosen. There are changes of tightness of the relationship between social indicators. However, the basic cluster structures are extremely similar to each other in advanced and developing nations (Fig.15-17).

#### **5 MULTIVARIATE ANALYSES OF SOCIAL INDICATORS**

The results of the cluster analysis done separately on advanced and developing countries are basically the same. Regardless of the stage of development, the factors which influence expectancy of life span are "prosperity" and "information". In the advanced nations, however, it seems that expectancy of life span has somewhat separated from "prosperity" and "information". There seem to be other factors which influence expectancy of life span between the advanced nations.

## 5-1 Multiple Regression Analysis Of Social Indicators

In the regression analysis, expectancy of life span of each nation was regressed based on under 10 social indicators.

Looking into details, a positive effect was seen for social indicators of prosperity such as GNP, amount of electronic power generation, protein intake and total value of export. A high contribution was observed for all social indicators demonstrating "amount of information", such as the enrolled students in higher education and circulation of newspaper per population. Functions of medical care service were suspected of making important contributions, i.e., the population per number of hospital beds. Contributions of mortality rate were found to be low as well (Fig.18-20).

## 5-2 Principal Component Analysis Of Social Indicators

Analysis on principle components of social indicators was also done. Principle component analysis is a mathematical technique to induce a "summarized explanation" on subjects which can only be recognized as the multiple variables. Almost everything can be explained with the 1st through 3rd major principle components in the analysis done on advanced and developing countries.

The principal component analysis of the mixed group of advanced and developing countries, the first principal component showed an obvious relationship with expectancy of life span (Fig.21). In advanced nations, however, although a positive correlation exists, there is too much scattering (Fig.22). The relationship between expectancy of life span and principal components of social indicators in advanced and developing countries demonstrates satisfactory regression that is almost as equivalent to the result of the multiple regression analysis.

In analyzing the typical indexes of health level such as average life expectancy and health itself, it is obvious that a relationship existed between these indexes and the culture and prosperity levels in a nation. However, it is strongly suggested that this relationship was not a homogeneous one.

There is a controversial question. "Why is there a difference in expectancy of life span between countries in Europe, North America and Japan, where people share a high living standard?" A robust answer is not easy. The meaning of the principal components, backed up by strong facts, unfortunately cannot be interpreted with the types of social indicators that are easily obtained. Therefore, in order to investigate on the difference in expectancy of life span between the advanced nations, it is necessary to spread the aims of multiple data analysis to the items not yet taken care of - for instance, more detailed surveys on lifestyles, including food and so on. Such detailed comparative studies should be done.

## 6 CONCLUDING DISCUSSION

Because tedious information often mingles with multidimensional information, it is impossible to see essential points just by the power of human instinct. The solution is to apply a multivariate analysis. Although multivariate analysis of expectancy of life span and social indicators elucidated the relationship between expectancy of life span, prosperity, and information, the difference in expectancy of life span within the advanced nations could not be satisfactorily explained. This is because the variety of social indicators listed on the International Statistics Handbook do not include the information necessary for a precise analysis. There is a strong need for new research and analysis with international cooperation.

The explosive increase in population and chronic famine in the third world, together with ozone hole, acid rain, and global warming (green house effect) are the factors of megacrisis mankind has to face. Supply of accurate information is important for countermeasures against each of these problems. In the cooperation with the developing countries, "abolishment of ignorance and poverty" which we, the advanced nations have experienced, should be possible.

## 7 REFERENCES

1. Furukawa, T.: A cluster analysis of international social indicators of 1975's, Symposium "Culture and Life Span". Jpn.Soc.Gerontol., Oct. 5, 1979, Osaka.
2. Furukawa, T.: Relation of Culture and Life Span from an Analysis of Social Indicators, in "Social Structure in Aging World", (Jpn) edited by Oota, K., Abe, H., and Furukawa, T., Science Pub. Co., Tokyo, 1981.
3. Furukawa, T.: A Multivariate Analysis of the Relation between Health and Social Indicators, The International DISCOVER Symposium, May 10-14, 1982, Columbus, Ohio.
4. Furukawa, T.: Research on the Evaluation of National Health Status in Japan, (Jpn), Governmental Report of the Ministry of Health and Welfare, 1991-1992.
5. Furukawa, T.: Does Elongation of Japanese Life Span come to End?, in "Health and Disease in Aged" (Jpn) edited by Itakura H., Yoshimura, M., Yasumoto, N., Koseikan Pub.Co., 1992, Tokyo.

## 8. APPENDIX - NOTE ON THE MULTIVARIATE ANALYSES OF SOCIAL INDICATORS

The source data of multiple regression analysis and principal component analysis was "The International Data Book in 1990".

The multiple regression equations described in Chapter 5-1 appeared in the following:

- 1 A multiple regression equation on social indicators in some of advanced countries (Fig.18):

Referring to the values of simple correlation coefficients, 9 social indicators are used in an equation

Expectation of Life Span

$$\begin{aligned}
 &= -0.012 \\
 &+ 0.504 \times \text{Circulation of Daily Newspaper} \\
 &- 1.207 \times \text{Electric Power Generation} \\
 &+ 0.095 \times \text{GNP} \\
 &- 0.451 \times \text{Protein Intake} \\
 &+ 0.254 \times \text{Students in Higher Education} \quad \text{♂} \\
 &+ 1.225 \times \text{Numbers of Telephone}
 \end{aligned}$$

$$\begin{aligned}
 &+ 0.384 \times \text{Total Export Value} \\
 &- 0.218 \times \text{Population Density} \\
 &- 0.983 \times \text{Number of Production of Books}
 \end{aligned}$$

$$(r = 0.835, p = 0.144, F = 1.490, N = 15)$$

2. A multiple regression equation on social indicators in some of advanced countries (Fig.19):

Referring to the values of simple correlation coefficients, 8 social indicators with regard to medical services are used in an equation.

Expectation of Life Span

$$\begin{aligned}
 &= + 0.279 \\
 &- 0.105 \times \text{Electric Power Generation} \\
 &+ 0.051 \times \text{Protein Intake} \\
 &- 0.085 \times \text{GNP} \\
 &- 0.177 \times \text{Students in Higher Education} \quad \sigma^2 \\
 &- 0.026 \times \text{Total Export Value} \\
 &- 0.743 \times \text{Death Rate} \\
 &+ 0.381 \times \text{Population per Hospital Beds} \\
 &- 0.617 \times \text{Population per Physician}
 \end{aligned}$$

$$(r = 0.936, p = 0.074, F = 3.912, N = 15)$$

3. A multiple regression equation on social indicators in mixed group of advanced and developing countries (Fig. 20):

Almost same to the above analysis, 9 social indicators with regard to medical services are used in an equation.

Expectation of Life Span

$$\begin{aligned}
 &= + 0.159 \\
 &+ 0.089 \times \text{Circulation of Daily Newspaper} \\
 &- 0.089 \times \text{Electric Power Generation} \\
 &+ 0.113 \times \text{Protein Intake} \\
 &- 0.131 \times \text{GNP} \\
 &+ 0.219 \times \text{Students in Higher Education} \quad \sigma^2 \\
 &+ 0.069 \times \text{Total Export Value} \\
 &- 0.019 \times \text{Death Rate} \\
 &- 0.390 \times \text{Population per Hospital Beds} \\
 &- 0.040 \times \text{Population per Physician}
 \end{aligned}$$

$$(r = 0.978, p < 0.0001, F = 33.611, N = 24)$$

In this analysis, the social indicators used were described in a normalized scale, i.e., mean value = 0,  $\sigma = 1.0$ . For the convenience sake of the leaders, on the scattergrams the values in natural scale were also indicated on opposite axis.

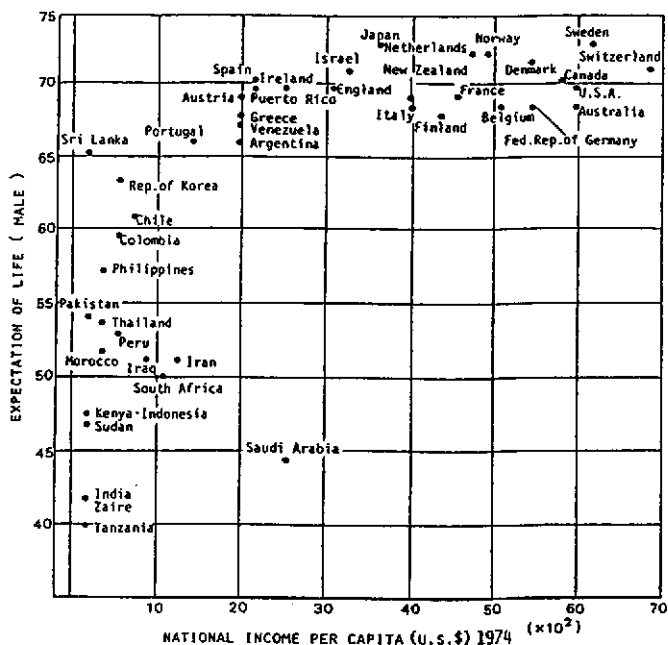


FIG. 1 RELATIONSHIP BETWEEN EXPECTATION OF LIFE (MALE) AND NATIONAL INCOME PER CAPITA. (  $N=49$ ,  $r=0.735$  )

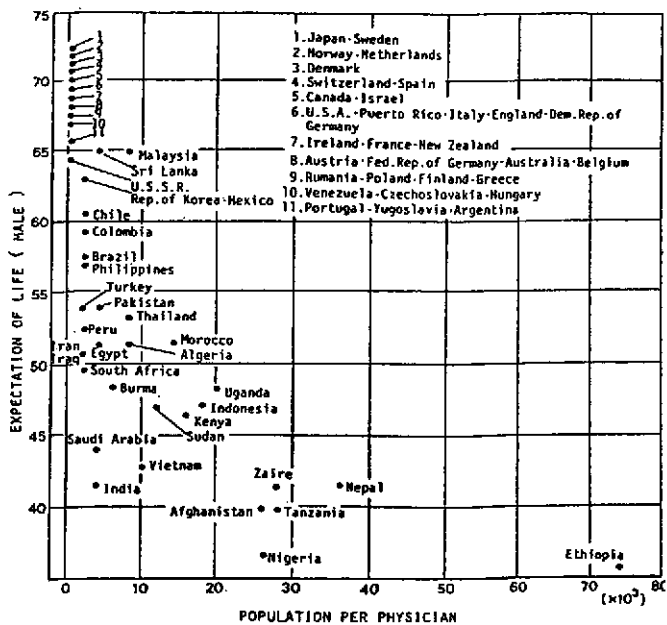


FIG. 2 RELATIONSHIP BETWEEN EXPECTATION OF LIFE (MALE) AND POPULATION PER PHYSICIAN. (  $N=64$ ,  $r=-0.714$  )

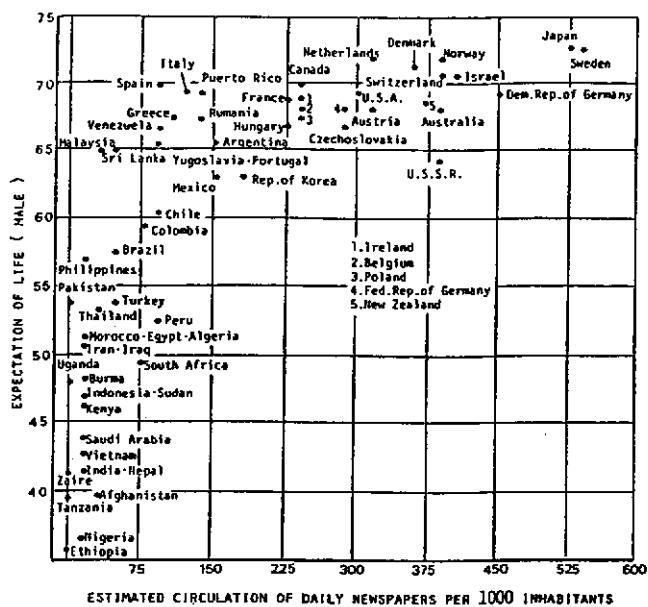


FIG. 3 RELATIONSHIP BETWEEN EXPECTATION OF LIFE (MALE) AND ESTIMATED CIRCULATION OF DAILY NEWSPAPERS PER 1000 INHABITANTS. (N=63,  $r=0.762$ )

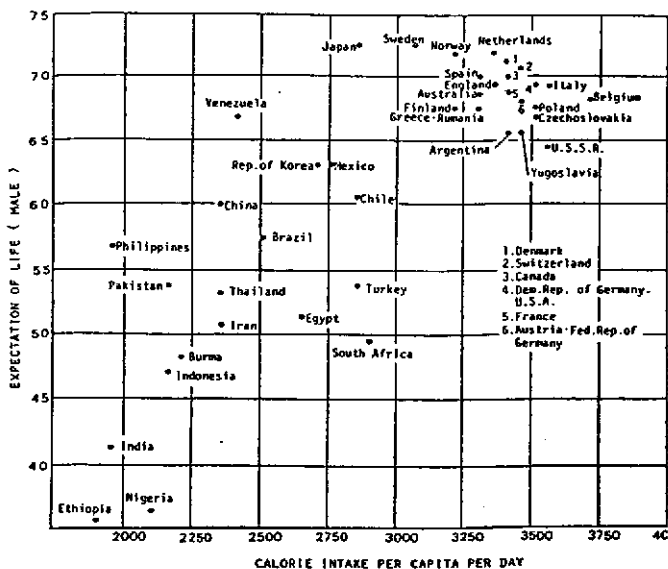
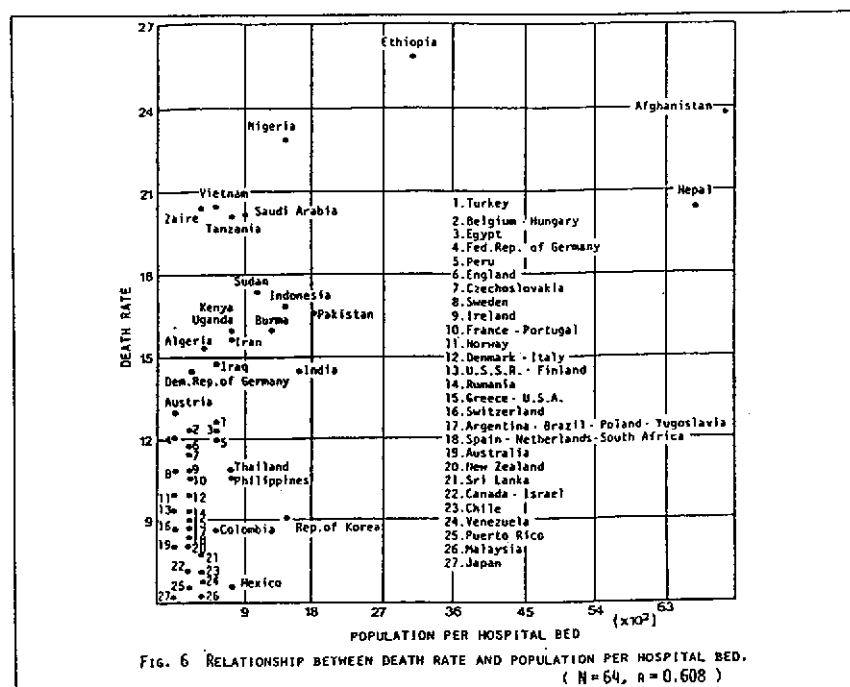
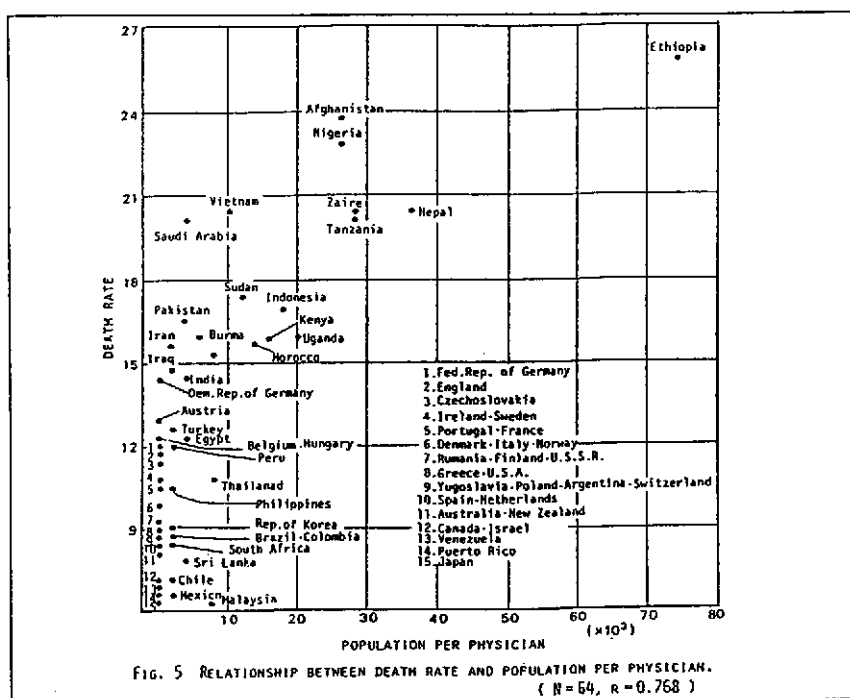


FIG. 4 RELATIONSHIP BETWEEN EXPECTATION OF LIFE (MALE) AND CALORIE INTAKE PER CAPITA PER DAY. (N=43,  $r=0.824$ )



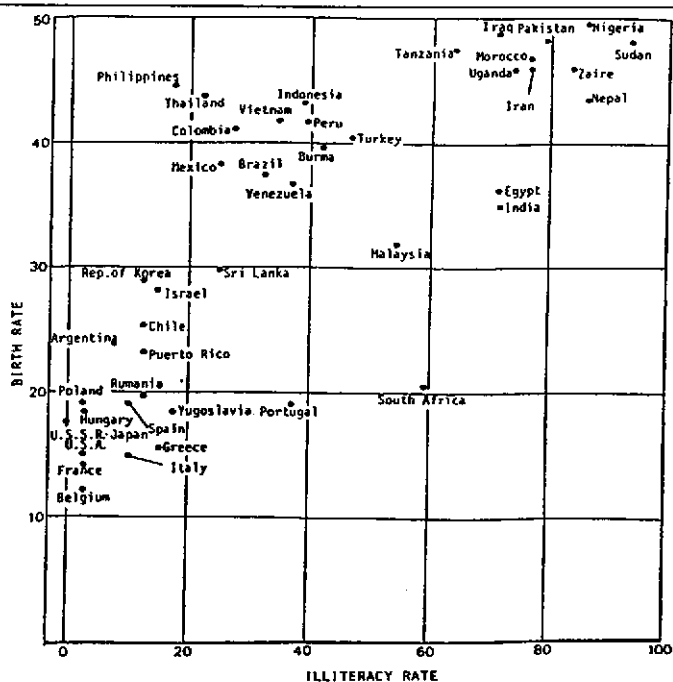


FIG. 7 RELATIONSHIP BETWEEN BIRTH RATE AND ILLITERACY RATE. (  $N = 45$ ,  $r = 0.788$  )

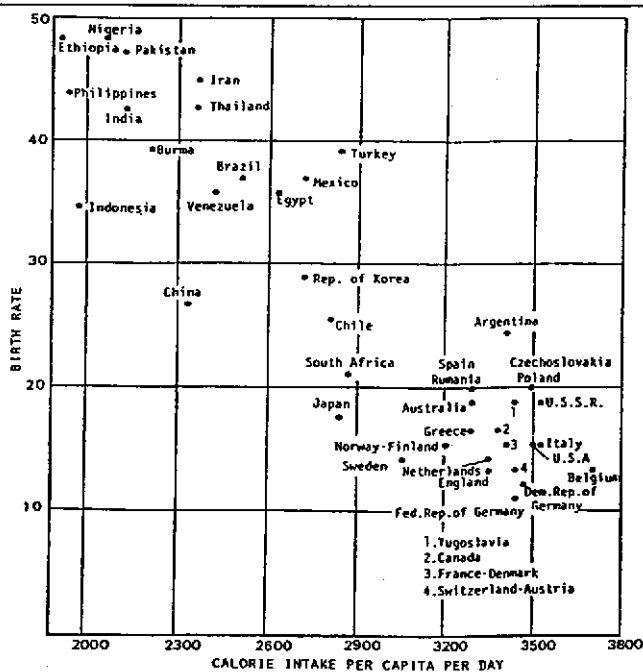


FIG. 8 RELATIONSHIP BETWEEN BIRTH RATE AND CALORIE INTAKE PER CAPITA PER DAY. (  $N = 43$ ,  $r = 0.902$  )

Figure 9

Cluster Analysis Based on the 1970's Data  
( 50 nations, 53 items )

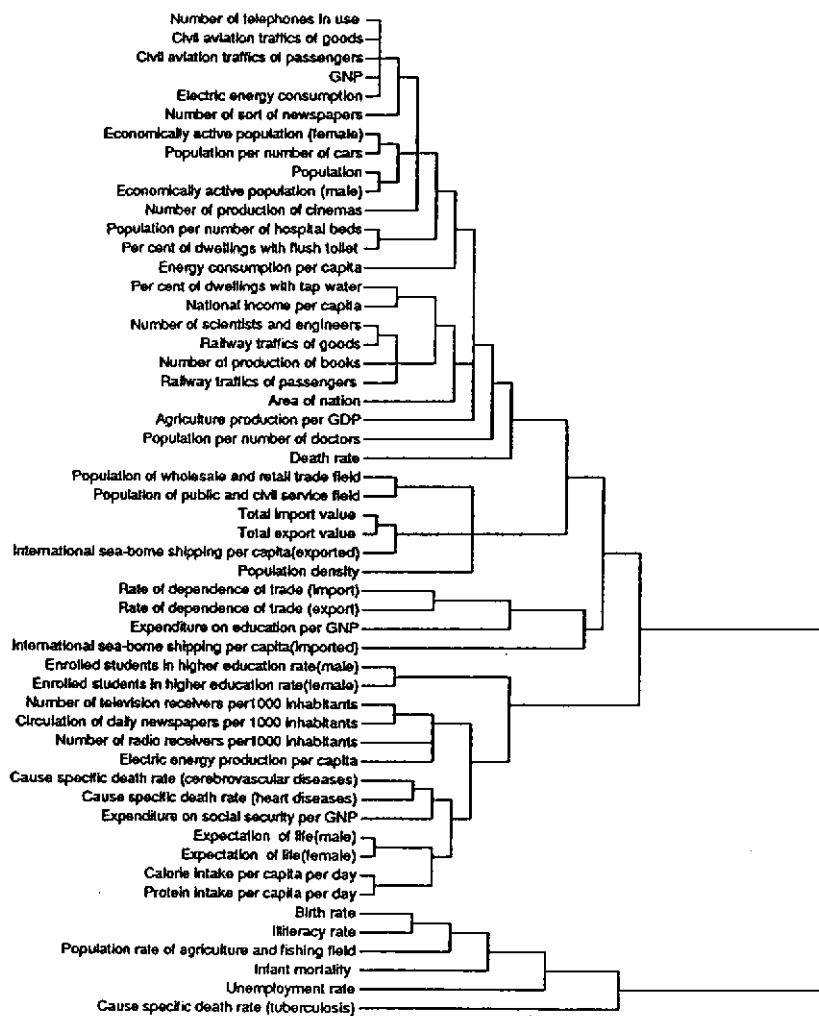


Figure 10

Cluster Analysis Based on the 1980's Data  
( 50 nations, 53 items )

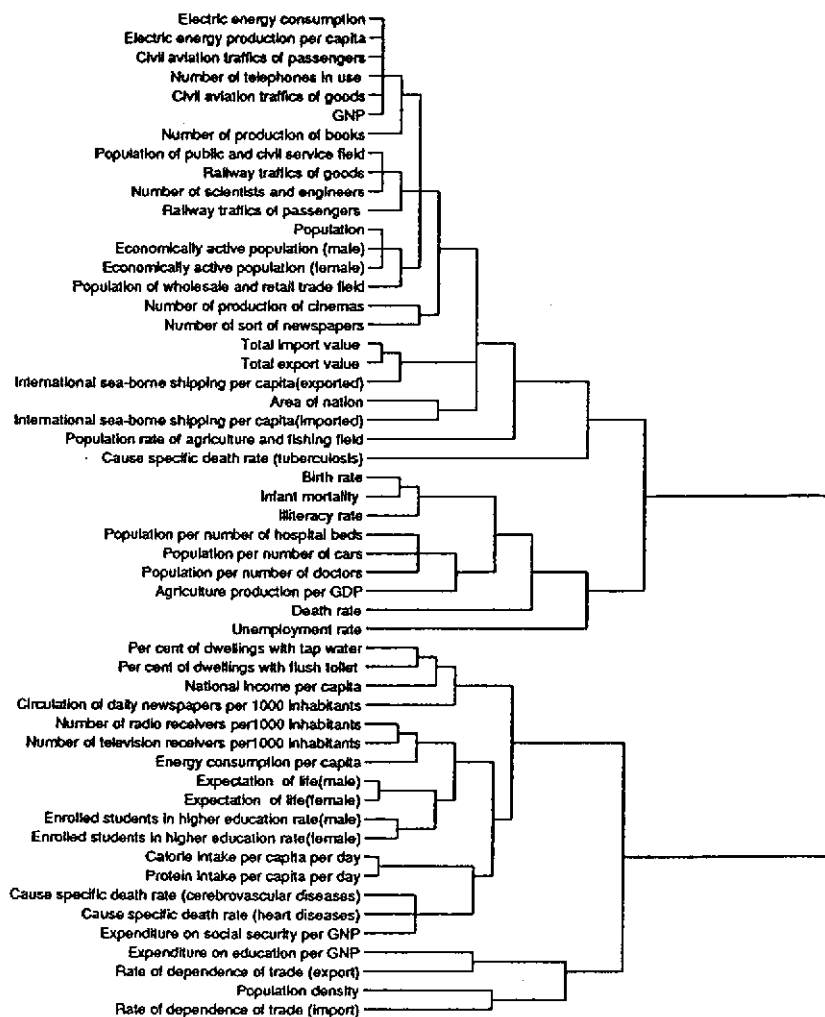


Figure 11

Cluster Analysis Based on the 1990's Data  
( 50 nations, 53 items )

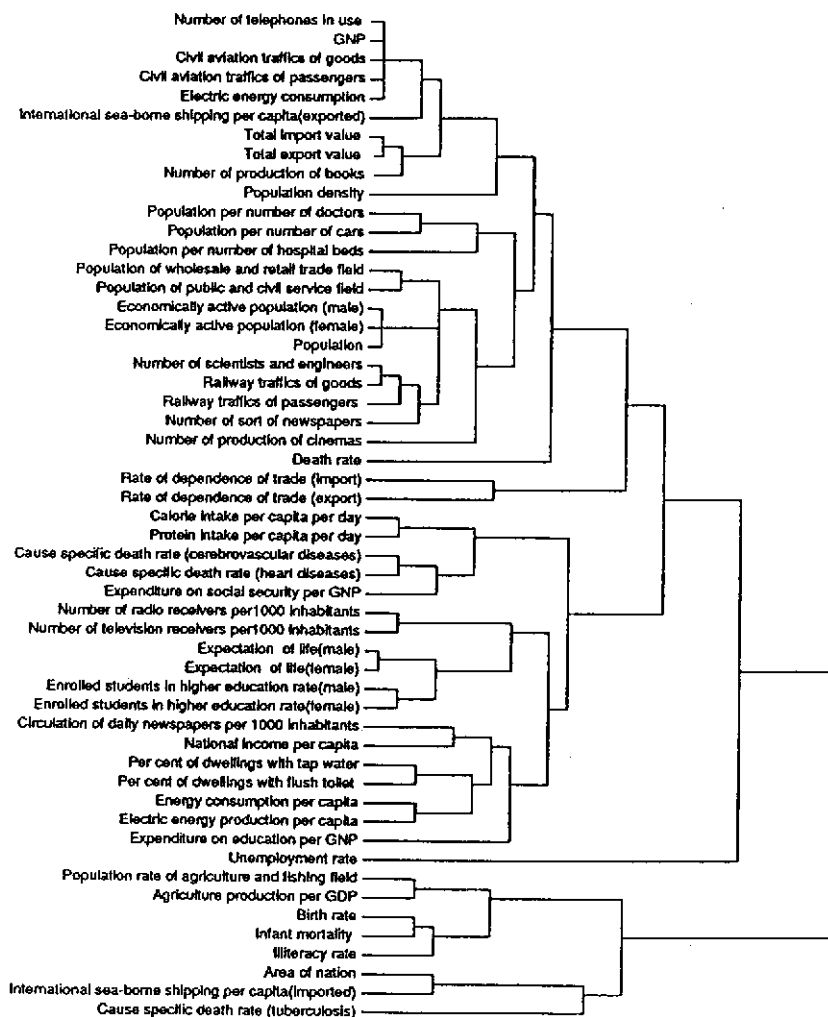


Figure 12

Cluster Analysis Based on the 1970's Data  
( 20 advanced nations, 15 items )

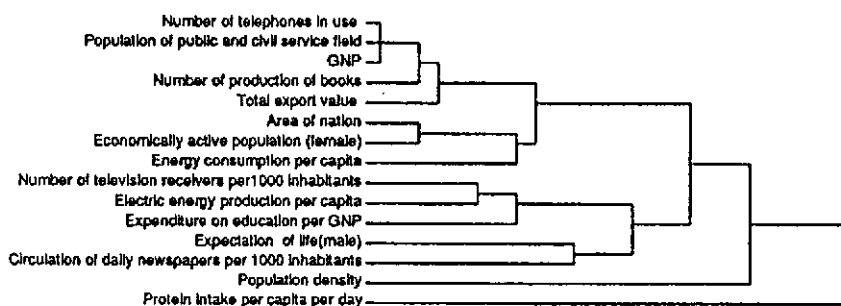


Figure 13

Cluster Analysis Based on the 1980's Data  
( 20 advanced nations, 15 items )

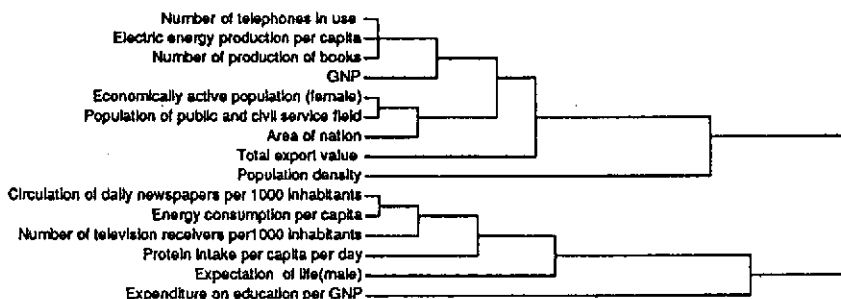


Figure 14

Cluster Analysis Based on the 1990's Data  
( 20 advanced nations, 15 items )

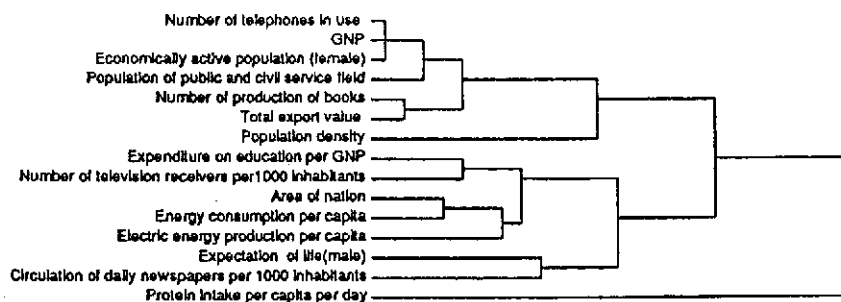


Figure 15

Cluster Analysis Based on the 1970's Data

( 20 developing nations, 15 items )

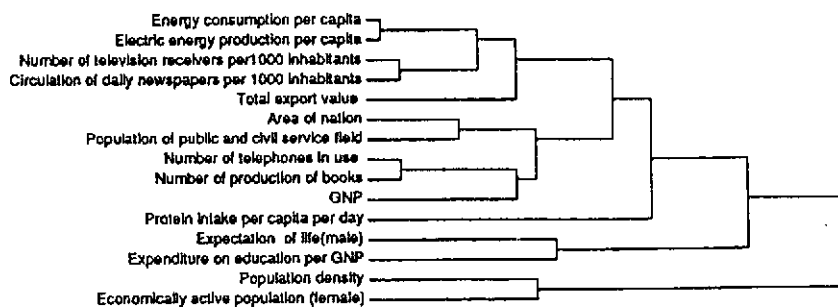


Figure 16

Cluster Analysis Based on the 1980's Data

( 20 developing nations, 15 items )

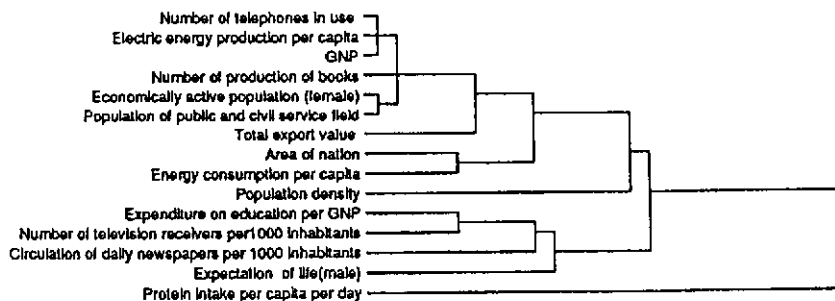


Figure 17

Cluster Analysis Based on the 1990's Data

( 20 developing nations, 15 items )

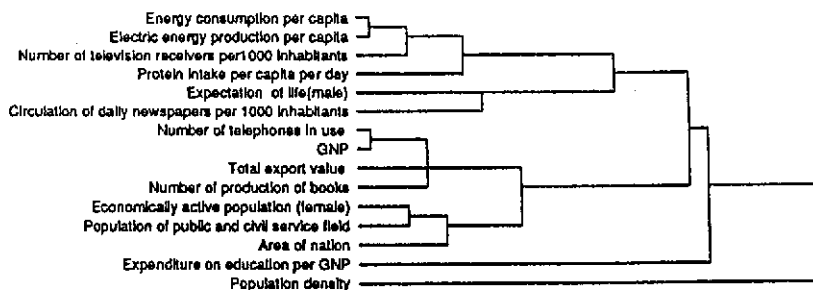
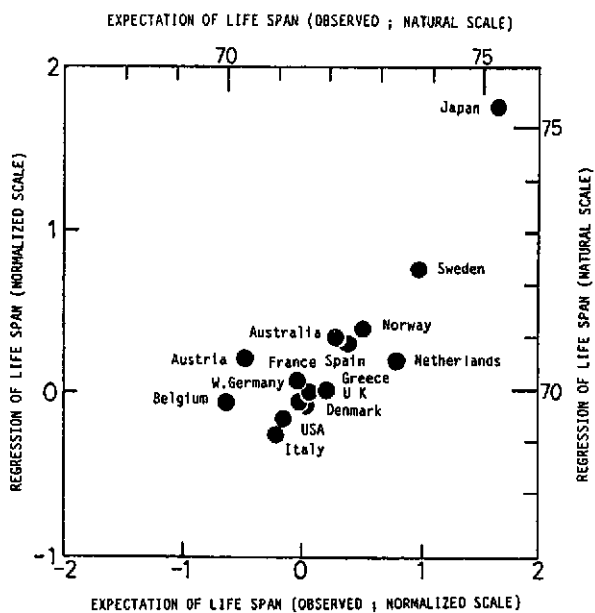
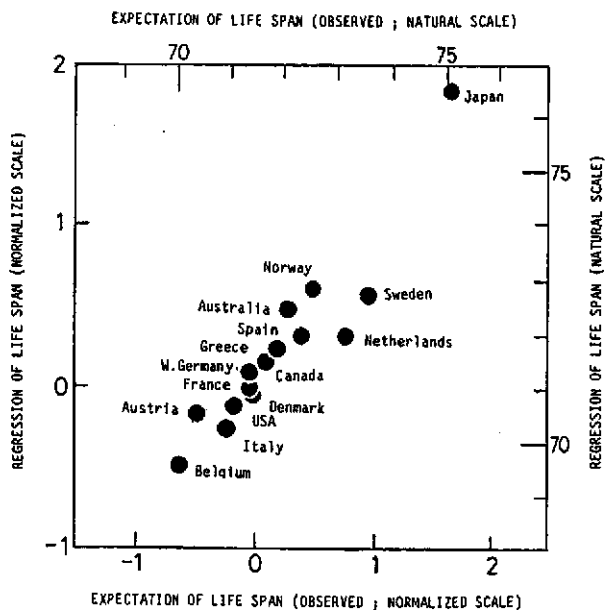


Figure 18



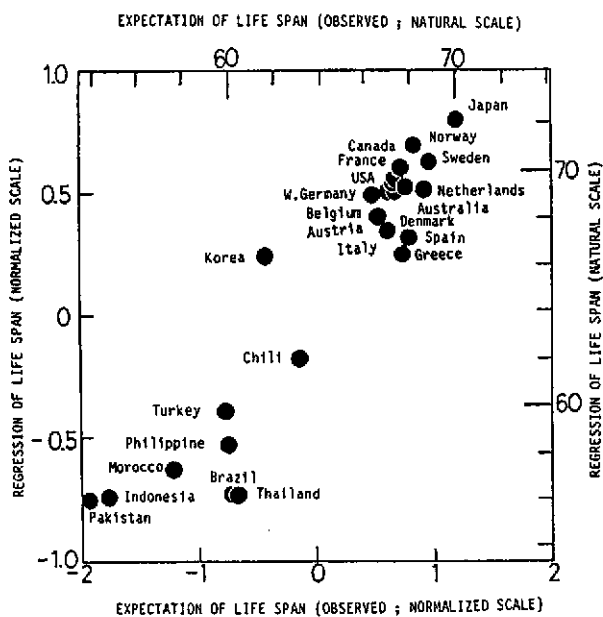
MULTIPLE REGRESSION ANALYSIS OF EXPECTATION OF LIFE SPAN  
(9 SOCIAL INDICATORS, 15 ADVANCED COUNTRIES)

Figure 19



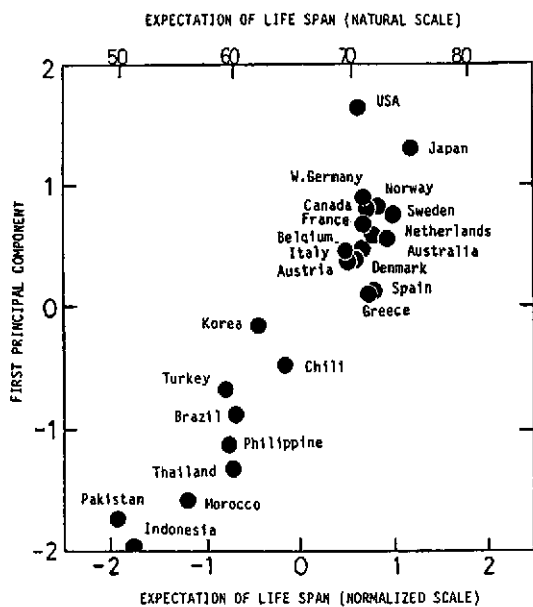
MULTIPLE REGRESSION ANALYSIS OF EXPECTATION OF LIFE SPAN  
(8 SOCIAL INDICATORS, 15 ADVANCED COUNTRIES)

Figure 20



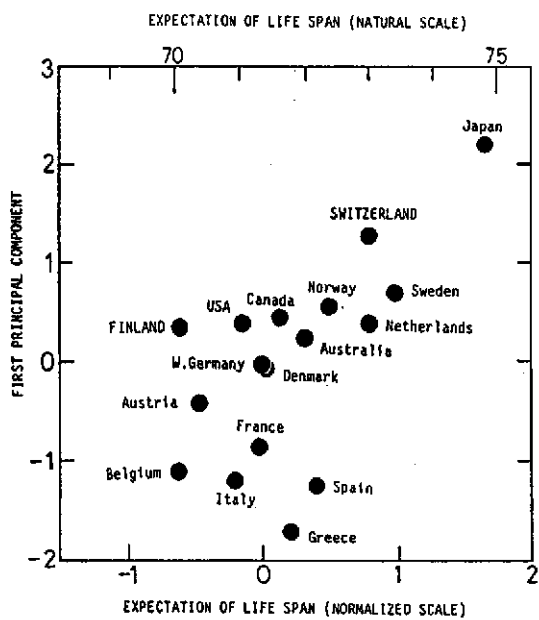
MULTIPLE REGRESSION ANALYSIS OF EXPECTATION OF LIFE SPAN  
(9 SOCIAL INDICATORS, 25 ADVANCED AND DEVELOPING COUNTRIES)

Figure 21



RELATION OF FIRST PRINCIPAL COMPONENT TO EXPECTATION OF LIFE SPAN  
(ADVANCED AND DEVELOPING COUNTRIES)

Figure 22



RELATION OF FIRST PRINCIPAL COMPONENT TO EXPECTATION OF LIFE SPAN  
(ADVANCED COUNTRIES)

# Social Differentials in Health Within and Between Populations

---

*M. G. Marmot*

## INTRODUCTION

**L**ife expectancy is improving in most countries. Between 1965 and 1990, life expectancy for men improved by 6 years (68 to 74) and for women by 6 years (74 to 80) in the rich OECD countries. There were even more marked improvements in much of the rest of the world: by 8 years for men (from 41 to 49), and by 9 years for women (43 to 52) in sub-Saharan Africa; by 14 years (46 to 60) for men and 14 years (44 to 58) for women in India. It is a big jump to go from life expectancy to health status, but if life expectancy is improving, it is a fair guess that there are attendant improvements in health status.

Why are things getting better? This is a question of much interest to demographers but, if one goes by the content of papers in the major medical journals, of little interest to doctors, with a few notable exceptions<sup>1</sup>. Presumably, we would all have ready answers to why life expectancy improved in the poor countries of the world: public health, medical care, food, education, economic development. (These answers, though obvious, may not be correct.) The situation is more complex when we ask why life expectancy improved in the richer countries of the world.

One obvious way things have improved is that countries have got richer. If we believe that there is a link between wealth and health, there is a ready explanation for the improvement in health. There are two problems with this rather simple approach. We do not understand very well the reasons for the link between wealth and health. Second, Wilkinson has shown that, taking countries as the unit of analysis, there is a close relation between GNP/capita and life expectancy up to a ceiling of about \$5,000. Beyond that, there is little relation between a country's GNP and the life expectancy of its inhabitants<sup>2</sup>. Wealth is important but it does not provide a complete enough explanation of why life expectancy has, overall, improved.

One way of approaching the question is to ask why improvements in life expectancy have not been uniform. This has two aspects: among industrialised countries, there is marked divergence in life expectancy; and within industrialised countries there are substantial socio-economic variations in life expectancy and other measures of health status. Exploring reasons for these within-country differences in health is the purpose of this paper, but it is useful to put this into perspective by comparisons among countries.

## Unequal improvements

Amid the general picture of improvement in health, there are some significant discrepancies. The first is among countries. Figure 1 shows life expectancy in 1965 and 1990 for a range of countries. During this period Japan leapfrogged its way to the top of the life expectancy range internationally<sup>3</sup>. Starting from a similar position to the UK (68 years), it had an improvement of 8 years compared to 5 in the UK. Another marked contrast is between Singapore and the former Soviet Union. The Soviet Union/CIS declined from 66 to 65 years, while Singapore increased from 64 to 71. Hungary from a position similar to other European

countries failed to see improvement from its 67 years. Similarly, Czechoslovakia which had similar life expectancy and mortality patterns to neighbouring Austria and Germany, failed to experience the improvement enjoyed by these other European countries from the 1970s<sup>4</sup>.

We examined a range of possible explanations for Japan's improvement<sup>5</sup>. Their dietary fat intake was increasing, with consequent rises in plasma cholesterol, they have a high prevalence of smoking and spend little on medical care. We concluded that their health success was tied up with their economic success. This might seem to be playing a role in Singapore's improvement and the lack of improvement in the countries of Central and Eastern Europe. As will be shown below, these differences may relate more to socio-economic differentials than to overall prosperity.

The second major departure from a picture of overall improvement is the divergence in mortality among socio-economic groups within countries. Figure 2 compares mortality rates in the USA in 1960 and 1986 according to level of education<sup>6</sup>. Although mortality rates declined for men and women regardless of education, the decline was more rapid in those of higher education. The gap between those of higher and lower education, already substantial in 1960, increased by 1986. A similar picture of widening gap between manual and non-manual occupations was reported for England & Wales<sup>6</sup>.

This picture of a widening gap in mortality between socio-economic groups at a time of overall improvement has attracted criticism<sup>7,8</sup>. It has been argued that, in the UK, the apparent higher relative mortality of those in lower social classes applies to a dwindling portion of the population. The US analyses allowed for this by comparing relative mortality for constant percentiles of the distribution in the two time periods. The disparity in health status between socio-economic groups is not disappearing as overall health status of the population improves.

### **Health, Health Care, Who Cares?**

It may appear as a contradiction to assert that the medical profession cares little about the health status of the population, when health is so much in the news. Health in the news largely takes two forms: titillating snippets that supply a dedicated readership with a wide range of conflicting advice on the lifestyle route to eternal life; and the funding and organisation of health care. The first of these, the latest newsworthy lifestyle advice, may have some impact on the health of a small minority of interested people; but neither it nor the second is likely to have much impact on the health status of the population as a whole. McKeown, one of the few to be interested in the determinants of the improvement in health over many decades, made a persuasive case that improvements came from nutrition and environmental change not from advances in medical care<sup>9</sup>.

Yet health care reform is on the political and public agenda in the USA because that nation spends about one seventh of its national income on medical care. Keeping costs under control will help individuals and companies who have to foot the bill; and extending coverage to those presently uninsured will bring the USA more in line with other advanced countries. The evidence suggests that important as it is to extend medical care coverage to the whole population, this is unlikely materially to influence socio-economic differences in life expectancy<sup>10</sup>.

Similarly, it will be difficult to make the case that health care (strictly medical care) reform will have much impact on the health of Americans as a whole. If we again take life expectancy as a measure of health status of a country, the USA, despite leading the world in medical care expenditure, trails rather badly. In 1991, life expectancy for men in the USA was 72 years which, along with the UK, Denmark and Ireland, was equal to the worst of the 22 rich (OECD) countries<sup>11</sup>. This pales by contrast with Japan (76 years), Sweden and Hong

Kong (75 years), and Norway and Switzerland (74 years). For women, the USA ranked 14 out of 22.

The UK spends relatively little of its national income on medical care, about 6% of GNP, but the debate similarly focuses on the funding and organisation of medical care rather than on the health status of the population.

It should be no surprise that the medical profession is concerned principally with the diagnosis and treatment of disease. It is what we are trained to do. There is a commitment to investigating the mechanisms of disease, with a view to improving treatment and targeting it at those at special risk. Related to this approach, advances in biological understanding of disease have been truly outstanding. Whether these will translate into major changes in the population burden of ill-health remains to be seen. To the extent that prevention has been discovered, by and large it follows the medical orientation to altering the risk status of individuals at special risk. As Rose has shown, unless an approach is taken to changing risk status of whole populations, the impact on public health will be limited<sup>12</sup>.

In many countries, there has been a recent welcome change from this purely medical approach. Governments have set health targets for the population's health. In the UK, the Health of the Nation chose five key areas and set targets for the year 2000 and beyond<sup>13</sup>. In laying out a strategy for how these targets might be achieved, the National Health Service was seen as central but there was explicit recognition that there are determinants of health that lie outside the health care system. Other sectors are important. The contradiction is that health is high on the agenda only of the Department of Health. Other areas of public policy that potentially have a profound influence on health do not usually have health as a policy goal. There is, for example, little explicit recognition that economic and social policies may have a profound influence on the health of the population.

One reason for the interest in socio-economic differences in health is the insight it provides into social and economic influences on health. Social variations in health status were acknowledged in the Health of the Nation, but the conclusion was that the knowledge base was insufficient to recommend specific actions to deal with them.

It is important therefore to understand the reasons for socio-economic differences in health. Taking actions that reduce these differences is not only important as a social goal, but without reductions in socio-economic differences the ability to meet health targets for the whole of society will be reduced.

## **SOCIO-ECONOMIC DIFFERENCES WITHIN COUNTRIES: THE WHITEHALL STUDIES**

There have now been a large number of studies showing a relation between various measures of social position and morbidity and mortality<sup>14</sup>. The task is to go from description to explanation. That is the task on which we are embarked in the Whitehall studies of British Civil Servants<sup>15</sup>. Studies based on vital statistical data have a limited range of information. In the two Whitehall studies, we have collected a range of information that allow us to test hypotheses as to reasons for persisting socio-economic differences in morbidity and mortality.

In the original Whitehall study, we classified men by grade of employment as an index of social position and they were followed for mortality. The results were as in Figure 3, updated from previous reports<sup>16</sup>.

The most striking finding from this figure is the social gradient in death rates. There is a step-wise relation between grade of employment and mortality. The Whitehall Study consists of a group of people, of one predominant ethnic group, all employed in stable office based jobs, not subject to industrial hazards, unemployment, or extremes of poverty or affluence; all

working and living in Greater London and adjoining areas. Yet in this relatively homogeneous population, we observed a gradient in mortality - each group with a higher mortality than the one a step higher in the hierarchy. The difference in mortality was threefold between highest and lowest grades. The task for explanation changes. The question is not why people at the bottom have worse health but why social differentials in health are spread across the whole of society.

As repeated above, health is not only a matter of life or death. In a second study of Civil Servants, the Whitehall II study, we have been approaching health as a combination of social, psychological and physical functioning. As one way of operationalising this, we have been examining the determinants of sickness absence rates i.e. absence from work due to sickness. Figure 4 shows sickness absence rates by grade of employment, with employment grade now measured as 6 grades from the highest administrators, Unified Grades 1 through 6, down to the lowest, clerical officers and office support staff. For both men and women, there is a dramatic gradient, rising sharply from the low rates among senior administrators to the high rates among lower grade office staff. Once again, the interesting contrast is not only between those at the top and bottom, but there is a continuous inverse relation.

These sickness absence data should stimulate our thinking. The relation between social position and sickness absence rates holds for short spells of absence of 7 days or less as it does for long spells that require a medical certificate. If all illness is a mix of biological, psychological and social, it might be that short spells of absence reflect more of the psychological and social, and long spells more of the biological aspects of illness. Even if it were the case that short spells reflect psychological and social functioning, not "true" illness, it is interesting that the social gradient in absence rates is identical for short and long spells. One interpretation of this is that the factors responsible for the social gradient in social and psychological functioning are similar to those responsible for the social gradient in "true" illness.

### Poverty or inequality

Much of the discussion about social inequalities in health has related to the health disadvantage of those at the bottom<sup>17</sup>. This is analogous to seeing social problems as a particular problem for a disadvantaged minority. There is little doubt that poverty, or deprivation, is likely to be bad: for health among other things. A dramatic example of this comes from New York<sup>18</sup>. Young blacks in Harlem have 6 times the mortality of the New York average. Mean income of blacks was less than one third the US average. The life expectancy of black men in Harlem was below that of men in Bangladesh.

The Whitehall data on mortality (Figure 3) and Whitehall II data on sickness absence (Figure 4) suggest that something other than absolute poverty is at work here. Each grade has worse health and higher mortality rates than the grade above it. Executive grade civil servants are not poor by any absolute standard yet they have higher mortality rates than administrators. Even clerical officers who are far from well-off, with earnings at or below the national average, are not poor by comparison with England at an earlier period in history, or with developing countries.

This social gradient in mortality suggests the operation of factors across the whole of society. Whether it is relative deprivation or relative lack of access to the fruits of a wealthy society, it is clear that explanations for socio-economic differentials in Britain in the 1990s must be broader than the notion of poverty advanced earlier in the century<sup>17</sup>.

This relates to the question with which we began: why has life expectancy improved and how does that relate to the general improvement in living standards? Above a certain level of material well-being, it is not easy to see how improvements would lead to further improvements in health. If shelter, clothing, nutrition are adequate, do further improvements lead to

further improvements in health? Is a colour television better for health than a black and white one? A holiday in Spain better than one on the English sea-side? The ability to buy fast food better than home cooking?

Wilkinson provides an answer<sup>2</sup>. He shows that, comparing countries, there is a relation between gross national product per head and life expectancy at birth only for poor countries: in 1984, below a threshold of gross national product per head of about \$5000, few countries had a life expectancy of 70 years or more. Beyond that level however, there is little relation between gross national product and life expectancy. The relation is much closer with measures of income dispersion. Using as a measure of equality of income distribution the share of total post tax household income received by the least well off 70% of families, there is a striking correlation with life expectancy. Wilkinson tried several measures of income distribution, the share of total income received by the least well off 10%, 20%, 30% etc. He found that the correlation with life expectancy increased progressively until the bottom 60-70% was reached i.e. the higher the share of total income enjoyed by the bottom 60 or 70%, the longer the average life expectancy. He also showed that changes in life expectancy correlated with changes in income distribution. Among 6 countries for which these data were available, Japan had the greatest increase in equality of income distribution and the greatest increase in life expectancy. Japan now has the most equitable distribution of income and the longest life expectancy of any OECD country. By contrast in the United Kingdom, the bottom 60 % of households had a declining share of total income and the population had a relatively small increase in life expectancy.

How are we to explain the relation between income inequality and life expectancy? One possibility is that greater income inequality implies a greater proportion of the population in poverty. If there were a non-linear relation between income and life expectancy - stronger at lower incomes than at high - the apparent relation between income inequality and life expectancy could possibly be explained by the differing proportions of those in poverty. Wilkinson points out that this is not the most likely explanation<sup>3</sup>. If the criterion of income inequality is taken as the share of income received by the bottom 10%, the relation with life expectancy is very much weaker than the criterion of the share received by the bottom 60 or 70%. Even if increased income inequalities did result in more households in poverty, the proportion of absolute poor in OECD countries is still too small to account for the size of the relation between income inequality and life expectancy. There must be an effect of income inequality on the large bulk of the population. In other words, the main influence on life expectancy among rich countries is likely to be of relative rather than absolute deprivation.

**The task for explanation then becomes how relative position in the social hierarchy is related to mortality, life expectancy and other measures of health.** We are dealing not with absolute poverty but with inequality.

### **One disease or many?**

A striking feature of social class differences in health is the generalisability of the findings across diverse pathological conditions. In the Whitehall study, the higher risk of death among lower grades applied to deaths from lung cancer, other cancers, CHD, cerebrovascular disease, other cardiovascular disease, chronic bronchitis, other respiratory disease, gastro-intestinal disease, genito-urinary disease, accidents and violence<sup>4</sup>. Findings such as these, suggesting that there may be common factors operating that cut across known causes of disease, have prompted the speculation that there may be factors that increase general susceptibility to ill-health<sup>19,20</sup>.

An alternative to a general susceptibility hypothesis, is that a variety of specific factors are operating to explain social class differences in mortality. Support for this view comes from

the fact that some cancers, notably those of the colon, brain, prostate, haematopoietic system, breast and melanoma, do not show the same social class variation as the causes listed above<sup>21</sup>.

A general susceptibility hypothesis implies that certain groups will be at higher risk of death whatever causes are operating. It does not deny the operation of specific causes. Diseases linked to smoking, such as chronic bronchitis and lung cancer, show a particularly strong social class gradient - stronger than cancers not linked to smoking. But the latter do show a social class gradient as do other diseases not linked to smoking. Put another way, the general susceptibility hypothesis means that there are factors operating that cut across our current system of classifying diseases. These will increase risk of death in addition to the effect of known factors such as smoking. This can account for the fact that an administrator who smokes 20 a day has a lower risk of lung cancer mortality than a lower grade civil servant smoking the same amount<sup>26</sup>, even after pack-years and tar content of cigarettes is taken into account; and for the gradient in mortality that occurs for CHD even among non-smokers<sup>22</sup>.

The production of this apparently increased susceptibility may well be operating at a social level. The whole life-course of people in different social locations is different, and insults to health may accumulate over the entire period from birth to death. That these influences on health cluster in such a way as to produce social groups at differing degrees of disadvantage with respect to most diseases is undeniable. Our current level of knowledge regarding this general susceptibility allows us to go little beyond this empirical observation, however.

## SOCIAL AND OTHER CAUSES OF DISEASE: CHD

Coronary heart disease serves well to illustrate that a biological and a social model of the causation of disease must be reconciled. There are two related challenges to the search for causes of the inverse association between socio-economic position and risk of CHD and other diseases: (i) the social distribution of CHD has changed - the evidence suggests that in the twentieth century in Britain and elsewhere, CHD changed<sup>23</sup> from being more common in higher to more common in lower socio-economic groups; (ii) much is known about the causes of CHD that relate to lifestyle of individuals, quite apart from their genetic predisposition.

Figure 5, using data from the Whitehall study shows one approach, that ignores social causation, to explaining why CHD is of epidemic proportions in most industrialised countries. We can define a low risk subgroup of civil servants as those in the bottom two quintiles of the population distribution of plasma cholesterol level, and in the bottom two quintiles of the blood pressure distribution, and who had never smoked. The mortality rates from CHD and other causes in this low risk group can then be compared with the mortality in the bulk of the population at "average" risk. The figure is a guide to what is potentially feasible biologically. If the association between risk factors and disease is causal, and the evidence suggests that it is, and if it were possible to have a British population in which the "average" level of risk was reduced to that in this currently defined low risk group, then CHD rates would be more than 60% lower than the present average. Other major causes of death would also be lower.

If more than 2/3 of all coronary deaths can be attributed to these three risk factors, how can we argue that there are important social determinants? First, Figure 5 is a guide to what is biologically feasible, not necessarily achievable practically. To move the mean level of plasma cholesterol or blood pressure to that of the mean of the bottom two quintiles, or to move the whole population into the "never smoker" category is not a short term feasible option. Second, given that the British population has uniformly high plasma cholesterol levels, this is not the reason for the marked social gradient in CHD. In Whitehall, mean plasma cholesterol

showed a small positive association with grade i.e. higher levels in higher grades; there was a small inverse association with blood pressure and although the differences in smoking were marked, the social gradient in CHD mortality among non-smokers was the same as among smokers. Figure 6 shows that adjusting for coronary risk factors explained about 25% of the social gradient in CHD. Another way of expressing this is that without a high level of risk factors, CHD would not be a major cause of death in Britain and among these civil servants. Given that the level of risk factors is high, something must account for the fact that for a given level of "exposure" to risk factors, something increases the susceptibility to CHD in a step-wise fashion, increasing with lower social status. It is for the cause of this increased susceptibility that we must search. The data showing a number of causes of morbidity and mortality have a similar social gradient suggests that some of the factors causing susceptibility may be common to a number of different manifestations of ill-health.

## EXPLANATIONS OF SOCIAL INEQUALITIES IN HEALTH

Recently, my colleagues and I reviewed possible explanations for social inequalities in health<sup>10</sup>. That review will be summarised here.

Although social differences in the provision of medical care loom large in the USA, they are less likely as an explanation in the UK because of universal access under the National Health Service. Further the improvement in mortality over time, greater for higher than for lower social classes, is the result of the decline in causes of death judged to be non-amenable to medical care, and the lack of decline in lower classes.

Analogous to the limited role of differences in medical care in generating social inequalities in health, is its limited role in generating international differences. When comparing Japan to England & Wales, for example, we noted that Japan like the UK spends a relatively small proportion of GNP on medical care. The decline in mortality in Japan was observed both for amenable and non-amenable causes of death<sup>1</sup>.

Access to good quality medical care is a right that should be enjoyed equally by all members of society. It is hard to make the case that it is differential access to or provision of medical care that is responsible for inequalities in health in European countries.

The evidence does not support health selection as an explanation i.e. that health may determine social position rather than vice versa. This was one of the possible explanations considered by the Black report and rejected as a major cause of social inequalities in health<sup>24</sup>.

### Factors operating early in life

A different version of the selection hypothesis suggests that while social selection based on health status is not a crucial contributor to health differentials, common background factors determine both social position and health in adulthood. This process has been termed 'indirect selection'<sup>25</sup>, and recognises that people bring with them into adulthood the results of influences from their earlier days: genetic factors, biological results of early experiences, educational, cultural, psychological and social factors. Since it has been argued that the major influences on health in adulthood operate early in life<sup>26</sup>, it is possible that both social position and health in adulthood are determined by common early life influences. The focus on the childhood origins of adult disease has been criticised precisely because influences from early life shape the lives people lead and the social environments in which they live and work<sup>27</sup>. It may be these conditions of adult life that are related to ill-health, and the importance of childhood conditions may therefore be indirect. It is clearly not easy to separate the direct effects on health of early and later life experiences.

Whitehall data relating height to mortality provide some insight into the possible separate effects of current and past environment. Height is influenced by environment, as well as by genes. Short height predicts adult mortality independent of grade of employment<sup>6</sup>, and it is reasonable to speculate that this may in part be a reflection of a persisting influence from early life. Grade of employment, which is to some extent an index of current social influences, predicts mortality independent of height. Thus two sets of influences may affect mortality risk, factors from early life and current influences.

Individuals may bring more than biological characteristics from early life to adult life. Education, in most studies is a powerful predictor of health status. This may be because education is a potent indicator of current adult social position. It may also be telling us something of the causal pathways involved. People with more education may have more life management skills. It may be these life management skills, carried forward into adulthood, that predict both adult socio-economic circumstances and health status.

### **Health related behaviours and biological risk factors**

As stated above, smoking was the only major established coronary risk factor that explained any substantial part of the social gradient in mortality in Whitehall. To the extent that lifestyle patterns do mediate the relation between socio-economic position and health, the question changes. Why are there links between social position and lifestyles such as smoking?

There must, of course, be biological pathways underlying the link between social position and disease rates. These biological changes may, in part, result from behaviours such as diet and smoking but may also be the result of other influences, stress for example, acting directly on the neuroendocrine system.

### **Material conditions**

The Black Report<sup>24</sup> emphasised the importance of material conditions as an explanation for social inequalities in health. In fact, Black referred to materialist or structural explanations: emphasising hazards to which some people have no choice but to be exposed given the present distribution of income and opportunity. This can be interpreted as broader than simply material conditions, and include psychosocial influences that are inherent in position in society.

The difficulty in understanding material explanations is to know how they operate. At a time when poor living conditions meant polluted water, crowded unsanitary housing with high rates of cross infection, and appalling conditions of employment, it was not difficult to see how these could be responsible for worse health among the socially deprived. This would be additional to the effects of inadequate diet which is part of the material conditions of life. As conditions improved, mortality of all social groups improved. But why do the social gradients persist? Are we to understand that there are residual effects of bad housing with damp and infection, as well as air pollution and other material conditions that, although they affect the lower social groups less than they used to, still affect them more than the higher social strata? If there are such residual effects it is no surprise that they do not affect all social strata equally. Can this be the whole explanation?

The Whitehall data and Wilkinson's data on income inequalities are relevant here. In the Whitehall and Whitehall II studies, morbidity and mortality varied linearly with grade of employment. It is possible that the worse health of the second highest grade compared with the highest could be the result of worse housing, poorer diet for children, or greater pollution, but it seems unlikely that comfortable "middle class" people in Britain are suffering from the

effects of material deprivation. Similarly, comparing rich countries, Wilkinson showed that it was not differences in wealth that predicted differences in life expectancy, but differences in inequality of income distribution.

One is drawn to the view that, in addition to the multiple influences already discussed, there must be an influence of relative position in society. What may be important is not absolute deprivation but relative deprivation. This would account for a social gradient in ill health, because each group, while not necessarily suffering from greater effects of bad housing etc., will have "less" than the group above it. This would account for the widespread finding of social inequalities in health in societies with very different levels of health. The social gradient in ill-health will vary in magnitude depending on the magnitude of the relative differences in "deprivation".

In a society that has met the subsistence needs of its members, what do we mean by "less" or "relative deprivation"? In addition to the factors explored above, we should look to psycho-social factors for part of the answer.

### **Psycho-social factors as potential explanations**

If relative deprivation is to be operationalised, what does it mean? How could a lower position in the social hierarchy cause heart disease and cancer, suicide and gastrointestinal disease, if not by diet or smoking?

Whitehall II data show that lower position in the occupational hierarchy is associated with lower participation in social networks and activities and less social supports; with adverse psychosocial work environment: less sense of control over the work, lower use of skills, and less variety; less sense of personal control over health - external health locus of control; more financial difficulties. These are all related to mortality and morbidity in other studies<sup>28-31</sup>. Syme has suggested that increasing lack of control may be a linking factor that accounts for increasing health disadvantage as the social scale is descended<sup>32</sup>. Relative poverty, whatever else it represents is relative lack of control.

We are now beginning our exploration of the extent to which these can account for social gradients in health. These can be illustrated with our data on sickness absence. Figure 4 showed that there was a steep gradient in sickness absence rates according to employment grade. Three sets of characteristics were related both to grade of employment and to risk of sickness absence: personal health behaviour, such as smoking, and self rated health; psychosocial work characteristics: lack of control, low variety, low use of skills, low job satisfaction; social circumstances outside work: negative aspects of social support and difficulty paying bills.

Figure 7 shows sickness absence rates by grade, adjusted for age and fully adjusted for these characteristics, as a way of asking the extent to which these characteristics mediate the grade ill-health relationship. These factors account for about a third of the grade-sickness absence relationship.

At first glance, this is somewhat disappointing. We must first take measurement issues into account. Peto and colleagues showed that measurement imprecision leads to a 50% underestimate of the relation between blood pressure level and subsequent cardiovascular disease<sup>33</sup>. There is a similar underestimate for cholesterol level. Measurement imprecision of these psycho-social measures will therefore underestimate the proportion of social variance explained by at least 50%. Second, we have as yet, put little effort into understanding the interrelation of these psychosocial measures and the different meaning they may have at different points in the hierarchy. For example, we have the impression that having to work fast is a positive characteristic for people in high grades. It goes with the territory. It may represent psychological demands for people in low grades. Third, there may be some crucial vari-

ables that we have left out. If relative position in the hierarchy is important, then control and sense of fairness may be important both for mental and physical health.

The Whitehall studies are changing our approach to the study of the determinants of health. We are using them to study the inter-relation of psycho-social and behavioural measures, and the biological pathways by which they relate to major differences in health status.

## REFERENCES

1. Sagan LA, The Health of Nations. True Causes of Sickness and Well-being, New York: Basic Books, 1987), 1-233.
2. Wilkinson RG, "Income distribution and life expectancy" British Medical Journal **304** 1992, 165-168.
3. Marmot MG, Davey Smith G, "Why are the Japanese living longer?" British Medical Journal **299** 1989, 1547-1551.
4. Bobak M, Feachem RGA, "Health Status in the Czech and Slovak Federal Republic," Health Policy and Planning **7** (3) 1992, 234-242.
5. Pappas G, Queen S, Hadden W, Fisher G, "The increasing disparity in mortality between socio-economic groups in the United States, 1960 and 1986," New England Journal of Medicine **329** 1993, 103-109.
6. Marmot MG, McDowall ME, "Mortality decline and widening social inequalities," Lancet **i** 1986, 274-276.
7. Illsley R, "Occupational class, selection and the production of inequalities in health," Quarterly Journal of Social Affairs **2** 1986, 151-165.
8. Strong PM, "Black on class and mortality: theory, method and history," Journal Public Health Medicine **12** 1990, 168-180.
9. Mckeown T, The role of medicine, (London: Nuffield Provincial Hospitals Trust, 1976)
10. Marmot MG, Bobak M, Davey Smith G, "Explanations for social inequalities in health," Book: Society and Health, (forthcoming).
11. World Bank. World development report 1992, (Oxford: Oxford University Press, 1992).
12. Rose G, The strategy of preventive medicine (Oxford: Oxford University Press, 1992).
13. HMSO, The Health of the Nation (London: HMSO, 1992).
14. Marmot MG, Kogevinas M, Elston MA, "Social/economic status and disease," Annual Review of Public Health **8** 1987, 111-137.
15. Marmot MG, Rose G, Shipley M, Hamilton PJS, "Employment grade and coronary heart disease in British civil servants," Journal Epidemiology and Community Health **32** 1978, 244-249.

16. Marmot MG, Shipley MG, Rose G, "Inequalities in death - specific explanations of a general pattern?" Lancet **1** 1984, 1003.
17. M'Gonigle GCM, Kirby J, Poverty and public health. (London: Golantz, 1937).
18. McCord C, "Excess mortality in Harlem," New England Journal of Medicine **322** 1990, 173-177.
19. Berkman LF, Syme SL, "Social networks, host resistance and mortality: a nine-year follow-up of Alameda County residents," American Journal of Epidemiology **109** 1979, 186-204.
20. Cassel JC, "The contribution of the social environment to host resistance," American Journal of Epidemiology **104** 1976, 107-123.
21. Davey Smith G, Leon D, Shipley MJ, Rose G, "Socioeconomic differentials in cancer among men," International Journal of Epidemiology **20** 1991, 339-345.
22. Davey Smith G, Shipley MJ, "Confounding of occupation and smoking: its magnitude and consequences," Social Science and Medicine **32** 1991, 1297-1300.
23. Marmot MG, Adelstein MM, Robinson N, Rose GA, "Changing social class distribution of heart disease," British Medical Journal **2** 1978, 1109-1112.
24. Black D, Morris JN, Smith C, Townsend P, Whitehead M, Inequalities in health: The Black report; The health divide (London: Penguin Group, 1988).
25. Wilkinson RG, "Occupational class, selection and inequalities in health: a reply to Raymond Illsley," Quarterly Journal of Social Affairs **2** 1986, 415-422.
26. Barker DJP, "The intrauterine and early postnatal origins of cardiovascular disease and chronic bronchitis," Journal of Epidemiology and Community Health **43** 1989, 237-240.
27. Ben-Shlomo Y, Davey Smith G, "Deprivation in infancy or adult life: which is more important for mortality risk?" Lancet **337** 1991, 530-543.
28. Karasek R, Theorell T, Healthy Work: Stress, Productivity, and the Reconstruction of Working Life (New York: Basic Books, 1990).
29. Marmot M, Theorell T, "Social class and cardiovascular disease: the contribution of work," International Journal of Health Services **18** 1988, 659-674.
30. Berkman LF, "Assessing the Physical Health Effects of Social Networks and Social Support," in Breslow L, Fielding JE, Lave LB, eds., Annual Review of Public Health, 5th Edition. (Palo Alto, USA: Annual Reviews Inc, 1984) 413-432.
31. House JS, Landis KR, Umberson D, "Social relationships and health," Science **241** 1988, 540-545.

32. Syme SL, "Control and health: a personal perspective," in Steptoe A, Appels A, eds., Stress, personal control and health (Wiley & Sons, 1989) 3-18.
33. MacMahon S, Peto R, Cutler J, et al., "Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias," Lancet 335 1992, 765-774.

Figure 1

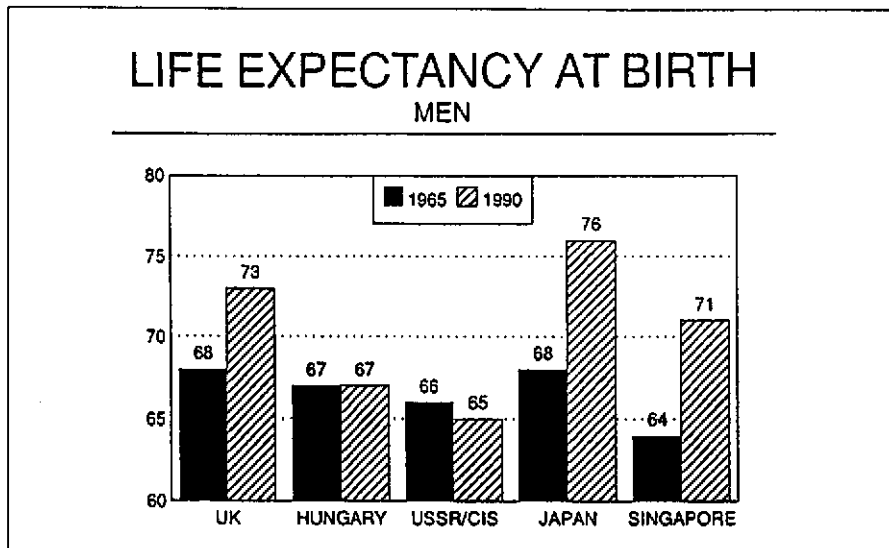


Figure 2

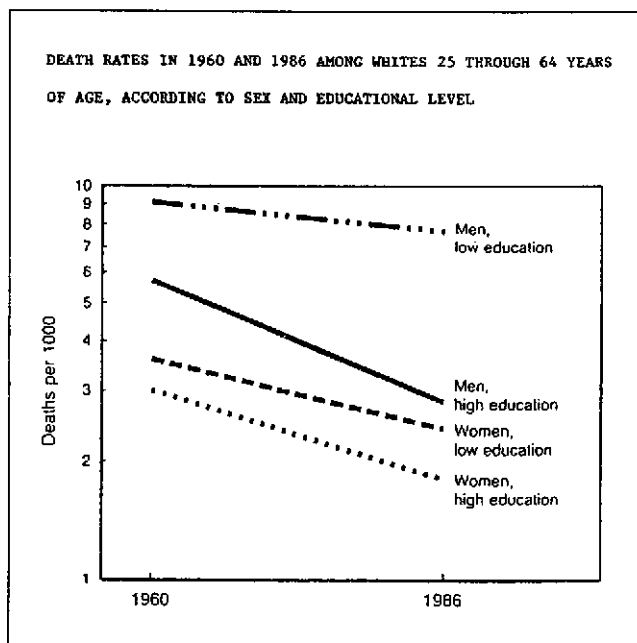


Figure 3

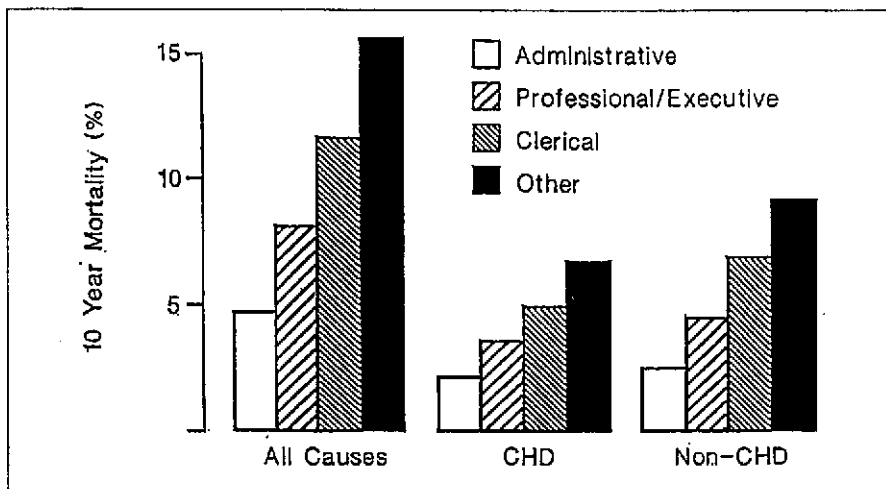


Figure 4

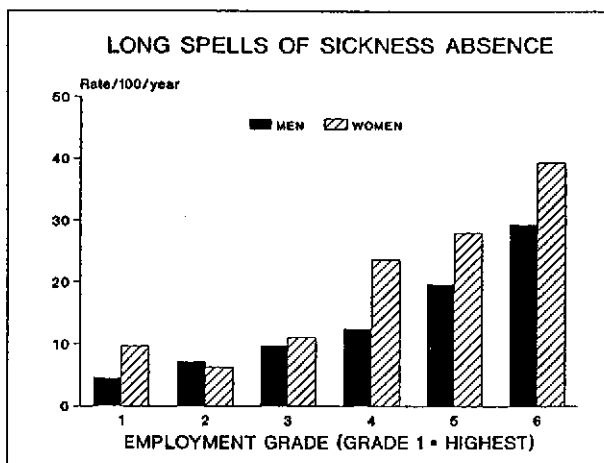


Figure 5

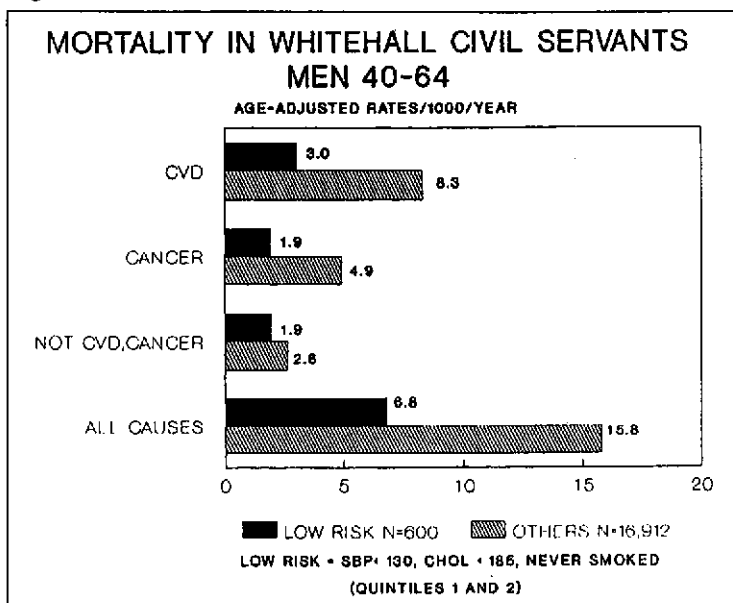


Figure 6

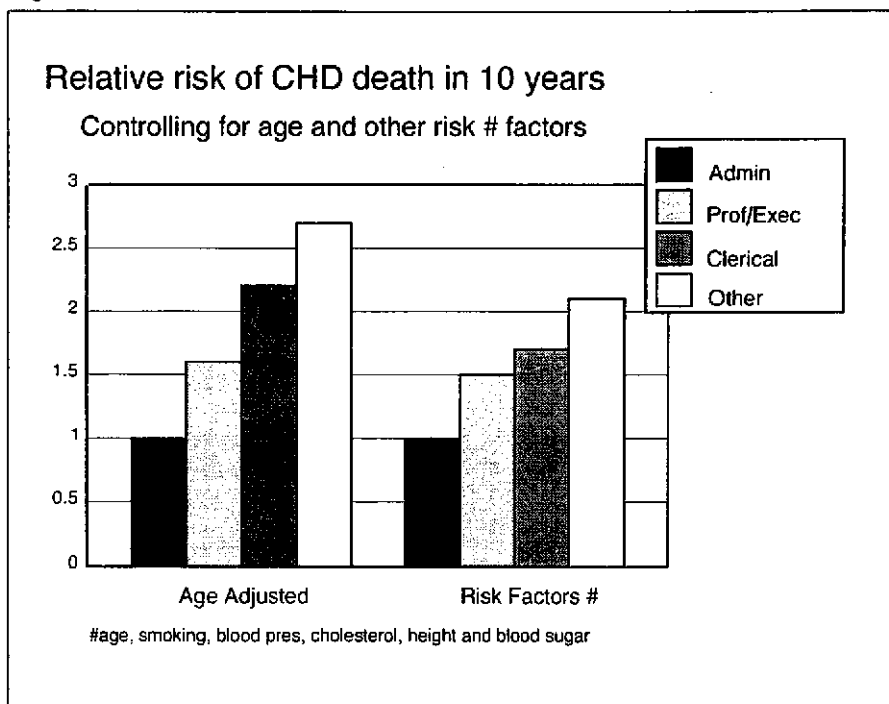
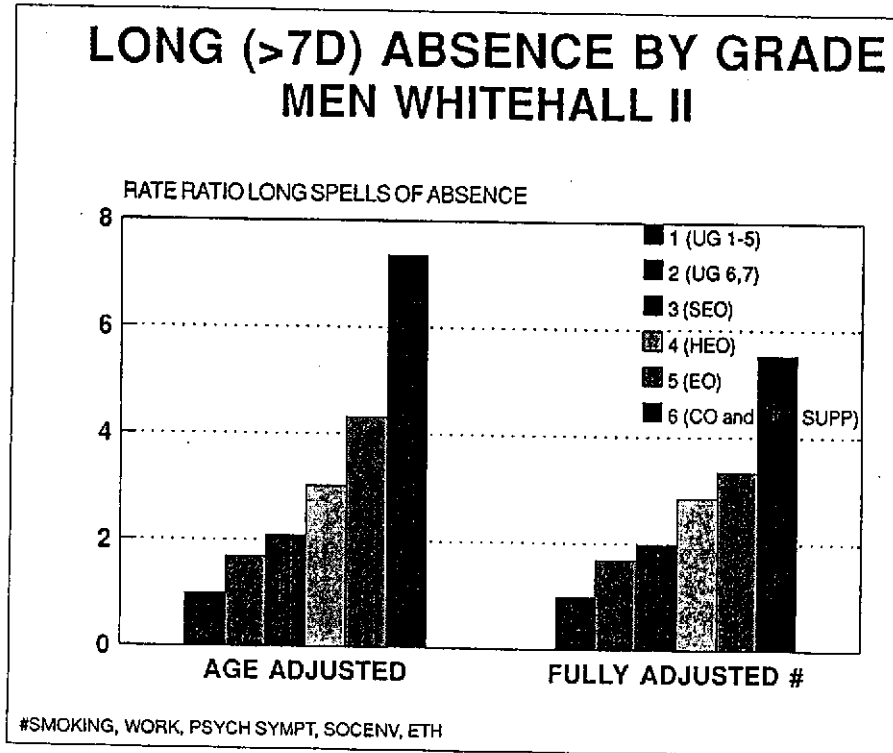


Figure 7



# Discussion of Papers by Michael Marmot and Toshiyuki Furukawa: Inequalities in Health: Bridging the Gap Between Individual and Ecological Approaches

---

*George A. Kaplan*

As we have seen in Professor Marmot's and Furukawa's presentations, the impact of social class on health is substantial, pervasive, and complex. In commenting on these two papers, I have taken as my task to attempt to draw some conclusions as to a conceptual and programmatic road map which might guide our understanding of how these approaches to the social determinants of health might fit together. By way of demonstrating what I believe to be a potentially useful approach, I will share with you some of the findings of my colleagues and myself. While the story is far from complete, the result of this activity has been to seriously challenge, in my mind, the paradigms we have used for understanding health inequalities, whether by class or geography, and to suggest, instead, a new approach more closely related to ecology than to contemporary epidemiology or demography.

Let me start with Figure 1 which shows the survival experience of a representative sample of Alameda County, California residents as a function of family income adjusted for family size, categorized according to government levels<sup>(1)</sup>. Over 19 years, those whose income level was "inadequate" had twice the risk of death compared to those whose income level was "very adequate." Results like this have, of course, been found in many studies. Following the standard analytic procedure, we statistically adjusted for a wide variety of factors - race, smoking, physical activity, obesity, prevalent disease, and others. The aim in such analyses is to see if income, in this case, has a statistically "independent" association with risk of death after taking into account differences in other risk factors which might otherwise "explain" this effect. Leaving aside this notion of "independence," which I shall return to later, even after adjustment for many potential factors there was still a gradient of increasing risk associated with decreasing income, with those with at the bottom level having a risk of death 63% higher than those at the top level.

The finding that income is independently related to risk of a health outcome, even after many potential explanatory variables are considered, is not an unusual result - in fact it is found more often than not. However, despite the familiarity of the finding, we are left in something of a conceptual muddle, unable to explain why there is an inverse gradient between income and health.

It is certainly possible that we did not have the correct measures to properly explain such an "independent" effect. Another possibility, which we chose to explore, was that we needed to change the level of explanation, to consider not individual factors, but instead socioenvironmental factors which might capture some critical aspects differentiating between people at different income levels. Our first step was to examine the impact of residence in a federally-designated poverty area, characterized by a broad range of disadvantage, on subsequent survival, again using the Alameda County Study cohort<sup>(2)</sup>. Those who lived in such an area had 55% higher risk of death over nine years compared to those who lived outside of this area.

This increased risk was again "independent" of a large number of other factors, including income and education, as well as behavioral and social risk factors. Interestingly, in these analyses, residence in a poverty area was a much stronger predictor of survival than was income or education. Turning our attention to a cluster of socioenvironmental variables reflected by residence in a poverty area seemed to lead to some intriguing results.

The impact of such area-type variables on health is not restricted to poverty, per se. Based on a factor analysis of a large number of census tract characteristics, we identified several clusters of census tract characteristics which were also related to the risk of individuals (1). Figure 2 presents the census tract characteristics for one of these clusters and death rates for respondents in the Alameda County Study based on a scale composed of these characteristics. Residence in census tracts with high proportions of older males, widowed males, and inadequate housing was associated with a 50% elevation in mortality risk. Interestingly enough, this association was "independent" of a large number of other factors such as race, income, education, marital status, smoking, obesity, social isolation, depression, and other factors. But again, we were no farther along in understanding these "independent" effects.

Still, the emphasis on area indicators as determinants of mortality outcomes turned our attention to geography (3). As Figure 3 shows not surprisingly, excess mortality, represented by standardized mortality ratios (SMR) greater than 1.00, is not distributed randomly in Alameda County. What is of interest, is that the same areas of the county have high SMRs for both men and women, and for causes of death which occur early in life, in middle age, and in old age.

In order to understand some of the reasons for this particular spatial patterning, we conducted a telephone survey in Alameda County, collecting a large amount of data on behavioral, social, psychologic, health status, and other indicators. When we examined the distribution of these measures by zip code, we found that on virtually every measure, those who lived in the high SMR areas were worse off. Furthermore, when we aggregated these characteristics into categories of demands and resources, those who reported high levels of demands and low resources were almost 10 times more likely to live in the high SMR areas.

It is possible, of course, to raise many lines of objection to the direction our efforts have taken. For example, poor measurement of confounders may lead to an overestimation of the independence of other factors, area characteristics may not represent the risk factors for individuals, etc. I believe these are largely technical issues, which obscure what may be the main lesson of this exercise. In the health sciences, there is an unremitting and unrelenting search for the cause, for the single factor or mechanism which will explain a particular disease. This search for specific, single, causal factors may seriously miss an important point illustrated in a circuitous way by the examples I've presented. In the world of biological as well as social affairs, complex, multi- and inter-determined, causally reciprocal processes are the rule rather than the exception. Our search for single, independent causes does damage to this reality. For example, where low income individuals or high SMR zipcodes are characterized by higher rates of unmet need and inadequate resources across a wide spectrum of domains, no explanation which focuses on a single, underlying cause is likely to be very realistic, successful, or informative.

At this stage, I would argue that we need to proceed in two directions. First we need to take seriously the complexity of the environments in which people live, to characterize the nature of this complexity, and create a new typology of ecologic niches which captures both the homogeneity within similar environments and the heterogeneity between environments. Much of this can be informed by consideration of the geographic patterning of health outcomes. Second, we need to develop new approaches to our understanding of the biology of health and disease which are appropriately complex. Isolating critically important aspects of the environment was much simpler when single sources of contagion were involved, isolating

environmental properties reflecting complexity of systems will be more difficult. For example, perhaps there are particular physiologic response patterns which are associated with residence in areas characterized by high levels of demands and low levels of personal, social, and institutional resources. Or, perhaps there are particular pathophysiologic processes which are part in parcel of living in environments which are non-responsive to human needs, or in which there are global patterns of inequality.

What this amounts to is a repudiation of the billiard-ball physics model of the determinants of health and disease in individuals and populations which has characterized much of the discourse on this topic. Note also that this approach implies a very different approach to intervention than is usually taken. Instead of focusing on increased screening or medical care, a single risk factor, or even a collection of risk factors, interventions would be broad scale, influencing, in addition, issues of job creation, community development, communication and information flow, environmental design and architecture, and other measures which might be at the core of the community's health.

Such an approach may very well help us to understand the patterning of health outcomes by social class and place, and even more important, to do something about it. While some would argue that such an approach is beyond the pale of the health sciences, I would simply point out that it has its precedents in a quote from Virchow who argued "Medicine is a social science and politics is nothing but medicine on a grand scale."<sup>40</sup>

## REFERENCES

1. Haan MN, Kaplan GA, Syme SL. Socioeconomic status and health: old observations and new thoughts. In: Bunker JP, Gomby DS, Kehrer BH, eds. *Pathways to Health: the role of social factors*. Menlo Park, CA: Henry J. Kaiser Family Foundation, 1989; 319-420.
2. Haan MN, Kaplan GA, & Camacho T. Poverty and Health: prospective evidence from the Alameda County Study. *American Journal of Epidemiology*, 1987; 125: 989-98.
3. Kaplan GA, Cohen RD, Cirillo P, Davis WB, Bartlett J, Johnston M, Wright WE, Mendlein J. Behavioral, psychosocial, demographic, and health factors associated with residence in high SMR zip codes. In: *Proceedings of the 23rd National Meeting of the Public Health Conference on Records and Statistics*. July 1991. DHHS Publication No. (PHS) 92-1214. Washington, D.C.: National Center for Health Statistics, 1992.
4. Ackernecht EH. *Rudolf Virchow: doctor, statesman, anthropologist*. Madison, WI, 1953.

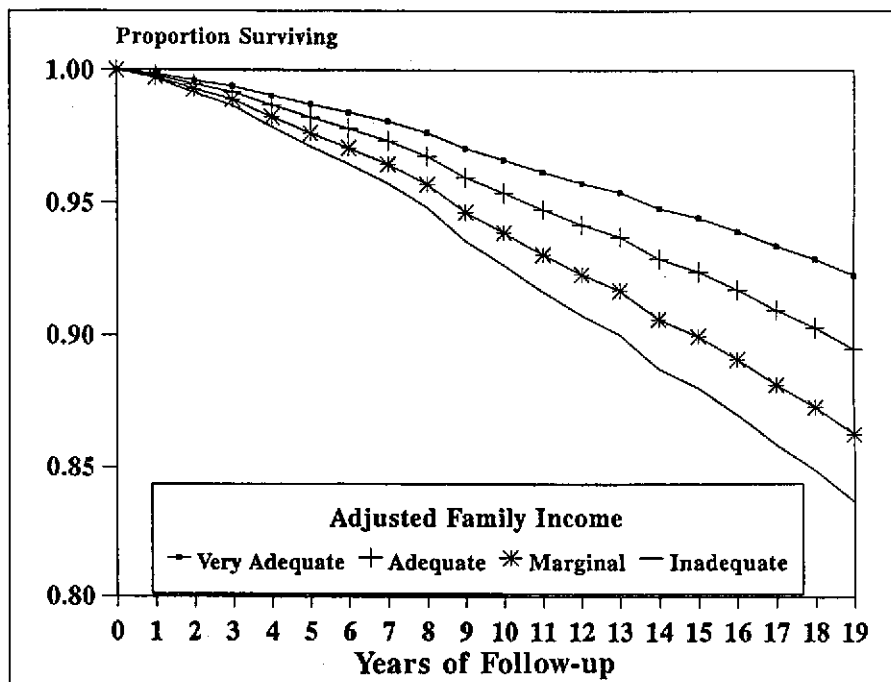


Figure 1 Family income level and survival over 19 years: Alameda County Study

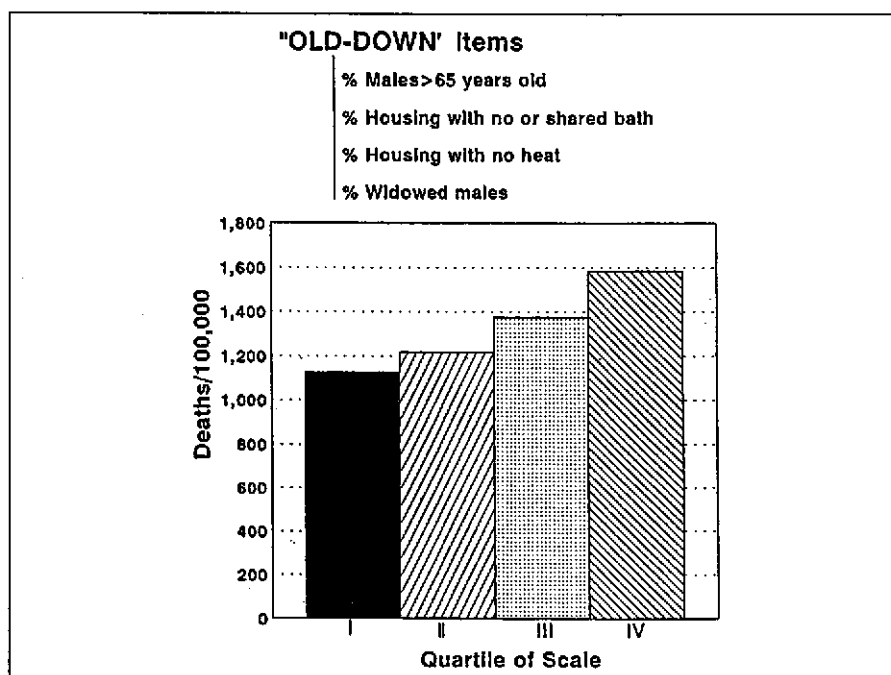
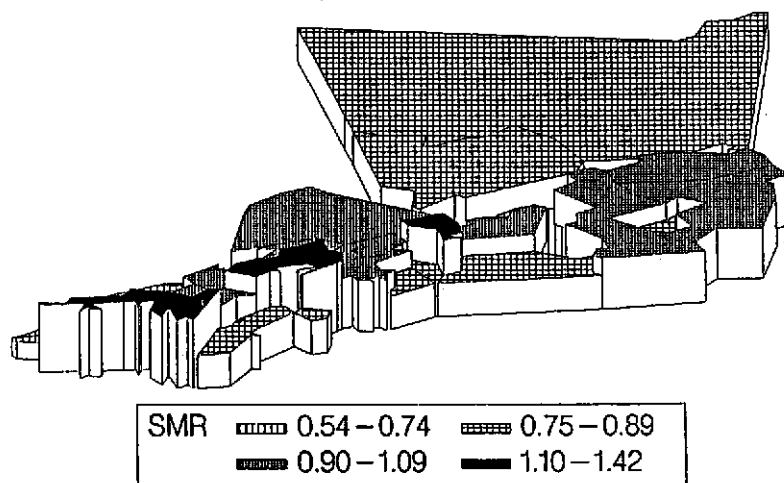


Figure 2 Death rate by quartile of "Old-Down" scale: Alameda County Study

# SMRs for All Cause Mortality by Zipcode Alameda County, California: 1984 – 1988



**Figure 3** Standardized Mortality Ratios for Alameda County zipcodes, 1984-1988

# Discussion of Papers by Toshiyuki Furukawa and Michael Marmot, concerning "Variations in Health Status Across and Within Populations."

---

*John Frank*

## Overview

**T**hese two papers are helpfully complementary, dealing as they do with the question of inter-nation variance in life expectancy across developed and developing countries (Prof. Furukawa's paper) and within-population variance in individuals' mortality and morbidity (Prof. Marmot's paper). Specifically they tell us what is currently possible, in terms of explaining socioeconomic gradients in human health outcomes at the macro- and micro-levels. They also are more than candid in admitting how much is not explained, at least in terms of independent variables that we are presently capable of measuring and analyzing. They therefore give us both the "extraterrestrial prospective" of someone looking down from space at gross patterns of health status across human societies, as well as the more microscopic pattern of differences in individual human beings' health status within a given setting.

Pleasing as these honest and comprehensive assessments are of what we have learned about socioeconomic gradients in health status, in both contexts, they somehow leave this observer, at least, with more questions than answers. Let me pose a few of them here:

## METHODOLOGICAL ISSUES

**Choice of Outcome Measures:** The first question that arises for me is whether or not mortality, and particularly life expectancy, are appropriate outcome measures for capturing the health status of human populations as a whole, particularly as may be related to current socioeconomic conditions. Here are some reasons why these measures may be problematic:

- First, there is the narrow technical problem that life expectancy is a rather complicated summary measure of (current) age-specific mortality rates and therefore inherently not likely to give simple linear secular trend-lines, as logarithmic transforms of age-standardized mortality rates tend to do.
- More broadly, one has to worry that mortality is subject to particular problems in acting as a proxy for the fuller range of human health outcomes; one of these is latency. Death has been termed, "the end of a long process called life". This implies, in turn, that cumulative exposures and experiences over many decades may contribute to the timing and nature of the end of our days, particularly in developed countries where the vast bulk of the deaths occur among the elderly. Indeed, if Barker et al. are correct, it may be that even in utero exposures are relevant to the timing and cause of death in later life.

More fundamentally, one worries that the way which we end our days may be a poor reflection of the way in which we spent most of them. It is increasingly common for elderly

people to suffer from lethal conditions which have very little to do with their health in the previous decades, - decades which may have involved considerable suffering (for example from mental illness or arthritis). Therefore the problem of relevance. Does it really matter how you die or even that you die at 80 or 85? Is it not more important how you lived, before that?

All these arguments are not new. They are part of the progressive march towards obsolescence of mortality as a key health status indicator in human populations that have reached a certain stage of socioeconomic development. Marmot's work on sickness absence from work is a significant advance in this respect. However, we need more analyses of quality of life, and particularly the direct measures of human suffering, which often entail the use of "softer" concepts such as loneliness, sadness and/or family and social functioning.

**Choice of Analytic Methodologies and Null Hypotheses:** There is some confusion, it would appear from the literature, about what we would expect social inequalities in health status, including those captured by mortality rates, to do over a period of time. This is indicated by our collective tendency in epidemiology to choose, in some situations, to represent time trends in mortality-by-social-group on a logarithmic scale, as opposed to an arithmetic scale. What is it that we are putting forward as the null hypothesis, exactly? Are we expecting cross-sectionally computed standardized mortality ratios to effectively stay the same over time, in which case the logarithmically transformed graphs of mortality rates would stay the same vertical distance apart over time, even if the arithmetic versions of these graphs showed "narrowing gaps"? Although these questions of presentation have not entered directly in today's papers, they are critical to the discussion of such economic gradients and their patterns over time.

#### ISSUES IN INTERPRETATION

Prof. Furukawa has eloquently suggested that the really interesting variables that might help us explain life expectancy differences internationally are probably just not available in routinely published statistics. The problem is that one is often still uncertain, upon discovery of a variable that shows this sort of strong relationship, whether one is studying a fundamental cause or simply another epi-phenomenon, carried along as an associated proxy for a truly important (but unmeasured) variable still beyond our grasp. I recently saw an unpublished analysis of developing country mortality rates in which availability of basic sanitation was by far the most important correlate of infant and child mortality internationally. While one might want to believe that latrine availability was truly potent determinant of infant and child mortality in the tropics, some years of living in East Africa (where villagers often forbid children's use of latrines in order to "keep them clean"), tell me that is most unlikely to be the case. In fact, latrine availability was probably acting in the analysis in question as a proxy for many other features of the physical and social environment in which infants and children live, that contribute more directly to their health and to their death rates.

To again move to a broader and more substantive issue, the juxtaposition of these two papers raises the obvious question, "Why should correlates of mortality (or any other health outcome) at the individual-human-being-level be similar to those at the community/nation level?" A clear example is the work of Wilkinson, quoted by Prof. Marmot, showing the strong relation of certain measures of income distributional equity, at the national level, to health statistics such as mortality rates and life expectancy. Equity is an attribute of a population, not of an individual, and so can only appear in the one sort of analysis.

To go a little deeper into this, I wonder if we couldn't learn something from well-established modes of thought in population ecology. Scientists in that field have long been aware that certain measures of the complex interactions between species and species groups, inside a population or community of plants and animals, may be much more sensitive indicators of the health and well-being of the whole community than the specific health status of a given species or of individuals within it. A simple example would be the degradation of complex species diversity, in situations where environmental pollutants tend to greatly reduce the number of species present. Only by examining the changes in species distribution can one see these effects in their subtle early stages. [Note the analogy to Wilkinson's findings concerning societal income distribution and health indices.] An even more subtle example is the discovery that changes in the nitrogen handling and cycling of some forest communities seem to be the very first indications of "envirostress". These changes cannot be seen by examining the individuals among the species present in the community, or by counting them, as in the example above. They can only be seen by conducting a sensitive analysis of "community nitrogen metabolic" measures that reflect the function of the community as a whole.

Surely there is an metaphor here for human societies, in that the health of whole communities of organisms like ourselves must be greater and more complex than the sum of the parts - particularly since we are so evidently social animals.

#### FINAL THOUGHTS

If I had to draw to any conclusions from this most stimulating session, to guide further research, it would be that somehow we need to do a better job of bridging the gap between macro-health-status studies across nations and micro-health-status studies, among individuals in a given population. We need to build-up sequentially our understanding of the relevant interactions between individuals, and groups of individuals, that help determine health status at the aggregate level. It strikes this observer that what is missing is some serious integration of qualitative social science methods with the quantitative methods so eloquently demonstrated by our two presenters today. We particularly need to focus in on contextual details that might help us understand in-depth studies of special populations and their members much better. An example from today's session might be "What do we really know about the kind of place that Whitehall is to work in, compared to other similar offices elsewhere in the public sector, whether in the U.K. or other developed countries?" Therefore one hopes that today's juxtaposing of these macro- and micro-views of social gradients in health status would lead us to work harder in future on the central problem of how they might be better meshed, and thereby further our understanding of the determinants of health, from the subcellular to the clinical, through to the extraterrestrial perspective.

# The Social Environment and Health

---

*S. Leonard Syme*

There is a growing interest in identifying factors in the social environment as important determinants of health and well being. This interest comes primarily as a result of the fact that we have not been very successful in identifying risk factors at the individual level. In fact, we have had three major problems with our work at the individual level and all three of these problems have led us inevitably to think more seriously about the social environment as an important focus for our work.

The first problem is that we have done such a poor job in identifying risk factors for disease. As one example of this problem, consider our achievements in the area of coronary heart disease – the disease that has been studied most aggressively for over 40 years with the highest level of financial support the world has ever seen. During these 40 years of incredible world wide effort, a large number of seemingly important risk factors have been identified. The three that everyone agrees on are cigarette smoking, high blood pressure and high serum cholesterol. Dozens of other risk factors also have been proposed but not everyone agrees about them. Included here are such risk factors as obesity, physical inactivity, diabetes, blood lipid and clotting factors, stress, and various hormone factors. When all of these risk factors, combined, are considered together, they explain about 40% of the coronary heart disease that occurs<sup>1</sup>.

How is it possible that after 50 years of massive effort, all of the risk factors we know about, combined, account for less than half of the disease that occurs? Is it possible that we have somehow missed one or two crucial risk factors? While this is of course possible, it should be noted that the relative risk of the new risk factors would have to be enormous to account for the other 60% of the coronary heart disease that occurs. It seems not very likely that we would have missed one or two risk factors of such enormous power and importance. And, it must be said, our record of success in the area of coronary heart disease is one of the very best; the results for other diseases is far less impressive.

I do not mean to suggest by these remarks that the risk factors we have identified are unimportant. They are important and they have been useful in the prevention of coronary heart disease, but, clearly, there are other very important issues involved that we do not yet understand.

The second problem we face in epidemiology is that even when we do identify disease risk factors, it has proven very difficult for people to make changes in their risk behavior. An example of this problem is provided by the Multiple Risk Factor Intervention Trial (MRFIT)<sup>2</sup>. This project was an experiment designed to reduce the death rate from coronary heart disease in the United States. It was to be the largest, most ambitious and most expensive experiment ever designed to see if by getting people to change high risk behavior, we could reduce the coronary heart disease death rate.

This trial required that we screen 500,000 men in twenty-two cities across the United States to find 12,000 men in the top risk category of coronary heart disease by reason of their cigarette smoking, high blood pressure and high serum cholesterol. A random half of these men worked with us in our clinics to lower their risk while the other half were sent back to their own doctors. After six years of massive effort with the men in the special care group, disappointingly few of them changed their high risk<sup>3</sup>. To be specific, 42% of the smokers

stopped smoking which is probably the best record ever achieved in a smoking cessation program. But 58% continued to smoke. Of the men with hypertension, about half had it under control at the end of the trial, but about half did not. Our record in reducing serum cholesterol by dietary intervention is probably not worth even mentioning: We achieved a 6.7% reduction in serum cholesterol levels.

So some changes took place, but in this group it should have approached 100%! This group of men was selected in the best way we know to be perfect compilers. The limited success we did achieve is, I argue, probably the best we can ever hope to achieve in one-to-one intervention programs. It was the high-water mark. The message from this extreme example is clear and well known: it is very difficult for people to change high risk behaviors even when they really want to and even when we do all that we can to help them.

The third problem we have in epidemiology is that even when people do successfully change their high risk behaviors, new people continue to enter the at-risk population to take their place. For example, every time we finally helped a man in the MRFIT project to stop smoking, it is probable that, on that day, one to two children in a school yard somewhere were for the first time taking their first tentative puffs on a cigarette. So, even when we do help high risk people to lower their risk, we do nothing to change the distribution of disease in the population because, in one-to-one programs like MRFIT, we have done nothing to influence those forces in the society that caused the problem in the first place.

We epidemiologists tend to look at individuals in order to find causes of disease even though it is clear that this will not be helpful in understanding the distribution of disease in the population. This point was forcefully made at the turn of the century by Emile Durkheim, the French sociologist<sup>1</sup>. Durkheim studied suicide – a behavior that clearly is the result of problems experienced by individuals. Indeed, suicide might be considered as one of the most personal and intimate behaviors possible. In spite of this, Durkheim noted that suicide rates exhibited a patterned regularity over time and place. Thus, he said, suicide rates are consistently higher in certain countries and in certain groups, over time, even though individuals come and go from these groups. If the causes of suicide are rooted in the individual, how can it be that there exists a patterned regularity in groups? Durkheim reasoned that there must be something about the group that somehow promotes a higher or lower rate in the group. This factor would not, of course, account for which individuals in the group committed suicide but it would account for the fact that the rate in the group was consistently high or low.

The value of this approach is that it can lead to a more effective preventive medicine. This is not to deny the importance of working with individuals one at a time: this clearly is worthwhile and important. But it is at least as important to recognize that this approach is very limited because it does nothing about those forces in society that cause our problems in the first place and that will continue to provide a fresh supply of at-risk people, forever. If epidemiology is to be useful in developing a preventive approach, it will need to study the community and the population. As matters now stand, almost all epidemiologists study large numbers of individuals in communities. This is not epidemiology. It is clinical medicine in large groups.

How did this unfortunate circumstance come to pass? One reason is that the clinical tradition runs so strong that it has come to dominate not only clinical medicine, but epidemiology and public health as well. In epidemiology, we typically use a clinical classification of diseases in our research. Thus we study diseases such as heart disease, cancer, and AIDS. These are clinical conditions that are of concern to clinicians- to those who must work with individuals on a one-to-one basis to treat or control their disease. Heart disease, cancer, and AIDS are not public health problems, they are clinical problems. And yet much of the funding of epidemiologists and much of the way epidemiologists organize themselves nevertheless is centered around these clinical disease categories.

Let us think about the approach that epidemiologists used in the control of infectious diseases. That approach to infectious disease was organized around such diagnostic categories as water-borne, food-borne, air-borne, and vector-borne conditions. This diagnostic classification scheme is of little use in clinical practice, since it does not give any insight into treatment possibilities. But the job of the epidemiologist is not to treat individuals anyway. What insight this classification scheme does provide is to target those aspects of our environment that are generating disease and it suggests at what points in the environment our interventions might be deployed. This cannot be done with words such as "coronary heart disease", "cancer", and "AIDS".

In short, epidemiologists have subtly adopted a disease classification scheme from the clinical model that inevitably focuses attention on clinical states of health and that inevitably draws attention to the affected individual.

When social, community and environmental forces are considered – and, of course, they are – they are thought about to shed light on the individual and his health. The focus is not where it ought to be – on the social and physical aspects of our environment as objects of concern in their own right.

There are at least two reasons for why it is so difficult to think in terms of the social environment. As indicated earlier, the clinical tradition in our work is so pervasive it overwhelms all other approaches. In the United States, it certainly has overwhelmed the National Institutes of Health (NIH). One can apply for money from the NIH to study arthritis and metabolic disease, heart disease, cancer, and eye disease, but one cannot request funds to study health at the National Institutes of Health. Only diseases of clinical relevance. I have nothing against categorical support but I do question why almost all the money goes to clinical categories and not to public health categories. Of course, it does not help that we do not have a well-reasoned and useful set of categories to offer.

The second reason that it is difficult to think about the social environment as a determinant of health is more subtle but perhaps even more important. It may be that a community approach to infectious disease is easier to develop because most of us are exposed to infectious agents whether or not we want to be. A contaminated water or food supply puts all of us at risk, as does toxic air or infected mosquitoes. It is clear that public health agencies should look after these things on behalf of us all. This clearly is a public health issue. In contrast, the way we eat, drink, smoke, drive, sit, run and work can be seen as being our own affair and not anyone else's business. These behaviors, it could be said, are private matters and are not the province of public health concern. In this view, we are responsible, each of us as individuals, for the heart disease, cancer, and AIDS that we get.

As individuals, all of us ultimately are responsible for our own health and our own behavior. But it is naive to think that we are free agents in this. All of us are influenced by forces in the community that shape our choices and preferences. When we go to the large food market, it almost always is true that the bad foods are at eye level, badly labelled, and reasonably priced, while healthy foods with helpful labels are tucked away on remote shelves and are more expensive. Do we really have a free choice in that circumstance? The way in which food market people arrange foods slants that choice in certain directions. We would have more of a free choice if both good and bad foods were placed side-by-side at competitive prices.

How about the teenager being bombarded in subtle and not so subtle ways about the attractiveness of smoking? That teenager really does not have a free choice until we in public health provide an equally compelling argument against smoking. The environment at present is still pro-smoking and we need some good environmental interventions to redress that imbalance.

When one enters a 2,3, or even 4 story building, at least in the United States, it is easy to notice that the elevator is always placed in the center of the building and that it often has music, carpeting, and soft lighting. The stairs, if one can find them, are at the end of the building with bare floors and harsh lighting.

Just as bad water and food affects all of us, unhealthful forces in our society influence our choices and opinions and thereby affect us all. It makes no sense to ignore a contaminated water supply and instead, tell people to boil their household water supply. It make no sense either to urge people to change their unhealthful behaviors when we leave untouched those forces that prompted much of that behavior in the first place.

To this point, I have discussed the 3 major problems in our epidemiologic research: our difficulty in identifying disease risk factors, our difficulty in helping people change their risk once they know about it, and the fact that, even when people do change, new people keep coming to take their place. I have emphasized the importance, for an effective preventive approach, of a community approach to health and disease. It is, of course, easy to talk about the need for a community approach but it is not as easy to see how we might in fact go about doing this type of work. I think the first thing we need to do, following Durkheim, is to take advantage of patterned regularities in morbidity and mortality rates.

The most impressive patterned regularity of all is social class. Everyone knows, and has known for hundreds of years, that people lower down in social class have higher rates of virtually every disease and condition<sup>5-7</sup>. In spite of this universal recognition, we know almost nothing about the reasons for this phenomenon. The list of possible explanations is long and well-known. It includes poverty, bad housing, unemployment, poor nutrition, inadequate medical care, and low education. We do not know the relative importance of these various factors because we do not study social class. Social class is of such overwhelming power that we epidemiologists, in our research, typically "hold it constant" so that we can study other things. If we did not, social class would swamp all other factors and we would not be able to see the role of any other issues. In consequence, we know virtually nothing about the various sub-components associated with social class.

But there is another, even more important, reason for our failure to study social class. This reason is that we do not feel that there is anything that can be done about it. Social class, we say, is a product of vast historical, economic and cultural forces and, short of revolution, it is not something one targets for intervention. So we give up and instead urge people to lower the fat content of their diets. The problem with the view that there is nothing to be done about social class is that it is based, not on facts, but on speculation. If research were to show that people in the lower social classes had higher rates of disease because they were poor, it might be true that interventions would be difficult. But we have no evidence that lack of money is, in fact, the major culprit and, without knowledge, it seems premature to conclude that social class is too difficult to consider or deal with.

The study of British civil servants by Marmot & colleagues provides a brilliant example of how we might better approach this issue<sup>8</sup>. They show in their study of British civil servants that those at the very bottom of the civil service hierarchy have heart disease rates four times higher than those at the top. After adjusting for such heart disease risk factors as blood pressure, serum cholesterol, smoking, physical activity, and so on, the difference between these groups is still threefold. But in this study, they show that those one step down from the top of the hierarchy, civil servants who are professionals and executives, such as doctors and lawyers, have heart disease rates that are twice as high as those at the very top – those at the very top being upper class directors of agencies, all of whom have been educated at Oxford and Cambridge, and whose career usually ends with knighthood.

It is not surprising that those at the bottom have higher rates of disease than those at the top, but it is surprising that doctors and lawyers, one step down from the top, also have higher

rates. Doctors and lawyers are not poor, they do not have bad houses or bad medical care, and they do not have poor education or poor nutrition. It is not just that those at the bottom have the highest rates of heart disease: there is a gradient of disease from the top of the British civil servant hierarchy to the bottom.

At first I thought this phenomenon might be a finding that was somehow unique to the British civil service. It is not. We recently completed a review of this issue and we found a similar gradient almost everywhere in the world and for virtually every disease that has been studied<sup>9</sup>. The problem posed by this finding is the following: we can imagine why those at the bottom have higher rates of disease but how can we explain a gradient? How can we explain the findings that those one or two steps from the top have higher rates of disease even though they do not suffer from the problems experienced by those at the bottom? People one or two steps from the top do not experience poverty, or poor nutrition, or problems of access to medical care, or bad housing, or poor education.

I do not know the answer to this. The only hypothesis I have been able to come up with is that as one moves down the social class hierarchy, one has less control of one's destiny. By this I mean less opportunity to influence the events that affect one's life<sup>10</sup>. At first, I thought this idea about control of destiny was a brilliant, original idea. It is not. In fact, it was a little embarrassing to realize that this great truth that I had discovered had been a topic of research by psychologists and sociologists for dozens of years. Many scholars have studied the concept of mastery, self-efficacy, locus of control, learned helplessness, controllability, predictability, desire for control, sense of control, powerlessness, hardiness, competence, and so on.

It is intriguing that so many different scholars, from different backgrounds and with different research objectives, should come up with ideas that are so similar to one another. More recently, a group of epidemiologists led by Robert Karasek<sup>11</sup> from the United States and Tores Theorell<sup>12</sup> from Sweden have shown that rates of coronary heart disease are higher among workers who experience not only high job demand but low discretion and latitude for dealing with these demands.

I do not know if the idea of control is correct or not, but I do know that some idea like it is necessary to deal with the difficult problems we face in trying to prevent disease. We need an idea that will help us understand why disease rates are higher in certain groups than others – over time – even though people come and go from these groups.

Since I do not at present have a better idea than that of control, we are using it in helping us understand the health of bus drivers in San Francisco, California. Several previous studies have noted that bus drivers have a higher prevalence of hypertension, as well as diseases of the gastrointestinal tract, respiratory system and the musculoskeletal system as compared to workers in other occupations<sup>13</sup>. These results have been obtained from studies from different transit systems, under different conditions, in several countries. Based on these findings, it has been suggested that certain aspects of the occupation of bus drivers may create an increased risk of disease for workers in that occupation.

From a clinical view point it is of value to identify drivers with disease in order to treat them. It would also be of value to teach drivers about better posture, more healthful eating habits and alternative ways of dealing with job stress. However, from an environmental perspective, it would perhaps be more useful to identify those aspects of the job itself that might be changed to prevent those diseases to prevent these diseases.

In our study of drivers, their exposure to noise, vibration and carbon monoxide fumes is being monitored but particular attention is being paid to the social environment of the driver. For example, in preliminary studies of drivers, the "tyranny of the schedule" has been forcefully brought to our attention. Drivers must keep to a specific schedule, but in almost every instance, this schedule is arranged without realistic reference to actual road conditions, and in

fact cannot be met. If this and other characteristics of the job that are associated with disease can be identified, it may be possible to introduce interventions, not merely among bus drivers, but directly on those environmental factors associated with the job. For example, it may be that by changing the way in which schedules are arranged, the bus company would be able to earn more money than it loses because of lower rates of absenteeism, sickness, accidents, and in particular, turnover of employees.

In the case of bus drivers, a clinical focus either on hypertension, respiratory problems, gastrointestinal tract diseases or musculoskeletal disorders clearly is useful. However, from an environmental and preventive perspective, it might be useful to group together these different diseases and conditions associated with common work exposure so they can be studied as related phenomena. If this is not done, the circumstances they share will not likely be appreciated.

In conclusion, there seems now to be enough evidence to warrant a serious research program on the social environment as an important determinant of health. Our work with individuals has been useful and productive, but this approach by itself clearly will not lead to an effective program of health promotion and disease prevention. A new initiative focusing on the environments in which we live must now become a priority for us all.

## REFERENCES

1. Marmot M, Winkelstein W. Epidemiological observations on intervention trials for prevention of coronary heart disease. Amer. J. Epidemiol. 101:177-181, 1975
2. Multiple Risk Factor Intervention Trial Research Group. The Multiple Risk Factor Intervention Trial. Prev. Med. 10:387-553, 1981
3. Multiple Risk Factor Intervention Trial Research Group. The Multiple Risk Factor Intervention Trial - Risk factor changes and mortality results. JAMA 248:1465-1476, 1982
4. Durkheim E. "Suicide: A Study of Sociology" (Edited and translated by G. Simpson). Free Press, Glencoe, Ill., 1951
5. Antonovsky A. Social class, life expectancy and overall mortality. Milband Mem Fund Q 45:31-73, 1967
6. Syme SL, Berkman LF. Social class, susceptibility and sickness. Am J Epidemiol 104:1-8, 1976
7. Kitagawa EM, Hauser PM. Differential Mortality in the United States. Cambridge, Mass. Harvard University Press, 1973
8. Marmot MG, Rose G, Shipley M, Hamilton PJS. employment grade and coronary heart disease in British civil servants. J Epidemiol Community Health 3:244-249, 1978
9. Haan MN, Kaplan GA, Syme SL. Socioeconomic status and health: Old observations and new thoughts. In Bunker JP, Gomby Df, Keherre BH (eds): Pathways to Health: The Role of Social Factors. Palo Alto, Calif. H.J. Kaiser Family Foundation, 1989

10. Syme SL. Control and health: A personal perspective. In Steptoe A, Appels A (eds): Stress, Personal Control and Health. New York: John Wiley, 1989
11. Karasek R, Baker D, Marxer F, Ahlborn A, Theorell T. Job decision latitude, job demands, and cardiovascular disease: A prospective study of Swedish men. Am J Public Health 71:694-705, 1981
12. Theorell T, Alfredsson L, Knox S, Persk A, Svensson J, Waller D: On the interplay between socioeconomic factors, personality and work environment in the pathogenesis of cardiovascular disease. Scand J Work Environ Health 10:373-380, 1984
13. Winkelby MA, Ragland DR, Fisher JM, Syme SI: Excess risk of sickness and disease in bus drivers: A review and synthesis of epidemiologic studies. Int J Epidemiol 17:124-134, 1988

# The Lifelong Impact of Childhood Experiences: A Population Health Perspective

---

*Clyde Hertzman*

A belief in the lifelong significance of childhood experiences is implicit in many late twentieth century institutions. It reflects itself in the special protection of the physical, social, and ideological environment of children, and in the extent of investment in education for the young. Yet despite its protean manifestations, there is no working consensus on just exactly how, and to what extent, the experiences of childhood influence later life.

Somewhat romantically, we could envision a bygone era when a working consensus did exist. The movements to eliminate child labour and bring about universal publicly funded education must have been sufficiently well articulated by their advocates; understood by the population; and accepted by those with antagonistic material interests to have overcome the prevailing norms of the time. Some insights of that era have endured. Most of us, if challenged, could recite commonly accepted wisdom about why reintroducing illiteracy and semi-slavery among children might bode poorly for them, and for society, in the decades to follow.

An analogous claim cannot be made for the present; when societal change threatens both the social environment of the child and the economic viability of the family and community. For example, across North America approximately 50% of single parent families live in poverty<sup>1</sup>, and the overall proportion of children so affected will likely continue to rise with the rising proportion of families headed by a single parent (and with economic stagnation). Among two parent, two wage earner families, still the most common type in Canada<sup>2</sup>, time allocation and child care arrangements are in a state of chaos<sup>3</sup>. Fewer than half of parental couples work fully overlapping, regular work weeks, making family time rare. Young children often find themselves cared for in three to five different places each week. High school drop-out rates hover around 30%<sup>4</sup>. Yet society has no working consensus on the significance of such problems and has not collectively committed itself to addressing them.

This state of indecision embodies a paradox, because rapidly accumulating evidence is revealing an impact of childhood experiences on subsequent health, well-being, and competence which is more diverse, profound, and longlasting than was ever understood in the past.

## **Population Health and Human Development**

"The impact of childhood experiences on subsequent health, well-being, and competence" is a broad topic, and there are many potential avenues of entry into it. One useful avenue is the population health perspective, which addresses the question, "What makes some populations healthier than others?" This line of inquiry begins with a big fact: life expectancy increases with increasing socioeconomic status<sup>5</sup> in virtually all developed countries<sup>6</sup> for which data are available<sup>7</sup>. The pattern is not, typically, a simple difference between rich and poor. Rather, it is a monotonic "gradient," wherein successive population quantiles, graded

according to increasing levels of income or education, or by occupational class, enjoy increasingly long and disability-free lives. This line of inquiry leads to the further observation that the magnitude of socioeconomic gradients in life expectancy have remained largely unchanged throughout the twentieth century, *despite a complete change in the principal causes of death*. This is counterintuitive from a biomedical perspective, since it suggests that the fundamental causal pathways are not dependent upon risk factors for specific life-threatening conditions, such as smoking, drinking, and compliance with seat belt legislation. Rather, the implication is that the diversity of conditions of life can somehow become directly embedded in human biology, such that human vitality can be directly affected by social hierarchies in a consistent manner across wide expanses of space and time<sup>7</sup>.

Insights into the biological dimension come from recent advances in psychoneuroimmunology and psychoneuroendocrinology; fields of inquiry which examine the ways in which perception can affect physiologic function and resistance to disease. For instance, it has been demonstrated that the neurological system can "talk to" the immune and endocrine systems which, in turn, can affect resistance to disease and the function of vital organs, respectively<sup>8,9</sup>. This provides a modicum of biological plausibility to the notion that the conditions of life, filtered through a perceptual screen, could affect vitality through a wide variety of pathological mechanisms. But this is only part of the story. What is missing is an understanding of how differing human experiences, unfolding over time, initiate and sustain processes which lead to differing levels of vitality among different social classes. A developmental perspective which begins at the very beginning of life would seem indispensable here. In practise, this means paying particularly close attention to insights gleaned from longitudinal studies, especially those which begin at birth and follow large population samples for decades into the future.

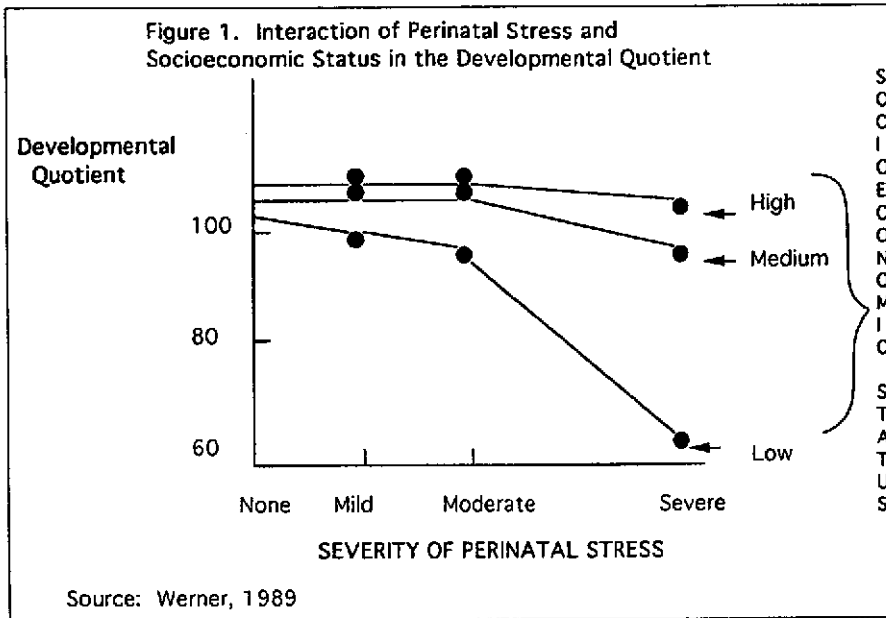
## Complexity

From the perspective which has just been described, health, well-being, and competence are intertwined. The study of well-being becomes the study of the perceptual screen. The study of competence becomes the study of how individuals and groups respond to the conditions of their lives and change them through problem solving. The study of health becomes the study of vitality, stratified by various levels of disability. This conceptual framework, in turn, offers an approach to developing a working consensus on how, and to what extent, the experiences of childhood influence later life.

Relevant research and insights are to be derived from a bewildering variety of disciplines, levels of population aggregation (from the cell to the nation), and species of life. This situation creates opportunity, which is self-evident, and danger, which requires some explanation. That is, there is a high probability that reductionist syntheses of information will intersect with conflicting ideologies to produce protracted, and ultimately futile, debates. Consider the following examples, which highlight different aspects of the complex issue:

### Example 1

In the Kauai birth cohort study<sup>10,11</sup> it was found that severe perinatal stress (i.e. complications of pregnancy, labour, and delivery) *did not* impair the physical and psychological development of children from high and middle socioeconomic status families. Figure 1 shows that, by 20 months, the average "developmental quotient" for low socioeconomic status children who had experienced severe perinatal stress was much below similarly stressed children from higher socioeconomic status families. In contrast, low socioeconomic status children who had experienced mild or moderate perinatal trauma were developmentally much closer



to their more affluent counterparts<sup>10</sup>. The special importance of these findings is that they suggest that certain aspects of high socioeconomic status not only protect healthy children from future risks and vulnerabilities but can actually reverse the latent impact of risks which already exist.

There is emerging evidence that a protective socioeconomic effect may also exist in relation to exogenous threats, such as environmental lead exposure. Lead is particularly useful model of such a threat because it is a ubiquitous environmental pollutant which is toxic to the developing nervous system in ways which can lead to cognitive deficits and nonadaptive behaviour patterns in school<sup>12</sup>. A recent study explored the question of whether or not high socioeconomic status can protect against the developmental effects of lead exposure during pregnancy<sup>13</sup>. Figure 2 gives the average mental development index at 24 months for high and low socioeconomic status children from that study, separated by low, medium, and high lead levels in umbilical cord blood. It shows that the threshold for negative developmental impacts of lead occur at lower levels of exposure among low socioeconomic status children than among children of higher socioeconomic status. As might be expected, at high levels of exposure the developmental effects of lead appear to overwhelm socioeconomic effects and emerge as the principal determinant of developmental delay<sup>13</sup>.

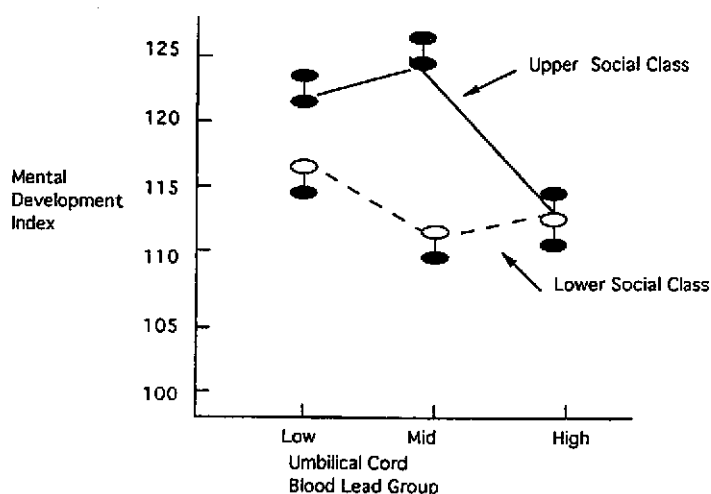
The above examples, which suggest that high socioeconomic status is a powerful buffer against both endogenous and exogenous threats to successful human development, would seem to suggest a broad social policy corollary. Those elements of high socioeconomic status which provide the buffer should be pinpointed and ways found to provide them to children who otherwise would not benefit from them.

### Example 2

Can education in the first two decades of life protect against dementia at the end of life? Until recently, this suggestion was treated as wild speculation<sup>14</sup>, but sufficient evidence has now accumulated for the matter to be taken up in the mainstream of neurology. In a recent

review<sup>15</sup>, the association of low education with an increased prevalence of clinically diagnosed dementia was found to be a consistent finding in studies carried out in six different settings around the world. An analysis of mental status scores in a six year follow-up from the Established Populations for Epidemiologic Studies of the Elderly in the United States showed that education protected against decline in performance on the Pfeiffer mental status test. The hypothesis is that those with more education have greater brain reserve in the form of an increased density of neural interconnections (increased synaptic density) in areas of the brain associated with learning and memory. This increased reserve would have the effect of postponing the time at which declines in brain reserve associated with aging would reach a threshold for dementia. A delay in onset of approximately 5 years could account for halving the prevalence of dementia at any given age<sup>15</sup>.

Figure 2: Mental Development Index at 24 Months of Age by Umbilical Cord Lead Level and Social Class



Source Bellinger et al, 1993, figure 1 (d)

Two crucial questions arise here: Does education actually increase synaptic density in the right places? and Can the effect of education in the first two decades of life be modified by later cognitive activities? Although the first question is currently unanswered, there is some evidence with respect to the second, namely, that cognitively demanding occupations may have a protective effect against dementia<sup>15</sup>. A further question arises from the observation that the association is stronger for comparisons between illiterate people and those completing secondary school than it is between those completing primary versus secondary school. Does this indicate that the effect is broadly dose-responsive, thus fulfilling an important epidemiologic criterion for causation, or does it suggest that there is a critical period of time in brain development, found around the usual time that individuals are taught to read, when life-long protective neural interconnections are most efficiently made?

If the "dose-response" model of educational effectiveness were valid, and if cognitively demanding occupations prove protective against dementia, there would be reason to extend

the social policy corollary described following example 1 to the rest of the life cycle. If the "critical period" hypothesis were correct, then a targeted investment during this time in childhood would appear to be more appropriate.

### Example 3

A series of studies has been carried out which examined the lifelong impact of "handling" newborn rats<sup>16</sup>. Handling involves removing the mother from the litter, placing individual pups into a new cage for 15 minutes, then returning both mother and pups to their cage. This is done once per day for the first 3 weeks of life<sup>9</sup>. When compared to non-handled pups, this simple intervention was associated with improved function of the "stress system" throughout the life cycle (i.e. lower basal corticosterone concentrations and faster physiological recovery in stressful situations). These changes reduced the total lifetime exposure of the brain to corticosterone, which is toxic to nerve cells in a brain structure known as the hippocampus, and thus the rate of loss of nerve cells in the hippocampus was reduced in the handled rats over their life span. Cognitive functions are sensitive to relatively small degrees of hippocampal damage, and so by 24 months of age, elderly by rat standards, the handled rats had been spared some of the cognitive defects typical of aging. The significance of this was demonstrated by a test wherein rats had to find a submerged platform in a pool of opaque water, relying entirely upon visuospatial cues from the surrounding room. Nonhandled rats had a progressive deterioration in their performance with age; in contrast, no deficits occurred in aged handled rats (Figure 3). *Most relevant in this context is the final observation: the handling phenomenon could not be induced at later ages<sup>9</sup>.*

### **Models of the Lifelong Impact of Childhood Experiences**

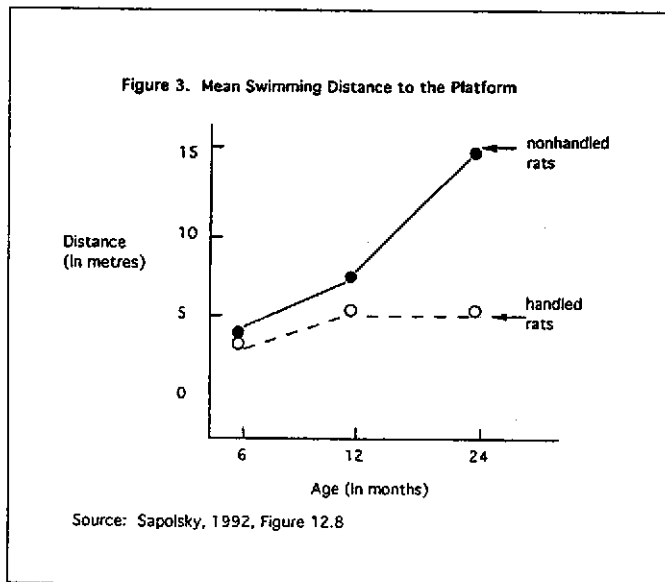
The above examples hint at two different, and sometimes conflicting explanatory models of the impact of childhood experiences in later life. The first, called the "latency" model, emphasizes the prospect that discrete events, which tend to occur very early in life, will have a strong *independent* effect in later life. The second, called the "pathways" model, emphasizes the *cumulative* effect of life events along developmental trajectories, and the ongoing importance of the conditions of life throughout the life cycle.

#### Latency Model

The latency model enjoys a biological precedent in the field of occupational and environmental carcinogenesis, where it has been demonstrated that exposures to known cancer agents early in life can cause specific types of cancer to emerge 10 - 60 years later, *without intervening exposure*<sup>17</sup>. The latency model has become closely associated with an eclectic mixture of research findings in diverse fields: a series of studies reputed to show that risks of cardiovascular mortality in the fifth decade of life are strongly associated with placental size and weight gain in the first year of life<sup>18,26</sup>; the putative effectiveness of certain early childhood intervention programs in changing the life trajectories of disadvantaged children<sup>27,29</sup>; and the demonstration of "critical periods" in brain development among primate species, wherein the developing organism is specially receptive to developing specific aspects of consciousness for comparatively short windows of time<sup>30</sup> but is comparatively refractory before or after.

The story of the handled rats gives the essence of this latter. Sapolsky<sup>9</sup> concludes that the age-specificity of the effect makes it a classic example of imprinting during a "critical period" in brain development. This example could set the reductionist imagination free to run wild. We could reason that the more important critical periods are, the more important latent effects are. Ultimately, the more important latent effects are, the more important targeted

strategies for child stimulation and social support are, at least during various, as yet undefined, critical periods in the development of the human brain. In other words, the experiential basis of socioeconomic gradients in health, well-being, and competence throughout the life cycle could be addressed through a series of magic-bullet-type social, emotional, and/or educational interventions: highly focussed in the developmental stage when they are applied and very specific in content.



### Pathways Model

The pathways model is most closely associated with a complex range of findings from longitudinal studies involving birth cohorts, adolescents, working populations, and the elderly. When the findings of these studies are overlapped in a temporal sequence to reconstruct the life cycle, a pattern emerges which highlights the enduring impact of socioeconomic status on health, well-being, and competence from cradle to grave. In highly simplified form, it goes something like this:

The socioeconomic status of one's parents helps situate the newborn in a specific community, which will exert a powerful influence on life chances in a complex, socioecological way<sup>31</sup>. Socioeconomic gradients are immediately felt with respect to infant mortality, which give way to differential levels of stability and security in early childhood, which, in turn, give way to differential readiness for schooling<sup>32</sup>. Differential school readiness precedes socioeconomic gradients in behavioural dysfunction in school<sup>33</sup> and ultimate school failure<sup>34</sup>. By early adulthood, as the nominal definition of socioeconomic status is transferred from ones parents to oneself, socioeconomic gradients begin to emerge in perception of well-being, which correlate with previous behavioural dysfunction and failure in school<sup>35</sup>. As adulthood unfolds, lower socioeconomic status becomes associated with the perception of high demands and low levels of control at work<sup>35</sup>. By the fifth decade of life, self-reported "high demand/low control" at work emerges first as a powerful predictor of disability and absenteeism<sup>36,37</sup>, and then as a predictor of premature mortality from most major causes of death<sup>38</sup>. Socioeconomic status gradients in mortality become most unequivocal at this time and persist at least into the eighth decade of life<sup>39</sup>.

## Latency, Pathways, and Worldviews

At first glance, it would appear commonsensical to view the latency and pathway models as complementary to one another. After all, there is no reason to suppose that latent factors only act independently, because any early life event which could exert a latent effect could also be the first step along a lifelong pathway which might have implications for health, well-being, or competence in the future. Conversely, any early childhood intervention designed to improve health and well-being in the long run will occur within a specific context which will provide a mixture of opportunities or barriers. Indeed, the closer the correlation between early life events and subsequent lifelong pathways, the more difficult the statistical problem of estimating the partial contribution of each.

Consider, for example, the findings of the longitudinal follow-up at age 19 of subjects from the Perry Preschool Study<sup>27</sup>, shown in Table 1. This study is significant because it was based upon one of the most comprehensive of the early intervention programs for children in American inner cities and because it was evaluated by comparing outcomes among children who were randomly assigned to "preschool" and "no preschool" groups. The data can be interpreted two ways. One interpretation would emphasize the remarkable improvements experienced by the preschool group in relation to the controls: higher rates of school graduation, higher rates of college attendance, fewer arrests, fewer teenage pregnancies, higher rates of employment, and lower proportions on welfare. The alternate viewpoint would emphasize the fact that most of the achievements of the preschool group, although impressive when compared with the controls, would not nearly match those of middle class children who had greater opportunities but no special preschool experience. After all, 31% were arrested or detained at least once, 18% were on welfare at the time of interview, and teenage pregnancy was still rampant among them.

*Table 1: Major findings at Age 19 in the Perry Preschool Study*

Category	Number * Responding	Preschool Group	No-Preschool Group	p
High school graduation(or its equivalent)	121	67%	49%	.034
College or vocational training	121	38%	21%	.029
Functional competence (average or above score)	109	61%	38%	.051
Ever classified as mentally retarded	112	15%	35%	.014
% of years in special education	112	16%	28%	.039
Ever detained or arrested	121	31%	51%	.021
Females only: teen pregnancies, per 100	49	64	117	.076
Employed at 19	121	50%	32%	.031
On welfare at interview	121	18%	32%	.044

Total n = 123

Source: Weikart, 1989, Table 1

On one level, this example illustrates the complementary nature of the power of latency, as demonstrated by the elements of effectiveness of the targeted "early life" intervention, and the power of the pathway, as demonstrated by the strict limitations to success which existed in the specific context of the study community. Yet despite the conceptual complementarity, the two interpretations presented above have a sharply conflicting emphasis. This paradox reveals an important aspect of the debate about latencies and pathways. At issue are deeply held beliefs about how human potential expresses itself, and, also, about the nature of the obligations which members of society have to each other. These deeply held beliefs are inextricably linked to individuals' choices regarding the facts which seem of greatest relevance to them, the approaches which they take to questions of causation, and the attractiveness which different policy responses hold for them. This is the substance of what Sylvia Tesh has described as "hidden arguments"<sup>40</sup>.

Evidence that hidden arguments are at work here is largely anecdotal, as it often is, since the phenomenon is commonly conveyed through the subjective responses of key opinion leaders to evidence which supports causal models they find threatening or objectionable. Such responses have been revealed in relation to the latency-based notions that the human brain, like other primates, may have critical developmental periods; and the evidence suggesting that placental size and weight gain in the first year of life can somehow pre-empt other socioeconomic influences on future health and well-being. The same phenomenon is no doubt at work among those who go to quite absurd lengths to deny the existence of socioeconomic status as a valid category for investigation and refuse to credit any analysis which addresses the determinants of socioeconomic gradients in health, well-being, or competence throughout the lifecycle.

### **The Way Forward**

Because of both the ideological and methodological barriers, the debate about latent effects versus pathways begins to look like a dead end. More information will not help, because in the near term there will likely be a deluge of new insights which will simultaneously strengthen the case for each model. What is needed is a pragmatic way out, because there are urgent questions whose answers depend upon a minimum level of common understanding, and will not await the outcome of the debate.

In a period of difficult economic transition, how can communities and societies protect and enhance the quality of the social environment in ways that promote health, well-being, and competence throughout the life cycle? Is it reasonable to believe that individuals can develop "coping skills," or that communities can develop "coping strategies" sufficient for them to overcome the population health implications of long-term unemployment? What might be the benefits of a social investment strategy that exclusively targeted certain transition points during the life cycle?

Put this way, the questions should serve to mobilize the rich knowledge base about the determinants of health, well-being, and competence throughout the life cycle, rather than dividing it. They build upon the principal point of consensus, which is the importance of the life cycle perspective. They are action oriented, and cannot be answered without social experimentation. Finally, they can accommodate small answers to big questions.

This latter may be the most important characteristic of all. When the urgent questions are conceived of as a choice between a cradle-to-grave social contract and an early childhood social investment strategy, we are quickly led back to the debate between latent effects and pathways and the hidden arguments underlying them. But the current knowledge base tends to support the pragmatic way out. It turns out that many of the most promising social interventions are rather modest in character: the value of positive mentors and good recreation

programs for the young, a few decent opportunities at a "second chance" to make a successful transition from childhood to adulthood, and a strong social support network to buffer the stresses of middle and late life.

## Endnotes

- 1 Canadian Council on Social Development, *Newsletter of the Centre for International Statistics on Economic and Social Welfare for Families and Children*, 1 (1992).
- 2 D. S. Lero, A. R. Pence, M. Shields, L. M. Brockman, and H. Goelman, *Introductory Report: Canadian National Child Care Study* (Ottawa: Statistics Canada, 1992).
- 3 D. S. Lero, H. Goelman, A. R. Pence, L. M. Brockman, and S. Nuttall, *Parental Work Patterns and Child Care Needs* (Ottawa: Statistics Canada, 1992).
- 4 This information was provided from a variety of sources, most recently by the British Columbia Ministry of Education.
- 5 Any analysis based on income levels, educational attainment, or occupational class is referred to here as socioeconomic status.
- 6 Developed countries refers to the member states of the Organization for Economic Co-operation and Development.
- 7 Clyde Hertzman, "Where are the Differences Which Make a Difference?" *CIAR Population Health Working Paper* 8 (1990).
- 8 W. Moyers, *Healing and the Mind* (New York: Doubleday, 1993).
- 9 R. M. Sapolsky, *Stress, the Aging Brain, and the Mechanisms of Neuron Death* (Cambridge: The MIT Press, 1992).
- 10 E. E. Werner, "Children of the Garden Island," *Scientific American* April (1989):106-111.
- 11 E. E. Werner and R. S. Smith. *Overcoming the Odds: High Risk Children from Birth to Adulthood* (Ithaca: Cornell University Press, 1992).
- 12 H. L. Needleman, D. Bellinger, "The Health Effects of Low Level Exposure to Lead," *Annual Review of Public Health* 12 (1991): 111-140.
- 13 D. Bellinger, A. Leviton, C. Waternaux, H. Needleman, and M. Rabinowitz, "Low - Level Lead Exposure, Social Class, and Infant Development," *Neurotoxicology and Teratology* 10 (1993): 497-503.
- 14 As recently as two years ago, reviewers for at least one relevant journal were treating such ideas as amusing fiction.

- 15 Robert Katzman, "Education and the Prevalence of Dementia and Alzheimer's Disease," *Neurology* 43 (1993): 13-20.
- 16 M. Meaney, D. Aitken, S. Bhatnager, C. van Berkel, and R. Sapolsky, "Effect of Neonatal Handling on Age-Related Impairments Associated with the Hippocampus," *Science* 239 (1988): 766.
- 17 This is a basic concept in occupational and environmental carcinogenesis which underlies an entire field of inquiry.
- 18 D. Barker and C. Osmond, "Infant Mortality, Childhood Nutrition, and Ischaemic Heart Disease in England and Wales," *The Lancet* (May 10, 1986): 1077-1081.
- 19 D. Barker, C. Asmond, J. Golding, D. Kuh, and M. Wadsworth, "Growth in Utero, Blood Pressure in Childhood and Adult Life, and Mortality from Cardiovascular Disease," *British Medical Journal* 298 (1989): 564-567.
- 20 D. Barker, C. Osmond, P. Winter, B. Margetts, and S. Simmonds, "Weight in Infancy and Death from Ischaemic Heart Disease," *The Lancet* (September 9, 1989): 577-580.
- 21 D. Barker, A. Bull, C. Osmond, and S. Simmonds, "Fetal and Placental Size and Risk of Hypertension in Adult Life," *British Medical Journal* 301 (1990): 259-262.
- 22 D. Barker, "The Intrauterine Environment and Adult Cardiovascular Disease," in *The Childhood Environment and Adult Disease*, Ciba Foundation Symposium 156, (Chichester: Wiley, 1991), 3-16.
- 23 D. Barker and C. Martyn, "The Maternal and Fetal Origins of Cardiovascular Disease," *Journal of Epidemiology & Community Health* 46 (1992): 8-11.
- 24 D. Barker, K. Godfrey, C. Osmond, and A. Bull, "The Relation of Fetal Length, Ponderal Index and Head Circumference to Blood Pressure and the Risk of Hypertension in Adult Life," *Pediatric and Perinatal Epidemiology* 6 (1992): 35-44.
- 25 D. Barker, T. Meade, C. Fall, A. Lee, C. Osmond, K. Phipps and Y. Stirling, "Relation of Fetal and Infant Growth to Plasma Fibrinogen and Factor VII Concentrations in Adult Life," *British Medical Journal* 304 (1992): 148-152.
- 26 D. Barker, C. Osmond, S. Simmonds and G. Wield, "The Relation of Small Head Circumference and Thinness at Birth to Death from Cardiovascular Disease in Adult Life," *British Medical Journal* 306 (1993): 422-426.
- 27 D. P. Weikart, "Early Childhood Education and Primary Prevention," *Prevention in Human Services* 6 (1989): 285-306.
- 28 F. H. Palmer, "Long-term Gains from Early Intervention: Findings from Longitudinal Studies," in E. Zigler and J. Valentine, eds., *Project Head Start: A Legacy of the War on Poverty* (New York: The Free Press, 1979).

- 29 S. L. Martin, C. T. Ramey and S. Ramey, "The Prevention of Intellectual Impairment in Children of Impoverished Families: Findings of a Randomized Trial of Educational Day Care," *American Journal of Public Health* 80 (1990): 844-847.
- 30 R. E. Kalil, "Synapse Formation in the Developing Brain," *Scientific American* (December 1989): 76-85.
- 31 M. Haan, G.A. Kaplan, and Terry Camacho, "Poverty and Health: Prospective Evidence from the Alameda County Study," *American Journal of Epidemiology* 125 (1987): 989-997.
- 32 R. Case and S. Griffin, "Rightstart: An Early Intervention Program for Insuring that Children's First Formal Learning of Arithmetic is Grounded in Their Intuitive Knowledge of Numbers," Report to the James S. McDonnell Foundation, 1991.
- 33 C. Power, O. Manor, and J. Fox, "*Health and Class: The Early Years*," (London: Chapman & Hall, 1991).
- 34 L. Pulkkinen and R. E. Tremblay, "Patterns of Boys' Social Adjustment in Two Cultures and at Different Ages: A Longitudinal Perspective," *International Journal of Behavioural Development* 15 (1992): 527-553.
- 35 R. Karasek and T. Theorell, "*Healthy Work: Stress, Productivity, and the Reconstruction of Working Life*," (New York: Basic Books, 1990).
- 36 M. Marmot, "Explaining Socioeconomic Differences in Sickness Absence: The Whitehall II Study," *Canadian Institute for Advanced Research* (Toronto: CIAR, 1993).
- 37 M. Marmot, G. Smith, S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner and A. Feeney, "Health Inequalities among British Civil Servants: the Whitehall II Study," *The Lancet* 337 (1991): 1387-1393.
- 38 M. Marmot, M. Kogevinas and M. Elston, "Social/economic Status and Disease," *Annual Review of Public Health* 8 (1987): 111-135.
- 39 M. Wolfson, G. Rowe, J. Gentleman and M. Tomiak, "Career Earnings and Death: A Longitudinal Analysis of Older Canadian Men," *CIAR Population Health Working Paper No. 12* (Toronto: CIAR, 1991).
- 40 S. N. Tesh, *Hidden Arguments* (New Brunswick: Rutgers University Press, 1988).

# Discussion of Papers by Len Syme and Clyde Hertzman: Towards a Sociocultural Epidemiology

---

*Ellen Corin*

Leonard Syme and Clyde Hertzman agree about two main challenges for contemporary epidemiology. Their first concern is that common epidemiological models cannot explain complex determinants of health and, more particularly, cannot account for the general gradient shape of the association between sociodemographic variables and various health problems. Since refinements to existing tools have produced a limited pay off, both authors call for a radical reframing of current paradigms to better illuminate the processes and chains of influence associated with health and disease. The second challenge derives from the growing use of epidemiological data for the rational bureaucratic planning of health services, a trend that has tremendously increased the power and responsibility associated with epidemiological research. Both authors note that the translation of epidemiological data into strategies of practice can no longer be considered unproblematic: on the one hand, the most carefully tailored programs have had quite limited efficacy; on the other hand, current programs presume a certain vision of reality or ideology that pre-selects the to-be-collected data, data that reinforce the vision of the reality underlying the preventive and curative practices.

This critical appraisal of contemporary challenges faced by epidemiologists is clarifying but the proposed solutions remain too timid.

With respect to the first challenge, one can deplore the authors' restrictive concept of the social influences on health and disease. Leonard Syme proposes to redirect the attention from individuals to social forces and the environment; Clyde Hertzman proposes to emphasize a long term developmental perspective and to reintegrate a time dimension into methods and analyses. However, both tend to rely on a simple notion of causality, to define processes in terms of variables and to take for granted the clear and unproblematic significance of key concepts.

The dramatic changes in the social fabric alluded to by Clyde Hertzman in his paper cannot be reduced to a simple, cumulative scheme of childhood experiences. For example, it appears difficult to account for the high suicide and school drop-out rates among the young, which characterize contemporary Quebec society, in terms of socialization practices without considering the parallel transformation in the modes of production, in the social organization of work, in values and in lifestyles, that frame a world without a perceived future for larger and larger groups of adolescents. The notion of competence, which both authors agree is an important element in their intellectual framework can not be reduced to either an individual feature developed through socialization practices or to discrete, potentially changing aspects, of a work situation. Both authors borrow arguments from studies in psychoneuroimmunology and social psychology to propose that the actors' perceived degree of control of their life could account for the gradient effect. This is interesting but if social process were overly psychologized, although they would be more amenable to empirical studies, the contextual analysis of social relations and structures would be lost. Individual competence cannot be dissociated from the actor's impression of being part of a larger frame where meaning can be assigned to things and where the person can locate him or herself. Additionally, groups and communities also devel-

op strategies for coping with difficult or unexpected situations; most of the time, individual and collective competence develop in a reciprocal relationship. In this perspective, the impact of "magic bullet" interventions would be of a limited value if the context of life remained unchanged in the long-term; it could even be counter-productive if it de-rooted individuals from their community framework or increased a greater feeling of incompetence in parents and communities.

The key challenge might be not to shift from variables measured at the individual level to variables measured at the social level, or to replace crude social variables by "thicker" social processes, but to examine how individual and social influences reinforce each other by an effect of "resonance" deriving from their convergence: how, for example, environmental or societal features reinforce the impact of problematic socialization practices or how these, in turn, reflect a more general feeling of hopelessness experienced by parents within the larger society. We need some kind of epidemiological imagination to understand how individual, relational and macrosocial forces echo and reinforce each other. We need to develop a theory of human agency that would consider what moves people to action in a given context.

If the notion of a gradient of social influences on health problems appears as an interesting starting point for elaborating new paradigms, one has to further understand what could be the significance of this notion in different sociocultural contexts. For example, one could hypothesize that the notions of "social hierarchy" or "social class" have a different significance in a hierarchical society where social differentiation is embedded within a complex set of reciprocal obligations and duties, in a highly individualistic society where people strive to achieve personal goals, or in a society that values the construction of a material and symbolic patrimony that is transmitted from generation to generation. As for the notion of competence, one could think that its meaning, its bearing in daily life and its markers differ in individualistic societies and in cultures where the person is essentially perceived as an element in a larger community frame where roles, status and expected actions are pre-defined by collective norms.

With regard to the second challenge which arises from recent developments in health policy, I would suggest that we need to rethink our own position in the development of knowledge and practice. Amazingly, the radical reframing called for by both Leonard Syme and Clyde Hertzman ends in ultimately reinforcing tremendously the power of experts. I cannot refrain from feeling uncomfortable when considering either Clyde's idea of "magic bullet" interventions targeted at social, emotional and educational life or Leonard Syme's idea of environmental interventions aimed at redressing people's behaviours. Advocating a pragmatic solution as Hertzman does is not sufficient. The limited efficacy of expert interventions, especially in underprivileged populations or in developing countries, can be considered as a stimulus for thinking. In the developing world, it is now largely recognized that strategies of intervention have to use and reinforce strengths and skills developed by the communities. In line with the general interest for competence claimed by both authors, it would appear important to pay attention to internal protective factors or to strategies developed by groups or communities to face problems in a given environment and to try to reinforce them through our actions. This means that attention should be devoted to characterize unexpectedly "well" sub-groups of people within larger vulnerable groups. A good example of this approach is offered by studies of the high rate of suicide among many Amerindian populations. Innovative programs have identified the features that characterize Amerindian communities with lower rates of suicide in order to design culturally and socially appropriate preventive strategies. To remain in the position of "those who know better" and know what to change and how to do it could be counter-productive to the intention of promoting control and competence in individuals and their communities. As we need a theory of human agency, we also need a theory of professional practice, one that questions the bias of our so-called neutral or better informed position: better informed, from which standpoints and at what costs!

# Discussion of Papers by Len Syme and Clyde Hertzman: The Social Environment and Health

---

*Akira Okamoto*

No one denies that the social environment affects the health situations of human beings nowadays. It is very difficult, however, to define and to classify what component of the social environment can be most or more specious as the determinant of health. The reason is that the social environment affects our health as the complex of its various components, such as social class, employment, income, education, housing, family, community, and so on.

Despite that the aspect of socioeconomic status is often adopted into the argument of health issues in western countries, it seems that there exists less significant difference in socioeconomic status, especially in annual income, among whole population in Japan than that of in western countries.

In Japan, the sociocultural aspect in life to life relations, life to non-life relations and even in non-life to non-life relations seems to be more considerable rather than the socioeconomic aspect.

Professor Noboru Kobayashi has advocated "Child Ecology" in such background in Japan. He illustrated a concentric circular diagram(Fig.1) in which he put a child and his family as a center of 4 ecosystems, namely micro-ecosystem, mini-ecosystem, meso-ecosystem and macro-ecosystem. These are not fixed. The alternation of child ecosystem by growth and development is shown in Fig.2.

In this child ecosystem, the interaction of the respective ecosystem should be investigated on the scientific basis. Through such investigations, the latent effect of the experiences in early childhood and the pathway effect along one's life stage can be clarified. From another view point, it must be possible that we can start from community in the survey of individual behaviors affected by social environment. Professor Syme mentioned that "the way we eat, drink, smoke, drive, sit, run and work can be seen as being our own affair and not anyone else's business." It is true that "these behaviors are private(individual) matter." I would like to emphasize, however, that these behaviors require places where they are performed. I would also like quote a sentence from "The Healthy City: Its Function and Its Future" by Professor Leonard J. Duhl. "Although there must be places to be born, to grow, learn, work, live, and die, there must also be places to play, or to 'just be', to waste time, to be with those we care about and who cares for us - to nurture the spirit and grow."

Where are places? They must exist in communities. I would like to conclude that the social environment as the determinant of health can work on individuals through the community, and through the ecosystem mentioned above.

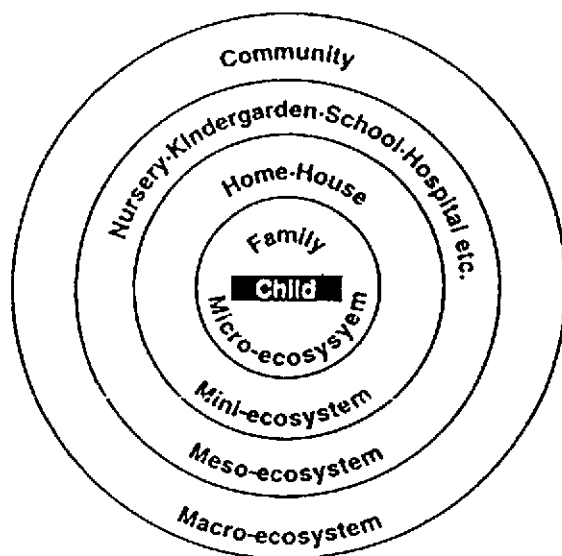


Fig. 1. Concentric circular structure of the ecosystem for child.

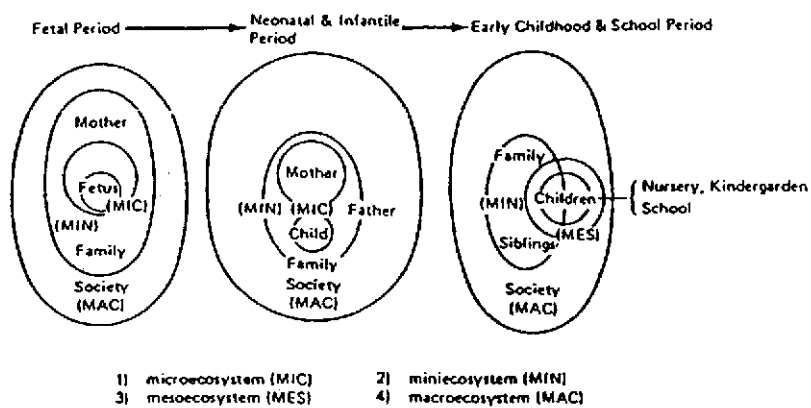


Fig. 2. Alteration in the pattern of child ecosystem by age.

# Discussion of Papers by Len Syme and Clyde Hertzman: Social Environment and Health

---

*Chris Power*

Many industrialised countries in recent decades have developed strategies to improve population health. So, using Britain as an example, strategies to reduce coronary heart disease focus on changes in risk factors, notably blood pressure, obesity, diet and smoking. As the papers here illustrate, such approaches are likely to have limited success. However, social influences tend to be excluded partly because they are not clearly defined or measured. This concern can lead us to emphasise less important but easily measured influences and thereby neglect those of greater significance.

Increasingly, however, the social environment is recognised as influencing health. This is suggested, as these papers argue, by the general pervasiveness of health gradients. Even so, the variability of gradients, over time and between locations, encourages optimism that health improvements can be achieved. For example, social differences in mortality appear to be smaller in the Netherlands, Denmark, Norway and Sweden, than in Finland and the UK, and those for France and Italy are even larger still. It is particularly notable that differentials are smallest in countries with well established social policies to improve living conditions amongst disadvantaged groups. Although the methodological problems in this work are daunting, there is scope for identifying relevant factors, including social policies, which impinge on health.

Studies in smaller communities use several concepts, identified by Syme (mastery, sense of control, learned helplessness etc) to investigate the role of social influences. Such concepts have been explored in relation to coronary heart disease amongst others, particularly in the workplace environment. However, it is not yet clear how demands of the workplace (and of other environments that are less well studied) interact with those environments that have already had an impact in early life.

Hertzman illustrates the impact of influences operating in childhood and thereby argues for a life course perspective in health research. This perspective has gathered momentum in Britain as three national birth cohort studies begin to be relevant to diseases in mid-life. Further impetus is provided by Barker's work linking circumstances in utero with adult diseases such as cardiovascular disease and diabetes. Hertzman outlines the difficulties involved in disentangling cumulative and latent effects. But such problems are not confined to longitudinal research as, for example, they are also apparent in studies of trends in disease over time.

Life course studies, collecting wide ranging information (on living and working circumstances, family structures and relationships, education, health etc) can establish the extent of clustering of influences over time. Mediating factors or 'stress buffering' effects of social structures can also be identified.

The British birth cohorts chart the effects, both short and longer term, of circumstances at birth and subsequently in childhood and adolescence. In the 1958 cohort there is clear social differentiation of health self-assessments at the relatively healthy age of 23. The social gradient is partly explained, statistically, by indicators of social adjustment earlier in life. Education is another major influence not just because of its link with subsequent social posi-

tion but also because of its relationship to ways of living. Thus social differentiation of health in early adulthood has much earlier origins.

Such evidence presses us to consider the role of social structures over the life course in efforts to improve population health.

# Biological Pathways that Contribute to Variations in Health Status: the Role of Early Experience

---

*Max Cynader*

**T**he title of these Proceedings is "Prosperity, Health, and Well-being", and within this volume is contained an important corpus of work regarding the determinants of health. One conclusion that seems to be certain from the large body of work presented in this volume is that happier people are healthier people, and that brighter and more successful people are also happier and healthier people. It is now well established that there is a strong correlation between success, intelligence, and long life and good health. As other chapters in this volume make clear (Hertzman, 1994), there is abundant evidence that the brain influences many other organ systems within the body, including the cardiovascular system (see Calaresu and McKittrick, 1991, for a review) and the immune system (see Sapolsky, 1992, for a review). The theme of this chapter is that an individual's probability of success, and indeed many of the abilities and attributes that an individual manifests, result from an interaction between heredity and environment. There is no question that one's abilities and talent are the result of one's genetic history as well as the exposure history that is unique to a particular individual. Studies have shown that identical twins, which share the same genetic makeup, share behavioral attributes as diverse as appreciation for particular forms of music, intellectual abilities, and even criminal tendencies. However, in addition to the undisputed role of genetic history in shaping our behavior, it appears clear that the exposure history of the organism plays a significant role. This presentation focuses on the role of the early environment in determining the later capabilities of the brain. Accumulating evidence shows that individual experience literally shapes the developing brain and our subsequent behaviour.

When we think about the development of the brain, a representation that is useful and evocative is that of Michelangelo's unfinished "Atlas", which is illustrated in Figure 1. The organism begins as a single cell, and there is then an extraordinarily complicated process of development in which cells divide repeatedly, migrate to their appropriate positions within the embryo, differentiate, and eventually produce the large volume of tissue that becomes the organism. Yet in parallel with this addition of neurons, neural pathways, and synapses (representing connections between cells) in the developing brain, there is an important process of **elimination**, or sculpting away, which occurs in parallel with these generative programs. An appropriate metaphor is that of an expanding ball which is being sculpted at the same time as it expands. Cell death plays a crucial and necessary role in neural development. Elimination of cells, connections between cells, and even of entire neural pathways, is an important feature of brain development. For example, more than half of the cells in the retina die before a child is one year of age. Over one-third of the neurons in the cerebral cortex are eliminated in the first three years of postnatal life (Meinecke and Rakic, 1992). In addition to a major expansion in synaptic connections that occurs during the first few years of postnatal life in the developing cortex, there is also massive elimination of synapses, connections, and even entire pathways. For instance, early in postnatal life there is a substantial pathway connect-

ing the visual and auditory cortex, which largely disappears later in postnatal life. What are the rules that govern the sculpting process? Stated in most general terms, neural development proceeds by **overproducing** neurons, connections, and pathways. Thereafter, development recapitulates evolution and only the fittest and most useful of these neurons, synapses and pathways are retained in the mature organism. It is important to understand that the needs of an individual neuron may not always coincide with the needs of the entire organism. Some neurons must die in order for the organism to survive. At a metaphorical level, one can think of each neuron as an individual, striving to grow, prosper, and accumulate as much synaptic territory and other neurons willing to listen to him, as possible. This type of biological aggression is valuable to neurons, because it enables them to survive in spite of the difficult developmental challenges that they face, and it seems intuitively reasonable that such a process should occur in a situation where many neurons will have to be eliminated.

Figure 2 tries to make some of the somewhat abstract notions discussed in the preceding paragraph more concrete, by way of an example. The figure shows a series of thin lines of different orientations. For the majority of viewers, the lines of different orientation all appear to be equally visible. Yet, for some individuals, the lines of a particular orientation are blurred and appear to be of lower contrast. The lower panels show how the lines of the top panel might appear to such individuals. This visual defect is frequently the result of uncorrected astigmatism. Astigmatism is a clinical condition in which an individual's cornea is shaped like a lemon rather than like an orange (Mitchell et al., 1973). Because the corneal curvature is unequal in the different directions, lines of different orientations cannot all be brought to focus at the same time. Such individuals cannot see lines of a particular orientation with clarity. If this condition occurs in adults, they eventually make their way to an Ophthalmologist or Optometrist, who prescribes a set of glasses in which the lenses are also lemon-shaped, but in the opposite direction to the natural astigmatism. The glasses then optically correct the corneal defect, and clear vision is restored. However, if the astigmatism occurs in very young children, then, unless the optical imbalance is recognized early and appropriate steps are taken, the process of Darwinian competition that was alluded to earlier takes place. We know from many studies of animal brains, that, within our cortex, we are born with a rudimentary set of cortical feature analyzers, each of which are involved with the analysis of lines of particular orientations. In the neonate, this initial set of rudimentary connections are set up independently of visual experience, and in normal development, each set of orientation-specific analyzers come to represent an equal area of cortex. Figure 3 is a view of the surface of the visual cortex, and illustrates with different shading the area of the cortical surface that is devoted to analyzing stimuli of different orientations. In the zone of cortex illustrated in Figure 3, there are roughly equal amounts of cortical territory devoted to analysis of each of the different ranges of stimulus orientation. However, with early astigmatism (like that described above) in which stimuli of one orientation are not seen as clearly as stimuli of other orientations, the number of neuronal analyzers for each orientation (which was initially equal) **changes**, and the amount of cortical territory devoted to the analysis of the orientation that was not seen well during early visual exposure shrinks substantially. Figure 4 (derived from Blakemore and Cooper, 1970) illustrates the incidence of cortical orientation-specific analyzers in two kittens which were reared under environmental conditions which allowed a preponderance of either horizontal, or vertical, stimuli to be seen during postnatal development. In the left panel of Figure 4, the kitten was reared in an environment containing preponderantly **horizontal** stimuli, and while the animal whose cortical orientation distribution is illustrated on the right was reared in an environment in which **vertical** stimuli predominated. In the figure, each line represents the responses of a single neuron within the visual cortex, and the stimulus orientation to which that individual neuron was most sensitive is represented by the orientation of the line in the figure. As can be seen, for the left-hand

panel, most cortical neurons that are encountered prefer stimuli that are oriented near horizontal, with virtually no neurons responding best to stimuli that are near vertical in orientation. Thus, in the kitten reared in a predominantly horizontal environment, the distribution of analyzers within the visual cortex devoted to analyzing horizontal, as opposed to vertical, has increased dramatically. Just the opposite situation obtains in the kitten whose data are illustrated in the right-hand side of Figure 4. Here, there is a strong preponderance of neurons which respond best to vertically-oriented stimuli, corresponding to the anisotropy in the post-natal visual exposure for this kitten. This example illustrates the principles alluded to above. If one does not get to use a set of analyzers within one's emerging brain early in life, then they will atrophy, and other competing analyzers will grow **at their expense**.

A condition such as early childhood astigmatism is frequently not recognized until a child goes to school and displays various difficulties in reading or other visually demanding tasks. At this stage, the child may be tested and fitted with glasses which provide a perfect compensation for the peripheral optical defect. **However, the ability to cure the peripheral defect, may not be sufficient to cure the defect in the central nervous system.** The use-dependent competitions that are described earlier depend on the exposure history of the organism, but only during a critical period early in postnatal development. During this critical period, the use to which a particular region of brain is put, determines how it will function for the rest of the organism's life. If curative measures are applied only after the critical period has passed, then, even though the optical defect may be cured, the pattern of feature analyzers that has been set up by the unusual visual exposure to which this individual's brain was subjected early in life will result in a permanent imbalance.

The scenario that is described above is potentially of great generality. The competitions that are set up may occur between visual inputs representing different stimulus orientations, or between visual pathways representing the two eyes within the cortex. Figure 5 shows a tangential (top) view of the surface of the visual cortex of the monkey (Hubel et al., 1977). Note the alternating stripes of light and dark. These stripes were visualized using a biochemical method to delineate the zones of cortex that respond to visual inputs represented by one eye or the other. In a normally-raised animal, such as the one illustrated in Figure 5, equal areas of cortex are devoted to each eye. In addition, if the two eyes are used together, extensive lateral connections that are initially present between each eye's territories are retained, and the capabilities of stereoscopic vision result. Figure 6 illustrates the binocular responses of a single cell within the visual cortex of a normally-reared animal. In this three-dimensional plot, one axis represents the spatial relationships between the two eyes, the second axis represents the timing relationships of stimulation to the two eyes, and the ordinate represents the response of the cortical cell. The spatial and temporal relationships between the two eyes are correlated with the depth of the visual stimulus and with its trajectory in three-dimensional space (Cynader et al., 1978; Cynader and Regan, 1978). The main point that the figure illustrates is that the responses of single cells within the cortex of normal animals are dramatically influenced by depth-related binocular inputs, and that individual cells may respond well to stimuli **only at a particular depth**. The situation illustrated in Figure 6, in which inputs from the two eyes are processed in a synergistic manner, occurs in normal animals. However, if one eye, for some reason, does not provide a visual image which is as effective as that of the other eye, or if the two eyes for some reason provide different and incongruent visual images at the same time, then the binocular synergy illustrated in Figure 6 does not develop. It should be emphasized that conditions under which the two eyes provide different images are not uncommon. If, for instance, a child develops strabismus, either convergent (cross-eyedness) or divergent (wall-eyedness), then the two eyes are looking in different directions and provide incongruent inputs to the cortex. Under these circumstances, the binocular synergy illustrated in Figure 6 does not develop properly. In other situations, the

inputs from the two eyes may be different, and the image conveyed via one eye may be less salient for the organism. Figure 7 shows an autoradiogram comparable to that of Figure 5, which illustrates the cortical surface in a monkey in which one eye was prevented from being used during development simply by patching the eyelid for the first few weeks of postnatal life (Hubel et al., 1977). In contrast to Figure 5 (in which inputs from each eye represent about 50% of the cortical territory), the vast majority of cortical territory in Figure 7 is occupied by the light-colored stripes, and the dark stripes are thin and broken up. The dark stripes represent the eye that could not be used for vision during early development, and the figure shows that this eye's share of cortical territory was reduced from about 50% to only about 15-20%. The influence of the other eye's inputs within the cortex has expanded correspondingly. The situation illustrated in Figure 7 occurs quite frequently in the general population, and appears to be part of the neural mechanism underlying Amblyopia or "lazy eye".

Like the problem of astigmatism, this disorder has a strong "critical period". If one eye fails to send the brain a useful message during early development, then its inputs are simply disconnected from the cortex. Extensive studies have delineated the timing of this critical period in both human and non-human species. Figure 8 shows that the eventual visual acuity that is attained via a disused eye depends on the age of the organism and on the duration of the period of disuse (Mitchell and Timney, 1984). By disuse we do not necessarily mean blindness; reduced vision can result from cataract, or refractive errors such as myopia (near-sightedness) or hypermetropia (farsightedness). In Figure 8, the ordinate represents the eventual visual acuity (on a log scale) that is attained. An acuity of 1 represents normal vision, whereas an acuity of 0.1 would represent 20/200 vision, and an acuity of 0.005 would represent virtual blindness or light perception only (denoted by LP at the bottom left of the figure). The horizontal black bars represent the duration of deprivation for individual subjects or groups of subjects. Thus, the subject illustrated on the top right of Figure 8 was prevented from using one eye for vision from 20 to 23 years of age. Despite this prolonged deprivation, visual performance through the deprived eye was essentially normal, once the peripheral defect was corrected. By contrast, individuals who were prevented from using one eye between 1 and 2 years of age, or just after 3 years of age, showed devastating effects on visual capabilities (reduction to light perception levels only) even after the peripheral defect was repaired. Overall, the figure makes clear that relatively short periods of visual deprivation imposed on one eye early in life can have devastating effects on future visual capacities. Equivalent deprivations imposed later in life have much weaker effects. The effectiveness of therapy is also directly related to the critical period. Therapy begun after the critical period has ended is virtually ineffective.

What can we do about the problems illustrated above? First, we need better diagnostic approaches to pick up these disorders early in life, while therapy is still possible. In addition neuroscientists worldwide are working to increase the plasticity and flexibility of the adult brain, to try to extend the critical period so that therapeutic approaches can be undertaken later in life. We have evidence that the critical period is not entirely fixed. It is in fact not really a period, rather it is a state of the brain. We know, for instance, that it is possible to extend the critical period in the visual cortex of animals well past its normal duration, simply by rearing animals in the dark. By **not using** the cortex throughout the naturally occurring critical period, it retains its capacity to be sculpted according to its eventual usage. Figure 9 illustrates the continuing plasticity of the visual cortex in animals which had been placed in a dark room at birth and then kept in total darkness throughout the naturally-occurring critical period (Cynader and Mitchell, 1980). These animals were thus prevented from using their visual cortex for vision during the normal critical period, and then after the normal chronological critical period was over, were allowed visual exposure through one eye only for a further 3-month period. The other eye was prevented from seeing by patching the eyelid. In a

normal animal of the same age, there would be no effect of disuse of one eye, but in these animals (which had been prevented from using the cortex throughout the critical period), the Darwinian competition that is described above occurs, and the eye that fails to provide useful information is ruthlessly suppressed within the cortex. In the figure, the ordinate represents the number of cortical cells encountered (and therefore the amount of cortical territory allocated to each eye). The abscissa is a 1-7 point scale showing the relative strength of inputs from each of the two eyes. In the animals illustrated in the top of the figure (kept in the dark for 6 months, 8 months, or 10 months), the deprived eye (which was not allowed vision after the animal was taken out of the dark room) controls relatively few cells, while the normal eye influences a far greater number of cells. The lower histograms show that no substantial effects are observed in control animals that were reared in the light for comparable periods of time instead of in the dark. In these cases, even if one eyelid is sutured, the deprived eye still retains its share of cortical territory.

The data of Figure 9 make clear that the critical period is not necessarily strictly linked to a particular chronological age. In this sense, it is not really a "period", but rather a **state** of the brain during which its long term organization can be shaped by the input that it receives. Let me not imply that not using one's brain is good for you. Even though the visual cortex of a dark-reared animal retains plasticity after the end of the naturally-occurring critical period, the functions of the cortex are compromised by this period of deprivation (Cynader et al., 1976). Nonetheless, the finding that the critical period can be extended suggests that the critical period itself is not simply a biological clock that is ticking away as the animal ages. Rather, the use to which the brain has been put plays a role in determining its biochemical and physiological status, and hence its ability to learn and modify its circuitry according to usage.

Intensive efforts are now underway in many neuroscience laboratories worldwide to try to understand what the essential nature of the brain state is that allows for learning and modifiability. New molecular biological techniques have identified a series of **genes that are turned on selectively** in cortical neurons at the height of the critical period, and which are then **turned off** as the critical period wanes (Prasad and Cynader, 1994). Some of these genes are summarized in Figure 10, which illustrates a prototypical neuron captured during its critical period. Each compartment of the neuron is "exploded" within one of the circles below the neuron, so that the neuron's cell body at the top left of the figure is expanded in the circle below it. The compartment shows a series of key molecules, including cell surface molecules such as contactin, TAPA-1, and the NK1 tachykinin receptors to be over-expressed in cortical neurons during the critical period. In adult animals, beyond the critical period, these genes are now **shut off**. Within the cell body, the nucleus and mitochondrion are then further exploded within the two circles at the bottom of Figure 10 to illustrate a series of key genes that are over-expressed in these structures only during the critical period. The neuronal terminal is exploded in the circle at the middle right of Figure 10, and illustrates a series of genes and gene products that are turned on selectively within this compartment of the neuron during the critical period.

Neuroscientists are trying to recreate the critical period by using gene therapy to try to turn on some of the genes that are shown in Figure 10, and are also trying to re-induce cortical plasticity by using therapies such as treatment with newly-discovered growth factors and other wonder molecules. In addition, behavioral strategies are being explored to try to extend the critical period into adulthood. We know that one way to extend the critical period is to not use a particular piece of brain at all, which is unfortunately not a very useful or practical approach. Another behavioral strategy that may be useful to try to achieve the same aim might be to **over-utilize** the plasticity mechanisms during the normal critical period within a particular piece of cortex. The idea is to subject a bit of cortex to continuous challenge dur-

ing the critical period. We have undertaken this by forcing a zone of cortex to receive dominant input from the left eye for a period in development, and then reversing inputs so that the right eye provides the dominant input for a subsequent period, then again left eye dominant, then right eye dominant, and so on. We do not yet have results from this experiment, but the notion is that by continuously exercising the plastic capabilities in a particular zone of cortex during the normally-occurring critical period, we may be able to prolong this plasticity. Just as exercising particular muscles strengthens them, so too **exercising the plasticity mechanisms** may in a similar way strengthen the capability for plasticity in situations in which it would normally decay. This idea may be of general importance for issues such as scholastic programs and for training of personnel in the workplace. Regardless of whether the eventual strategy for extending the critical period involves disuse, over-use, or a biochemical/molecular biological approach, it is clear that the ability to induce neural plasticity in a brain structure that would, under normal circumstances, no longer express it, would provide a potentially powerful tool to increase the effectiveness of virtually any training program that might be contemplated.

How general are these critical period effects? The examples presented thus far were selected from the sense of vision, but there is no doubt that similar use-dependent selection of particular pathways and neural circuits goes on in other parts of the brain. Examples abound in other sensory pathways including auditory research (hearing), and also in the senses of smell and touch (Brunjes, 1994). There is also strong evidence for critical periods in the development of higher cognitive functions such as language processing. For instance, the guttural sound "ch" is an important component of several languages, including Japanese, Spanish, and German, but it does not play an important role in English. During the "babbling phase" of language development in the first few years of life, all children make and use this guttural sound, but in children who grow up in an impoverished environment (learning only English), the ability to make this guttural sound is lost. At the crudest level, the expression "use it or lose it", while oversimplified, provides a strong metaphor to describe not only the use-dependent plasticity of vision, but also the use-dependent plasticity of the articulatory apparatus of language. There is abundant evidence that the best time to learn new languages is relatively early in life. Individuals who are exposed to a foreign language after they are eight or ten years old will eventually learn the foreign language, but they may never acquire the facility with this language that occurs naturally in young children. Abundant research shows that there is a time during early post-natal development when the brain is ready to accept language. The specifics of the language that is assimilated depends on the language (or languages) to which the developing brain is exposed. There is also evidence that other abilities which may be less obvious are regulated in development, and show a clear critical or sensitive period during development. These can include complex motor skills such as riding a bike, or skiing, or swimming. In addition, there is strong evidence that abstract intellectual abilities such as musical and mathematical capabilities are strongly age-related. By and large, if an individual mathematician has not shown flashes of great brilliance before the age of 25, then the probability is low that this individual will ever make important contributions to the discipline.

Importantly, it appears as though these critical periods that we have been studying do not occur in all parts of the developing cortex at the same time. Figure 11 illustrates a surface view of the cerebral cortex of the monkey. In the figure, the cerebral cortex, which is, after all, merely a crumpled sheet of tissue (about the size of a small coffee table in humans) surrounding the cerebral hemispheres, has been topologically unfolded, and the many different individual cortical areas that have recently been discovered (Knierim and Van Essen, 1992) are depicted. As mentioned, there is growing evidence that not all of these areas go through their critical periods at the same time, and in addition, there are suggestions that different

layers of the cortex may undergo critical periods at different times. Some of the cortical fields that are illustrated in Figure 11 show evidence of retaining plasticity even up to adulthood.

Some of the biochemical features that we have observed in the visual cortex (V1) at one particular age are in fact present in other cortical structures only at other ages. Neuroscientists are now trying to visualize the parts of the brain that are most plastic at any given age and in particular circumstances. It is easy to imagine the implications of this visualization capacity, because it offers the opportunity to direct particular training strategies to those parts of the brain that are most ready to learn. In conjunction with emerging technologies such as functional MRI, which enables us to **visualize** the activity of particular pieces of brain during specific tasks, it may become possible in the intermediate term to direct particular training and exercise to those parts of the brain that are most ready to accept it at any given age.

## STRESS AND CRITICAL PERIODS

Stress can take many forms. It may be physical, resulting from prolonged exposure to cold, or inescapable pain, or it may be mental resulting from having a supervisor that doesn't value an employee, a spouse who nags one, or the general feeling of powerlessness that comes from being unable to control one's own destiny. In the presence of short-term stressors, the body mounts an organized and normally effective response. A complex cascade of events, controlled by the brain, results in secretion of a series of releasing factors and hormones which increase the heart rate, alter the processing of glucose, dilate the pupils, and, stated most generally, prepare the body to meet that tiger lurking in the bushes, or the slings and arrows of everyday corporate life. We need the short-term stress response in order to catch that bus that is moving away from us and to evade the tiger. However, while there is no question the short-term response of the body to stress is beneficial and necessary, there is growing evidence that long-term stress and the body's response to it is counterproductive and deleterious. The deleterious affects are seen both inside and outside the brain. First, a large body of literature has now shown that chronic stress can cause depression of immune function, and also of other body systems that are controlled by the brain, such as the cardiovascular system. Second, chronic stress has a deleterious effect on brain function. There is evidence that repeated and prolonged exposure to stress hormones causes neurons to die. Subjecting animals to chronic stress (for instance by forcing them to endure inescapable shocks), results in cell death in various parts of the brain (Sapolsky, 1992). Interestingly, emerging evidence suggests that neurons in the brain that are **most ready to learn**, i.e., those that are within their critical periods, are the most **vulnerable** to the cell death and degeneration associated with chronic stress. As we know, the cortical plasticity of the critical period is due to a biochemical architecture within the cortex and to the expression of genes within individual neurons. This pattern of biochemistry and gene expression renders neurons ready to learn and alter their connectivity based on usage. Unfortunately it also appears to make them **more vulnerable** to the catabolic effects of stress hormones. This is an important point, because it suggests that the most plastic neurons will be those that will be most vulnerable to chronic stress.

In addition, stress and critical periods interact in another way. There is evidence (summarized in Hertzman, this volume) that the stresses to which we are exposed early in life, during a critical period, may modify our ability to **modulate and control** responses to stressors later on in life. There is evidence that rats that are subjected to mild stresses as neonates (for instance, by being handled repeatedly) show lesser, more controllable stress responses when tested as adults than do animals that have not been handled as infants. There thus appears to be a critical period for gaining effective neural control over the stress response.

## GENERAL CONCLUSIONS

The skills, abilities, dreams, and prejudices that make up each individual result from the genetic history of the individual and also from the particular environmental stimuli to which that organism was exposed. The feature analyzing and processing capabilities of the cortex are shaped by the input that it receives during particular critical periods early in development. Critical periods have been most extensively studied in sensory systems, but there is evidence for their role in higher brain functions as well. Even early **sensory** defects may cause devastating consequences for so-called higher functions. For instance if a child cannot see well early in life because of uncorrected astigmatism, then later on it is easy to imagine **reading difficulties** in that same child. A child who cannot read is disadvantaged in learning relative to other children in his peer group. This can lead to low self esteem and to a child who is turned off to school. Such a child is much more vulnerable to a large variety of poor outcomes including early pregnancy, low-wage employment, or unemployment, and therefore, as other chapters in this volume make clear, to poor health and enhanced mortality. Similarly early auditory defects caused by transient episodes of otitis media, may result in poor language acquisition. Poor verbal skills can lead to the same sort of deleterious cascade that is observed above. The best current strategies to prevent these deleterious cascades are to diagnose problems early and attempt therapy promptly. A more ambitious long-term goal is to recreate the plasticity of the critical period in the older organism, and then to try to cure long-standing deficits.

## REFERENCES

- Blakemore C, Cooper GF (1970) Development of the brain depends upon visual experience. *Nature* 228: 477-478.
- Brunjes PC (1994) Unilateral naris closure and olfactory system development. *Brain Research Rev.* 19: 146-160.
- Calaresu FR, McKittrick DJ (1991) Interaction of putative transmitters in central nervous pathways involved in the control of heart rate and arterial pressure. *Cardioscience* 2: 147-154.
- Cynader M, Berman N, Hein A (1976) Recovery of function in cat visual cortex following prolonged deprivation. *Exp. Brain Research* 25: 139-156.
- Cynader M, Gardner J, Douglas RM (1978) Neural mechanisms of stereoscopic depth perception in cat visual cortex. In S. Cool & E.L. Smith, III (Eds.), *Frontiers of Visual Science*, Springer-Verlag, pp. 373-390.
- Cynader M, Regan DM (1978) Neurones in cat parastriate cortex sensitive to direction of motion in 3-dimensional space. *J. Physiol.* 274: 549-569.
- Cynader M, Mitchell DE (1980) Prolonged sensitivity to monocular deprivation in dark-reared cats. *J. Neurophysiol.* 43: 1026-1040.
- Hertzman C (1994) The lifelong impact of childhood experiences: a population health perspective. *This volume*.

Hubel DH, Wiesel TN, LeVay S (1977) Plasticity of ocular dominance columns in monkey striate cortex. *Phil. Trans. Roy. Soc. (London)* B278: 377-409.

Knierim JJ, Van Essen DC (1992) Visual cortex: cartography, connectivity, and concurrent processing. *Curr. Opin. Neurobiol.* 2: 150-155.

Meinecke DL, Rakic P (1992) Expression of GABA and GABAA receptors by neurons of the subplate zone in developing primate occipital cortex: evidence for transient local circuits. *J. Comp. Neurol.* 317: 91-101.

Mitchell DE, Freeman RD, Millodot M, Haegerstrom G (1973) Meridional amblyopia: evidence for modification of the human visual system by early visual experience. *Vision Research* 13: 535-558.

Mitchell DE, Timney B (1984) *Handbook of Physiology - The Nervous System*. III. Post-natal development of function in the mammalian visual system. Chapter 12, pp. 507-555.

Prasad S, Cynader M (1994) Identification of cDNA clones expressed selectively during the critical period for visual cortex development by subtractive hybridization. *Molec. Brain Research* (in press).

Sapolsky RM (1992) *Stress, the aging brain, and the mechanisms of neuron death*. Cambridge, Mass: MIT Press.

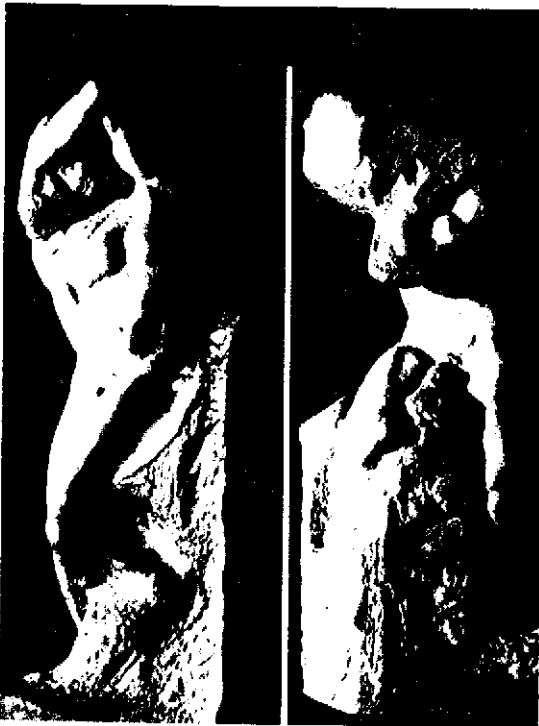


FIGURE 1

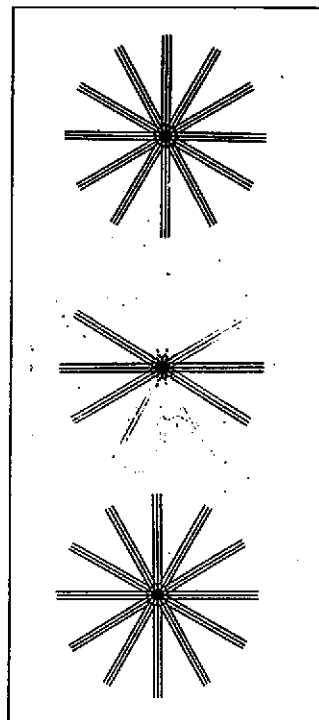


FIGURE 2

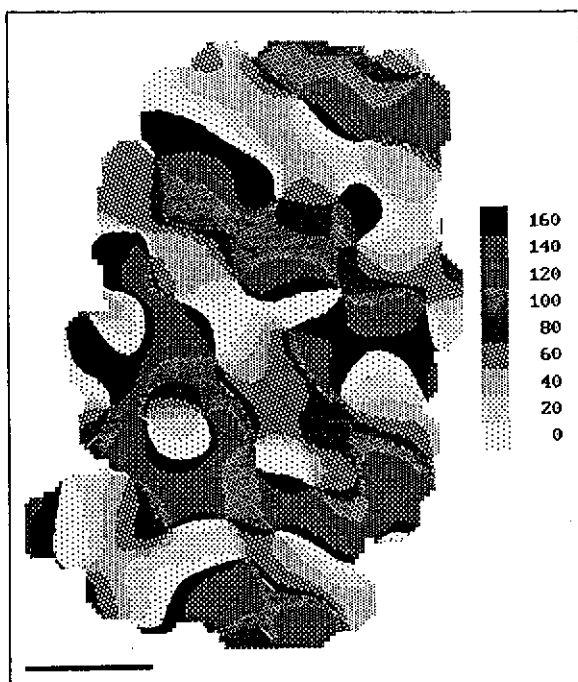


FIGURE 3

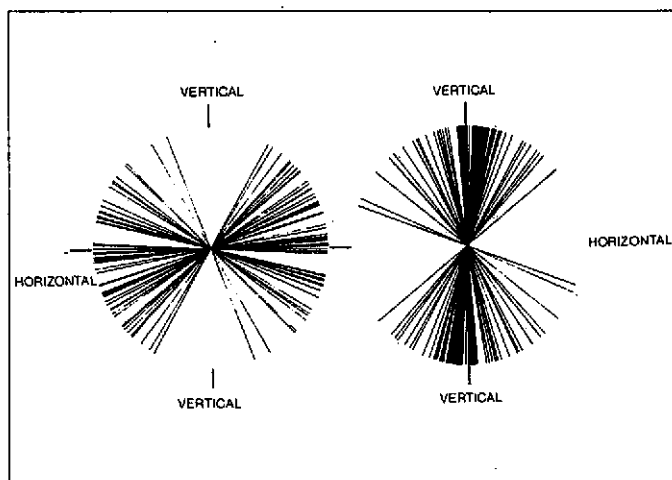


FIGURE 4

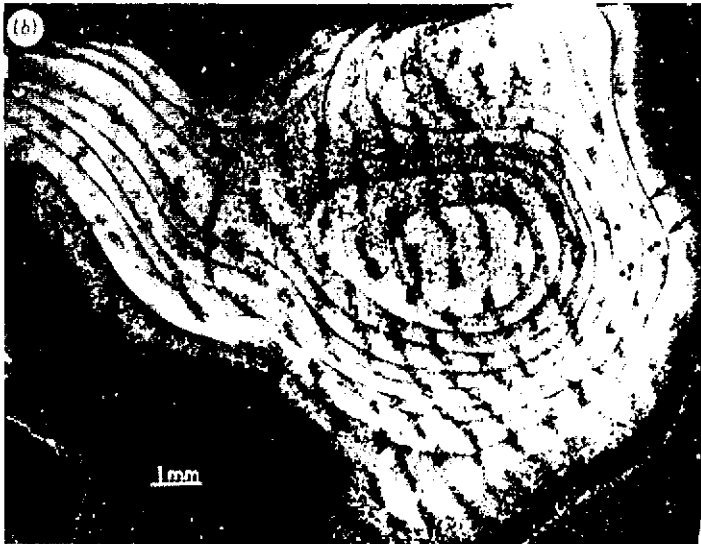


FIGURE 5

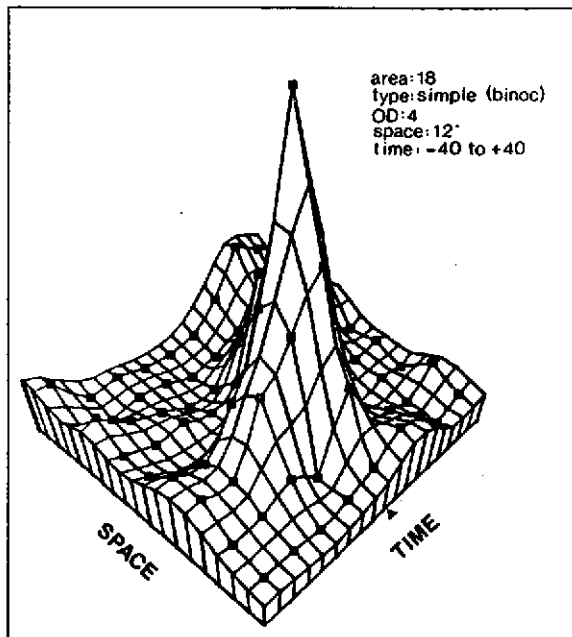


FIGURE 6

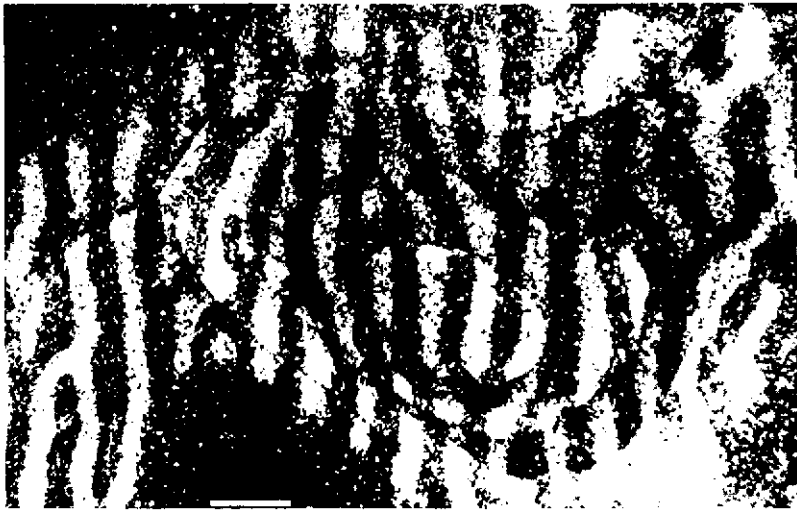


FIGURE 7

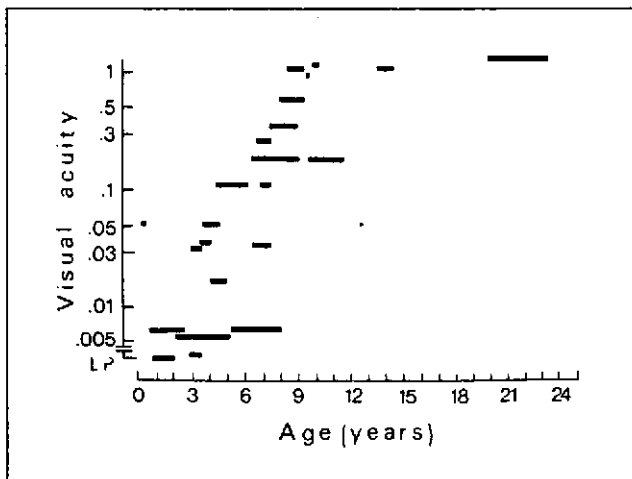


FIGURE 8

### THREE MONTHS MONOCULAR SUTURE

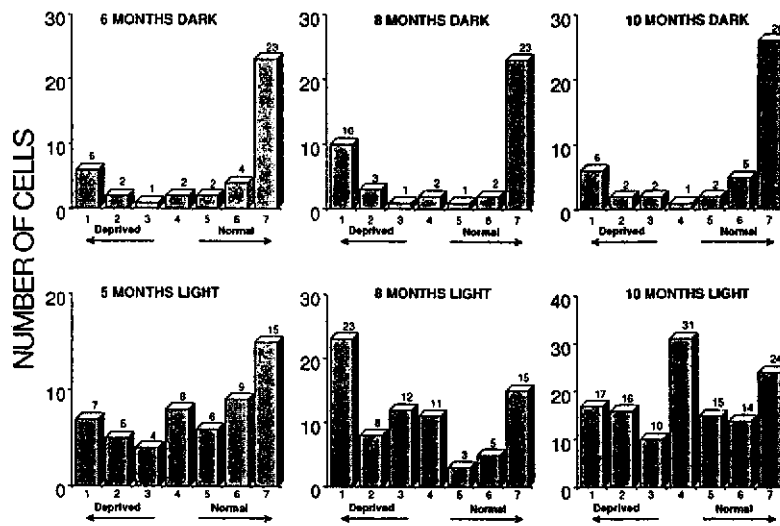


FIGURE 9

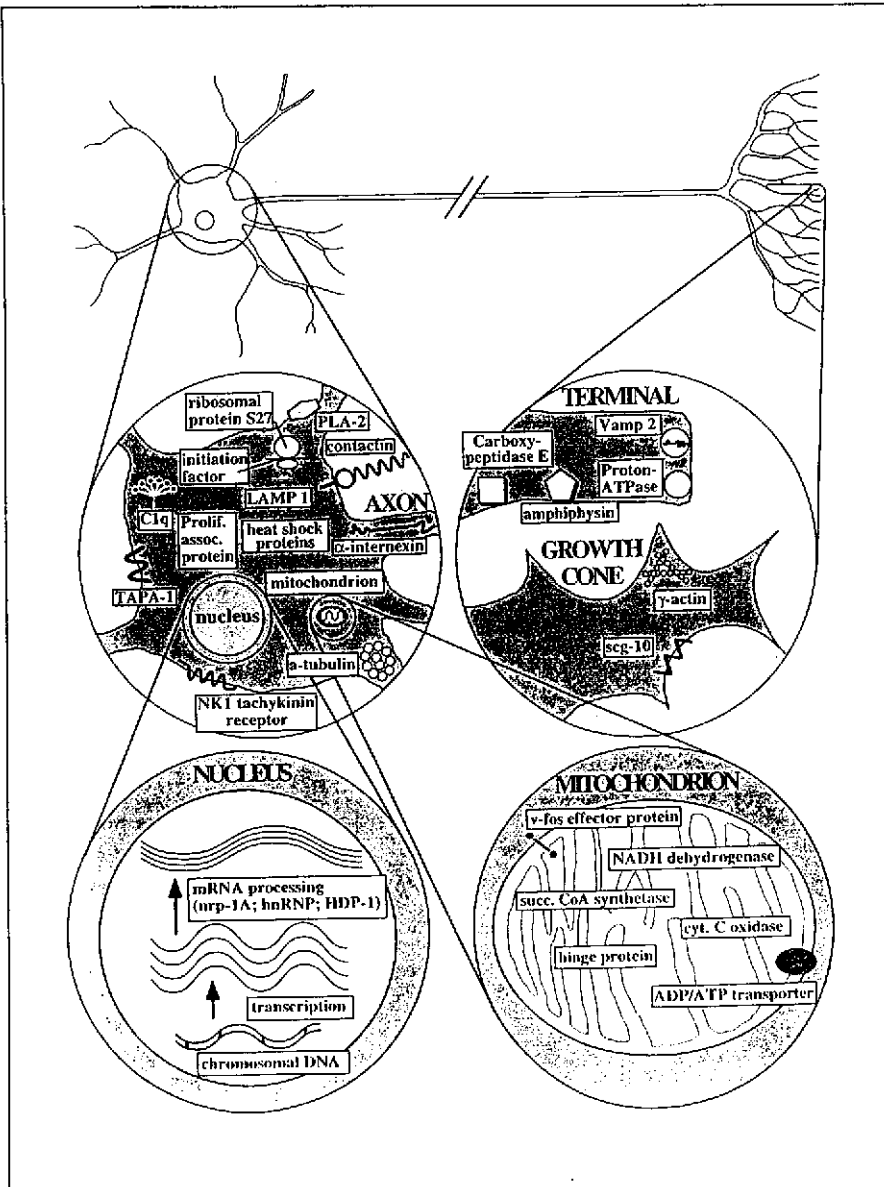


FIGURE 10



# Chromosome Alterations as Indices of Lifestyle

---

*Kanehisa Morimoto*

A great deal of epidemiological and clinical studies have shown that ways of daily living as well as exposure to environmental carcinogens determine the risks for cancer development, unspecific aging, and for genetic effects in the coming generations<sup>1,3</sup>. Chromosome alterations, which are directly-visible changes in the DNA, have close associations to those health risks<sup>4,6</sup>. Lymphocytes in the circulating blood are human cell-samples that can most easily be obtained, carrying information on both exposure doses and genetic effects of the mutagenic and carcinogenic factors. Thus, short-term cultures of phytohemagglutinin (PHA)-stimulated human lymphocyte are widely used to detect chromosome-damaging agents and possible human exposure to muta/carcinogens, and to study the immune response of blood<sup>7,9</sup>.

Among various types of chromosomal changes, chromosome-structural aberrations and sister-chromatid exchanges (SCEs) are widely-used indicators of chromosomal DNA damage; the former are efficiently induced by the chemical adducts and modifications of DNA (see Fig. 1) while the latter are produced directly by DNA-strand(s) breaks. Because the formation mechanisms of these two types of chromosome alterations are totally different, both endpoints are necessary to fully assess human genetic risks<sup>2,4,9</sup>.

## **1. Effects of Individual Lifestyles on Chromosome Alterations in the Peripheral-Blood Lymphocytes**

### **1.1 Effects of Total Lifestyles on Chromosome Alterations (SCEs)**

We have performed a series of experiments to see the correlations between those chromosome alterations (SCEs) in peripheral lymphocytes and individual lifestyles. One-hundred and twenty males aged 40 years or older who took multiphasic health check-ups at an industry were the subjects of study. Healthy lifestyles, or "good health practices" examined were: 1) not smoking, 2) not drinking too much alcohol, 3) doing physical exercise regularly, 4) sleeping more than 6 hours per night, 5) keeping nutritional balance in meals, 6) not snacking, 7) having breakfast everyday, and 8) not having too much perceived stress (see Table 1)<sup>10,11</sup>. The persons were categorized into 3 groups having good moderate and poor lifestyles by the number of good, health practices they do. Mean frequencies of base-line SCEs (the level of chromosome damage) in lymphocytes from men with poor lifestyles were shown to be 20 - 80% higher than those in cells from men having good lifestyles (data will appear elsewhere in detail).

### **1.2 Effects of Each Way of Daily-Living on Chromosome Alterations**

It has also been tested how large the contributions of each of daily lifestyles would be as a cause of the induction of chromosome alterations. Among 8 health practices in Fig. 1, cigarette smoking, alcohol drinking, and perceived mental stress were found to be three major factors for the induction of chromosome alterations. About two-thirds of the induced SCEs could be explained by exposures to these three lifestyles.

## 2. Lifestyles Determining Susceptibility to Mutagen Induction of SCEs

Experiments have also been conducted on samples from the same subjects to examine whether the lifestyles could have an effect on the sensitivities of peripheral-blood lymphocytes to the induction of chromosome damage by environmental mutagens/carcinogens exposure. For treatment, mitomycin-C (MMC), a potent carcinogen, was first dissolved in distilled water and diluted by phosphate-buffered saline (PBS). Aliquots of this freshly made solution were added to the SCE-observing culture to give a final concentration of  $3 \times 10^{-6}$  M. Cells were treated with MMC for entire culture, i.e., 72h. The results show that cells from male persons who have good lifestyles show lower SCE frequencies than cells from those who have poor (unhealthy) lifestyles. They suggest that lifestyles might influence chromosome susceptibility to environmental-carcinogens exposure.

## 3. Lifestyles Determining Sensitivity to Chemical Inhibition of Repair of Radiation-Induced Chromosome Damage

In our daily-living environments, we have been exposed to a mixture of suspected hazardous agents containing substances having activities to inhibit repair mechanisms of chromosome - DNA damage. To investigate the effect of lifestyles on repair kinetics of radiation - induced chromosomal damage, we have quantitatively examined the inhibitory effects of cytosine arabinoside (ara-C), a typical inhibitor of DNA damage, on the rejoining of chromosome breakage.

Those who were under chemotherapy or received radiation exposure were not included in our sample population. For irradiation, freshly drawn blood (0.3ml) mixed with 4.7 ml culture medium was irradiated with  $^{137}\text{Cs}$ - $\gamma$ -rays or tritiated water (HTO)- $\beta$ -ray at  $37^\circ\text{C}$  (total dose; 2 Gy, dose-rate; 0.5 Gy/h). For ara C-treatment, aliquots of ara-C solution dissolved in distilled water were added to the culture medium during irradiation to give a final concentration of 50  $\mu\text{M}$ . When the irradiation was finished, cells were washed with PBS three to five times and mixed with complete medium containing PHA to start the culture.

The enhancement ratio (E.R.) was calculated for each subject:

$$\text{E.R.} = (\text{D}+\text{R})_{\text{radiation+ara-C}} / ((\text{D}+\text{R})_{\text{radiation}} + (\text{D}+\text{R})_{\text{ara-C}}),$$

where (D+R) radiation+ara-C is D+R frequencies induced by irradiation and ara C-treatment, and so on. When mean ara C-enhancement ratio was calculated for three HPI categories, a clear, but not statistically significant, tendency was observed. Our results strongly suggest that it is necessary to take into consideration not only hereditary predisposition but comprehensive lifestyles to assess the repair inhibition in human population.

## 4. The Genetic Factor that can Determine the Drinking Behaviour and the Sensitivity to Chromosome-Damage Production

Drinking of alcohol is now widely known to be one of the major life-styles that could exert ill-health effects to healthy people. Alcohol is first metabolized to acetaldehyde which causes directly various types of chromosomal DNA lesions and alcohol-related diseases, and then further detoxified to much-less toxic metabolite, acetate<sup>12</sup>. Recent studies have revealed the existence of genetic deficiency in some enzyme species that can convert acetaldehyde to acetate in the human body. About 50% of Oriental people are deficient in the aldehyde-dehydrogenase 2 isozyme (ALDH2) that can most efficiently detoxify acetaldehyde<sup>12-14</sup>. It is therefore possible that cells in individuals that are genetically deficient in such an aldehyde dehy-

drogenase activity might show an elevated level of chromosome alterations in cases where they drank too much alcohol almost everyday.

We have performed a series of experiments to investigate now the genetic deficiency in ALDH2 affects the behavioural pattern for alcohol drinking and the sensitivity of peripheral lymphocytes to the induction of chromosome alterations by exposure to alcohol and alcohol-related chemicals. We here show 1) that lymphocytes from persons having unhealthy total lifestyles exhibit enhanced frequency of chromosome damage as manifested by an increase of SCEs and micronuclei, 2) significantly smaller proportions of individuals that were deficient in ALDH2 were habitual alcohol drinkers than those that were not deficient, and 3) lymphocytes from individuals that were missing in ALDH2 but still habitual alcohol drinkers, had significantly higher frequencies of SCEs than those even from ALDH2 proficient individuals that drank alcohol everyday.

#### 4.1 Variations in the Alcohol-Acetaldehyde Metabolism and Alcohol Drinking Behaviour

Individuals with different profiles of alcohol - and aldehyde-metabolizing enzymes also display different elimination rates ethanol and its toxic metabolite, acetaldehyde<sup>12</sup>. Differences in catalytic properties of polymorphic forms of alcohol and aldehyde dehydrogenase (ADH and ALDH) have been found to be responsible for the observed differences in alcohol elimination rates<sup>13</sup>.

The major pathway for the disposition of ethanol is its oxidation in the liver to two products, hydrogen and acetaldehyde, to which many of the effects of ethanol can be attributed. Three principal enzymes, cytosolic alcohol dehydrogenase, the microsomal ethanol-oxidation system (MEOS) located in the endoplasmic reticulum, and catalase located in peroxisomes, are known for the oxidation of ethanol.

Acetaldehyde is further oxidized to acetate, which is then converted to carbon dioxide via the citric acid cycle. Acetate may also undergo reactions to form fatty acids, ketone bodies, amino acids, and steroids via its activated form acetyl CoA. The enzymatic removal of aldehydes in humans is mediated by a number of unspecific enzymes like aldehyde oxidase, xanthine oxidase, and aldehyde dehydrogenase. However, the major oxidation of acetaldehyde in the liver and other organs is catalyzed by the NAD<sup>+</sup>-dependent aldehyde dehydrogenase (ALDH:aldehyde:NAD<sup>+</sup> Oxidoreductase, EC 1.2.1.3). The dehydrogenation of aldehydes into the corresponding tricarboxylic acids occurs as follows:



In recent years, a number of ALDH isozymes have been detected and characterized in human organs and tissues. At least five isozymes of ALDH coded by different gene loci have been detected in human organs and tissues which differ in their electrophoretic mobility, kinetic properties, as well as cellular and tissue distribution<sup>12</sup>. The various ALDH isozymes differ in their molecular size, subunit structure, and isoelectric point as well as in their chromosomal assignment. Besides, some of the isozymes show genetically determined structural variation.

#### 4.2 Determination of ALDH2 Deficiency by the Patch and "TAST" Tests

Rapid and simple screening method for detecting the ALDH2 deficiency is necessary for epidemiological studies in which a large number of subjects should be investigated. The patch test has been used widely for these purposes. Besides, we have developed a simple self-

administered questionnaire consisting of 13 short questions (TAST, see Table 2). We have used these 2 screening methods for detecting the ALDH2 deficiency in this study.

#### **4.3 Difference in Alcohol-Drinking Behaviour among ALDH2 Deficient and Non-deficient Individuals**

In our subjects that had kindly given an informed consent for cytogenetic investigation, ALDH2 deficiency testing and lifestyle-information gathering, about 50% (52 and 53% by the patch and TAST tests, respectively) were habitual alcohol-drinkers that drank averagedly 40 - 50 grams of net alcohol per day. Half of the subjects were found to be deficient in ALDH2 determined by the patch test (50%), or TAST screening (52%). We also observed a significant difference in the proportion of ALDH2 deficient persons between habitual and non-habitual alcohol-drinkers. Thirty-eight percent (by both tests) of the habitual drinkers were ALDH2 deficient while 64 or 68% were so among non-habitual drinkers. It is noteworthy to see that about 40% of ALDH2 deficient people that could not drink so much alcohol were screened as habitual drinkers.

We have also investigated such major confounding factors for these analyses as mean age of the subjects and the health practice index (HPI). No significant difference among these confounding factors was observed.

#### **5. Chromosome Alterations in Peripheral Lymphocytes from ALDH2 Deficient and Non-deficient Individuals**

##### **5.1 Baseline and Hydroquinone-induced SCEs as a Function of Daily Alcohol Consumption**

Based upon the lifestyle-information, we calculated net alcohol consumption per day separately in "Japanese sake", beer, whisky, and "Sochu", Japanese vodka. There exists a clear increasing tendency of both baseline and hydroquinone-induced SCE-frequencies with increasing daily consumption of alcohol. Such an increase in SCEs were more marked when plotted as a function of total net-consumption of alcohol.

##### **5.2 Baseline and hydroquinone-induced SCEs in lymphocytes from ALDH2 Deficient and Non-deficient Individuals**

When baseline and hydroquinone-induced frequencies of SCEs in peripheral lymphocytes were plotted as a function of daily consumption of alcohol, an increasing tendency of SCEs was more marked for ALDH2 deficient persons than for non-deficient ones. These differences were similarly observed in the sample data screened by the patch and TAST tests.

##### **5.3 Effects of ALDH2 Deficiency on Baseline and Hydroquinone-induced SCEs after controlling Smoking Habit**

We have already found habitual cigarette smoking to have several times larger effects than alcohol drinking on the induction of SCEs in peripheral lymphocytes. We thus analyzed the effect of alcohol drinking and ALDH2 deficiency in lymphocytes separately from cigarette smokers and non-smokers. The smoker's SCE data was corrected on the assumption that daily smoking of 20 cigarettes per day caused an increase of 1 SCE per cell according to our previous finding on the SCE frequency in lymphocytes from smokers and non-smokers populations<sup>3</sup>. It was generally confirmed that the effect of alcohol drinking on the SCE frequencies

was more proficient in ALDH2 deficient lymphocytes than in non-deficient ones even after controlling the effect of cigarette smoking. These general tendencies were similarly found both in sample subjects screened by the patch test and in those by TAST.

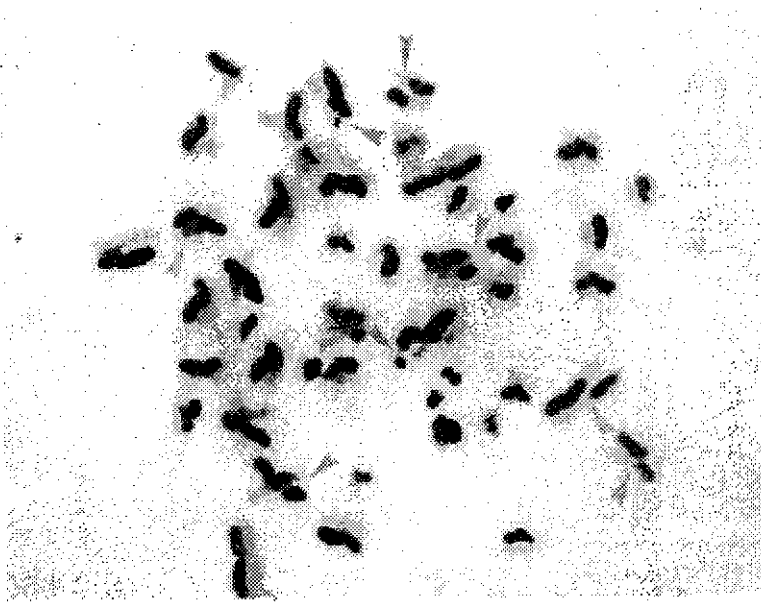
## Conclusion

I have proposed the importance of genetic health especially when we consider the future change of our health status. The level of chromosome damage in our peripheral-blood lymphocytes can be a good indicator of genetic health status because the peripheral blood is a sample that can be most easily obtained in the health check-ups, and because induced chromosome damage can be most efficiently detected in the peripheral lymphocytes. The lifestyles, or ways of daily living, were also shown to determine the rate of chromosome alterations in our body. These genetic changes might be an excellent index of the future status of our health as well as a risk indicator of the developments of cancers and circulatory diseases.

## References

- 1 Lisa F. Berkman and Lester Breslow, Health and Ways of Living (New York: Oxford University Press, 1983).
- 2 Kanehisa Morimoto, "Lifestyles and Health Situations: Importance of Genetic Health and Quality of Life; Position Paper for the Workshop on Social Learning and Coping Options," in WHO Document of the International Conference on Health Promotion (Copenhagen: WHO/EURO Office, 1986).
- 3 Kanehisa Morimoto, Kunihiko Miura, Tetsuya and Kaneko, "Health Situation and Chromosome Alterations," in Raymond Tice, Alexander Hollaender, Wo Lambert, and Kanehisa Morimoto, eds., Sister Chromatid Exchanges (New York: Plenum Press, 1984).
- 4 James Cairns, "The Origin of Human Cancers," Nature 289 (1981): 353-357.
- 5 George Klein, "The Role of Gene Dosage and Genetic Transpositions in Carcinogenesis," Nature 294 (1981): 313-318.
- 6 James German, "Oncogenic Implications of Chromosomal Instability," Hospital Practice (February 1973): 93-104.
- 7 Kanehisa Morimoto, and Sheldon Wolff, "Cell Cycle Kinetics in Human Lymphocytes Cultures," Nature 288 (1980): 604-606
- 8 Kanehisa Morimoto, and Akira Koizumi, "Proliferative Kinetics of Human Lymphocytes in Culture Measured by Sister Chromatid Differential Staining and Autoradiography," Experimental Cell Res 145 (1983): 349-356.
- 9 Kanehisa Morimoto, Kimio Kitamura, Haruo Kondoh, and Akira Koizumi, "Genotoxicity of Diesel Exhaust Emissions in a Battery of In-vitro Short-term and In-vivo Bioassays," in Nobru Ishinishi, and Akira Koizumi, eds., Carcinogenic and Mutagenic Effects of Diesel Engine Exhaust (Amsterdam: Elsevier, 1986).

10. Akihito Hagihara, Kanehisa Morimoto, "Personal Health Practices and Attitudes toward Non-smokers' Rights in Japan," Social Science and Medicine 33 (1991): 717-721.
11. Yukinori Kusaka, Hiroshi Kondoh, and Kanehisa Morimoto, "Healthy Lifestyles are Associated with Higher Natural Killer Cell Activity," Preventive Medicine 21 (1993): 602-615.
12. Dharam P. Agarwal, and H. Werner Goedde, Alcohol Metabolism, Alcohol Intolerance and Alcoholism (Berlin, Heidelberg, New York: Springer Verlag, 1990).
13. Shouji Harada, Dharam P. Agarwall, and H. Werner Goedde, "Aldehyde Dehydrogenase Polymorphism and Alcohol Metabolism in Alcoholics," Alcohol 2 (1985): 391-392.
14. Shouji Harada, Dharam P. Agarwall, and H. Werner Goedde, "Possible Protective Role against Alcoholism for Aldehyde Dehydrogenase Deficiency in Japan," Lancet ii (1982): 827.
15. Tatsuya Takeshita, Kanehisa Morimoto, Xiao-Quan Mao, Tomoko Hashimoto, and Junichi Furuyama, "Phenotypic Differences in Low Km Aldehyde Dehydrogenase in Japanese Workers," Lancet 341 (1993): 837-838.



**Fig. 1** Sister Chromatid Exchanges Observed in Human Lymphocytes

<b>Table 1:</b>
<b>Eight Health Practice Factors</b>
<ol style="list-style-type: none"> <li>1. Cigarette smoking (not smoking cigarettes)</li> <li>2. Consuming alcohol (not consuming alcohol every day)</li> <li>3. Eating breakfast (eating breakfast every morning)</li> <li>4. Hours of sleep (sleeping 7 to 8 hr per night)</li> <li>5. Hours of work (working less than 10 hr per day)</li> <li>6. Physical exercise (exercising at least once a week)</li> <li>7. Nutritional balance (eating a nutritionally balanced diet)</li> <li>8. Mental stress (keeping mental stress levels moderate)</li> </ol>
Note. The health practices recommended by Morimoto are given in parentheses.

**Table 2: ALDH2 phenotype screening test (TAST) used****Tokyo-University ALDH2-Phenotype Screening Test**

When you drink alcohol beverages, do you feel such subjective symptoms?  
Please select the appropriate answer for each item.

How to detect your ALDH2 phenotype.

- 1 Write down the coefficient you selected to each checking area.
- 2 You can get your TAST score by summing up all coefficients.
- 3 If your TAST score is...  
positive, then you are ALDH2-positive type.  
negative, then you are ALDH2-negative type.

Subjective Symptoms	always	sometimes	never	checking area
Facial flushing	<u>-10.04</u>	<u>5.22</u>	<u>8.95</u>	
Flushing elsewhere on the body	<u>-0.43</u>	<u>-2.98</u>	<u>1.20</u>	
Feel lchy	<u>3.37</u>	<u>-3.89</u>	<u>0.38</u>	
Become dizzy	<u>-0.58</u>	<u>-1.27</u>	<u>0.25</u>	
Become drowsy	<u>0.31</u>	<u>0.36</u>	<u>-1.03</u>	
Feel anxious	<u>0.00</u>	<u>-4.11</u>	<u>0.10</u>	
Get a headache	<u>-0.79</u>	<u>-0.07</u>	<u>0.01</u>	
Feel pounding in the head	<u>0.83</u>	<u>0.62</u>	<u>-0.24</u>	
Perspire	<u>-3.25</u>	<u>1.43</u>	<u>-0.44</u>	
Palpitations	<u>-1.88</u>	<u>0.04</u>	<u>0.26</u>	
Feel Nausea	<u>-10.07</u>	<u>0.19</u>	<u>0.03</u>	
Feel chill	<u>8.15</u>	<u>-2.42</u>	<u>0.14</u>	
Dyspnea	<u>-4.34</u>	<u>2.69</u>	<u>-0.19</u>	
Your TAST score:				

# Social and Biological Pathways that Contribute to Variations in Health Status: Evidence from Primate Studies

---

*Stephen J. Suomi*

## I. PRIMATE SOCIAL ORGANIZATION AND ONTOGENY

**H**uman beings, like all other primates, are inherently social animals, and much of what comprises the human experience involves social interactions, relationships, and hierarchical social organizations. Expressions of emotions, perceptions of the self, and the pursuit of goals rarely transpire in a total social vacuum. Even when individuals are physically alone, their thoughts and actions may be affected by social memories or social expectations. Interactions with others begin essentially at birth (if not prenatally), and virtually all individuals establish and maintain numerous social relationships in a variety of structured contexts throughout the whole of their lives. Thus, the influences of social factors on human activities are clearly ubiquitous and present from the very beginning — a true reflection of our primate heritage.

In point of fact, sociality is a fundamental characteristic of life in all primate species. It is the glue that bonds mothers and offspring to each other in long, complex relationships, and it provides the substrate for the extraordinarily complicated patterns of social interchange and organization that characterize most advanced nonhuman primate species (Novak & Suomi, 1991). The moment-to-moment social interactions, the day-to-day social relationships, and the long-term hierarchical social organization seen in every group of wild-living primates represent the product of countless years of evolutionary adaptation. The complexities inherent in these interactions, relationships, and hierarchical social group structures are remarkable, even by human standards.

Consider, for example, the case of rhesus monkeys (*Macaca mulatta*). In the wild, members of this highly successful species of Old World monkeys live in large social groups ("troops") that range in size from several dozen to over a hundred individuals. Rhesus monkey troops, be they large or small, all have the same basic social structure: each is organized around several multigenerational matriline (i.e., elderly females surrounded by two or three generations of their adult female offspring), along with unrelated males who have entered the troop as adults. This species-normative pattern of social group organization results from the fact that females stay in the troop in which they were born for their entire lives, whereas males remain in their natal troop only until puberty, at which point they typically leave to join all-male gangs (Lindburg, 1971). After spending several months (or longer) in an all-male gang, the young adult males move into different troops, where they typically take up long-term residence. Some of the young adult males will stay in their new troop for the rest of their lives, while others may subsequently change troops one or more times. However, no adult male is ever permitted to return to his natal troop once he has emigrated (Berard, 1989).

Rhesus monkey troops are also characterized by multiple dominance or status hierarchies. There is a hierarchy between the various multigenerational matriline, such that every member of the most dominant matriline (including infants and juveniles) outranks all members of the second most dominant matriline, who in turn outrank all members of the next most domi-

nant matriline, etc. There is also a hierarchy within every matriline, based on the general principle that younger daughters outrank their older sisters. In addition, there is a separate hierarchy among the adult males that have immigrated into the troop; the hierarchy is roughly based on relative tenure, in that the longer an individual male has been in the troop, the higher its ranking within the hierarchy is likely to be. Finally, there is a hierarchy among infants born within the troop — each infant, regardless of gender, shares its mother's relative rank, both within her own matriline as well as in the overall troop (Sade, 1967). Thus, a monkey's place within its troop is not determined so much by how big and strong it is but instead by who are its friends and relatives. These kinship and dominance relationships play major roles in organizing the activities of the troop as a whole, as well as influencing the interactions of individual monkeys with one another at any given moment.

Rhesus monkey infants spend their first weeks of life in physical contact with or in close proximity to their biological mothers, who provide them with nourishment, psychological warmth, and protection. In their second month of life the infants begin to leave their mothers for brief exploratory forays. By this time each infant has already established a strong attachment bond with its mother, and it uses her as a "secure base" from which to organize the exploration of its immediate environment. In the course of such exploration the infant encounters other members of its troop, and in the succeeding weeks and months it spends increasing amounts of time away from its mother engaging in extensive interactions with others, especially peers. In fact, from 6 months of age onward play with peers becomes the predominant social activity for young rhesus monkeys, and these play interactions become increasingly complex and involve specific sequences and patterns of behavior that appear to simulate virtually all adult social activities, including reproductive behavior and dominance/aggressive interactions. In other words, play with peers provides a medium through which specific behavior patterns crucial to normal adult functioning can be developed, practiced, and perfected long before they must become functional in adult life (Suomi & Harlow, 1975).

The onset of puberty in rhesus monkeys occurs near the end of the third year of life (for females) and the beginning of the fourth year (for males), and it is clearly associated with major life transitions for both genders. Although females remain in their natal troop throughout this period and thereafter, their interactions with peers decline dramatically, and they redirect much of their social activities toward other members of their own matriline, including the infants that they subsequently bear. Pubertal males, by contrast, leave their natal troop. The few individual males that refuse to leave voluntarily are eventually expelled physically, usually by members of other matriline within their natal troop. In either case, these young males typically join all-male gangs for varying periods before they attempt to enter a new troop. It should be noted that the time of adolescence and young adulthood is clearly the most dangerous period of life for a male rhesus monkey — in the wild, the mortality rate for males from the time they leave their natal troop until they are successfully integrated into a new troop approaches 50% (Dittus, 1979). It is also clear that individual males employ different "strategies" in their efforts to join a new troop, and that each strategy entails a different set of potential risks and benefits.

The above-described sequences of behavioral ontogeny and patterns of social troop organization have been observed not only in rhesus monkey troops living in the wild but also in groups maintained in captivity. These species-normative developmental sequences and group social structures are typical for many species of Old World monkeys, and they are similar in many respects to those seen in other nonhuman primate taxa (Lindburg, 1991). For any individual primate to survive — let alone thrive — in such a complicated social milieu it must not only develop a complex social behavioral repertoire but also acquire and retain detailed information about each of its fellow troop members, including that member's age,

gender, matriline, dominance status, and history of previous interactions. Failure to acquire and utilize such social information on a moment-to-moment basis will inevitably result in inappropriate behavior that the rest of the troop will not tolerate and ultimately will lead to that individual's expulsion and/or demise. Thus, complex social cognitive capabilities are prerequisites for successful social life in rhesus monkeys and other advanced primates. Indeed, some have argued that the large cortical regions that differentiate the brains of advanced primates from those of most other mammals have evolved primarily for coping with the myriad social-cognitive challenges posed by life in complex social groups (Humphreys, 1976; Seyfarth & Cheney, 1990).

## **II. INDIVIDUAL DIFFERENCES IN RHESUS MONKEY BIOBEHAVIORAL RESPONSES TO ENVIRONMENTAL NOVELTY AND STRESS**

While the same general developmental path has been followed from birth to maturity by countless generations of rhesus monkeys, there are numerous opportunities for any one monkey to take ontogenic excursions along the way. A major focus of recent research has involved characterizing differences between individual monkeys in how they follow this basic developmental trajectory. All rhesus monkeys begin life in constant physical contact with their mother, and all reduce the amount of maternal contact over their first half-year of life, but there are major differences between infants in the extent and timing of such contact reduction. Virtually all rhesus monkey juveniles establish playful relations with peers, but some initiate more play bouts and play longer and more vigorously than do others. The vast majority of rhesus monkey adolescent males emigrate from their natal troop and attempt to join new social units, but they vary from individual to individual in the precipitating factors and timing of emigration, as well as in the strategy they follow and the relative success they achieve in these endeavors. Thus, rhesus monkeys exhibit substantial differences in how they manage major species-normative developmental transitions. They also differ from one another in how they respond to environmental novelty and stress on a day-to-day basis.

A series of recent studies has disclosed that approximately 20% of rhesus monkeys studied in both laboratory and field settings consistently respond to novel and/or mildly challenging situations with extreme behavioral disruption and prolonged physiological arousal, compared with the rest of their respective social groups. Stimuli that typically generate interest and exploration in most monkeys instead elicit fearful and anxious-like behavioral responses, along with hypothalamic-pituitary-adrenal activation, sympathetic nervous system arousal, increased noradrenergic turnover, and selected immunosuppression, in these reactive individuals. In many ways these reactive rhesus monkeys resemble human infants and children classified as "behaviorally inhibited," not only in terms of their characteristic behavioral and physiological responses but also in their consistency in pattern of response across a variety of different novel and stressful situations throughout ontogeny (Suomi, 1991a).

A striking example of such developmental consistency in response to stressful events can be seen in the reactions of young rhesus monkeys to physical separation from their mothers and/or familiar conspecifics. Every rhesus monkey experiences many short-term maternal separations prior to puberty, but the nature and severity of response varies substantially across individual monkeys. For example, during the annual 2-month-long mating season, virtually all infants born in the previous year exhibit significant behavioral disruption when their mothers leave them repeatedly for 1-2 day periods to consort with different adult males. However, while most of these infants quickly get over their initial agitation and subsequently seek out the company of other troop members during their mothers' absence, about 20% of the infants instead become lethargic and withdraw from all social contact. This subgroup of infants – the very same individuals who display extreme wariness and anxious-like behavior

in the face of novel stimuli – remain behaviorally depressed until their mothers return (Berman, Rasmussen, & Suomi, in press). Laboratory simulations of these short-term maternal separations have shown that this depressive subgroup of monkey infants consistently exhibit significantly higher levels of plasma cortisol and ACTH, have higher and more stable heart rates, and display greater norepinephrine turnover and immunosuppression during periods of separation than their less reactive counterparts. Moreover, these differential patterns of behavioral, physiological, and immunological response remain remarkably stable across repeated separations throughout development. Monkeys who exhibit depressive reactions to separations as infants continue to show extreme behavioral and physiological responses to separation as juveniles and adolescents, while monkeys whose reactions to separation during infancy are mild are at very low risk for developing depressive reactions as juveniles or adolescents (Suomi, 1991b).

The same monkeys who display extreme behavioral and physiological reactivity to day-to-day environmental novelty and to challenges such as separation are also likely to encounter difficulties in adjusting to major species-normative developmental transitions. “High-reactive” monkeys are usually the last to leave their mothers as infants and the least likely to initiate play bouts with unfamiliar peers as juveniles. High reactive adolescent males are more likely to remain in their natal troop until they are forced out, rather than voluntarily leaving earlier; high reactive adolescent females are at the greatest risk to have difficulties rearing their first-born offspring. Thus, differences between monkeys in behavioral and physiological reactivity not only are consistent across a wide range of novel and challenging situations, but they are also remarkably stable throughout major developmental transitions. High reactive infants typically become high reactive juveniles, high reactive adolescents, and eventually high reactive adults. Given the impressive developmental stability of individual differences in behavioral and physiological reactivity, it is hardly surprising that an increasing body of evidence indicates that such differences are highly heritable (Suomi, 1991a,b).

Rhesus monkeys also differ dramatically from one another in their tendency to engage in impulsive, aggressive behavior. The appearance and socialization of aggression is an important species-normative feature of behavioral development in rhesus monkeys, as it is in all primate species. Aggression clearly serves multiple adaptive functions in habitats that contain not only predators but also other troops of monkeys who compete for resources necessary for survival; controlled aggression also sustains dominance hierarchies between and within matrilineal units within each monkey troop. However, uncontrolled and/or unpredictable aggression within a troop can drive its members apart and destroy it as a functional social unit. Therefore, aggression must be socialized – it must be minimized or at least largely ritualized in intragroup interactions, but it must also remain a viable response to counter external threats. Much of the socialization of aggression occurs in the context of play between peers, where young monkeys learn to moderate rough and tumble play patterns and to inhibit potentially explosive aggressive interchanges (Suomi, 1979).

However, there are marked individual differences in the degree to which such moderation is achieved throughout development. In particular, there are a few monkeys who as juveniles become involved in a disproportionate number of aggressive exchanges within their respective social groups. These individuals, mostly males, begin to exhibit inappropriate behavior in the latter half of their first year of life, at the time when most of their peers are becoming increasingly involved in complex play interactions. Their behavioral deficiencies stem not so much from an excess of aggressive behavior *per se*; instead, they appear to have a more general problem in inhibiting ongoing patterns of behavior as well as in responding appropriately to the behavioral initiations of other monkeys. In addition, these monkeys have apparent difficulties in perceiving the complex dominance hierarchies that characterize their social groups, often failing to display appropriate submissive behavior toward individuals that clear-

ly outrank them. As a consequence, they are frequent targets of aggressive attacks by other group members. Thus, these individuals, representing no more than 5-10% of their age group, are involved in a high proportion of the group's aggressive interactions, both as initiators and as targets (Higley, Linnoila, & Suomi, 1993).

The long-term prognosis for these socially inept young males is not good. During their juvenile years they become more and more isolated as peers avoid them and adults attack them with increasing frequency. They often are expelled from their natal troop long before the onset of puberty, and once they have emigrated are unlikely to be accepted in any other full-fledged troops, owing to their lack of basic social skills. Thus, these monkeys are likely to live a basically solitary existence for as long as they are able to survive.

Several recent studies have revealed that these socially inept monkeys tend to have unusually low cerebrospinal fluid concentrations of the serotonin metabolite 5-hydroxyindoleacetic acid (5-HIAA). Whether these unusually low 5-HIAA concentrations are in some way causing the inappropriate impulsive and aggressive behavioral patterns, whether they are a consequence of these behavioral patterns, or whether they merely represent some type of biological marker of a presently unknown etiology remains to be determined. What is known at present is that these uncharacteristically low levels of 5-HIAA can be detected as early as 6-7 months of age and appear to be remarkably stable throughout development, at least until early adulthood. Moreover, there is substantial evidence, albeit indirect, that differences between rhesus monkeys in CSF 5-HIAA concentrations are highly heritable (Higley et al., 1993).

In summary, while it is possible to describe a species-normative pattern of biobehavioral development for rhesus monkeys, there are marked differences between individuals in the details of their respective developmental trajectories. For example, approximately 20% of all rhesus monkeys consistently exhibit significant behavioral disruption and physiological arousal when confronted with novel or challenging situations, while approximately 5-10% appear to be impulsive and incompetent in their social interactions with peers and others. These behavioral tendencies appear relatively early in life (reactive individuals can be identified in the first weeks of life, whereas socially inept individuals can be identified by 6-7 months of age), they have relatively clear-cut physiological concomitants, they appear to be quite stable throughout much of development, and there is indirect evidence for a high degree of heritability. Most importantly, they have apparent differential long-term consequences for the relative health and well-being of the monkeys that display such tendencies (Suomi & Novak, 1991).

### III. LONG-TERM EFFECTS OF DIFFERENT EARLY SOCIAL ENVIRONMENTS

The relatively high heritability and impressive developmental stability of the individual differences in rhesus monkey stress reactivity and propensity for impulse aggression described above obtain only when the individuals being compared are reared and maintained in comparable physical and social environments. When monkeys who have been raised in different environments are compared with one another, these patterns are far less obvious. This is because developmental phenomena do not unfold in a vacuum but rather are subject to considerable influence by a host of environmental factors throughout ontogeny. In particular, individuals that grow up in environments which deviate significantly from the species norm follow developmental trajectories that also deviate from the species norm, regardless of the individuals' particular genetic pedigree.

For example, classic studies of early social isolation of monkeys carried out in the 1960's demonstrated profound and largely permanent social deficits and extreme behavioral abnormalities in individuals denied the opportunity to form attachments with conspecifics throughout the first 6 months of life (Harlow & Harlow, 1962). Subsequent studies demonstrated

that some recovery was possible in these monkeys via specially targeted therapeutic interventions, and that not all nonhuman primate species were as devastated by early social isolation as were rhesus monkeys. Nevertheless, it is clear that preventing the formation of early attachment relationships via isolation rearing severely compromises the formation of subsequent social relationships throughout development in rhesus monkeys and probably in most other primate species (Suomi, 1982).

More recent studies have examined the consequences of less extreme social deprivation experienced early in life. In these studies rhesus monkeys have been reared apart from their biological mother but in social settings where the formation of attachment bonds with other conspecifics, e.g., foster mothers or peers, is possible. In general, monkeys reared under such circumstances develop relatively normal behavioral repertoires and, in most cases do not display the extreme behavioral abnormalities characteristic of isolation-reared monkeys. However, the attachment relationships that they form with these other conspecifics are usually not entirely normal — indeed, they can be generalized as clearly “insecure” — and long-term problems often appear throughout development in these individuals.

For example, several recent studies have demonstrated that rhesus monkeys reared for the first 6 months of life in the company of same-age peers, but no adults, clearly develop strong attachment bonds to these peers. However, because peers are not nearly as effective as a typical monkey mother in providing a secure base for exploration or in reducing fear under conditions of novelty or stress, the attachment relationships formed with these peers are “anxious” in nature. As a consequence, while the peer-reared monkeys exhibit species normative motor development and complex social repertoires, they appear to be timid in the face of novelty, they are less likely to initiate play bouts with strangers than are their mother-reared counterparts, and they tend to be lower in social dominance hierarchies when grouped with mother-reared age-mates. Nevertheless, in familiar physical and social settings most peer-reared rhesus monkeys are largely indistinguishable from their mother-reared counterparts in most aspects of their social and nonsocial behavior (Suomi, 1991c).

In contrast, peer-reared rhesus monkeys generally display much more severe biobehavioral reactions to social separations than do mother-reared monkeys. As infants and juveniles, peer-reared monkeys exhibit higher levels of self-directed behavioral withdrawal, greater and more prolonged elevation of plasma cortisol and ACTH, higher concentrations of CSF MHPG, and greater suppression of a variety of immunological functions (Higley, Linnoila, & Suomi, 1993). These behavioral, physiological, and immunological differences persist throughout development, even when the peer-reared and mother-reared monkeys are living in the same social groups.

Follow-up studies comparing peer-reared with mother-reared monkeys later in life have demonstrated other long-term deficits resulting from peer-rearing during the first 6 months of life. For example, peer-reared adolescents display higher levels of aggression toward strangers, along with lower CSF levels of the serotonin metabolite 5-HIAA, than mother-reared counterparts. Peer-reared adolescents are also more likely to be expelled from their social group. Finally, peer-reared mothers are more likely to show inadequate care of their first-born offspring than are their mother-reared counterparts. Thus, although peer-reared monkeys develop sophisticated social repertoires and in most circumstances appear to be highly similar to mother-reared monkeys in most aspects of their behavior, they show clear-cut deficiencies in coping with novelty and stress throughout development (Suomi, 1991c). Put another way, peer-rearing appears to have the general effect of making individuals more stress-reactive, as well as increasing their likelihood of exhibiting inappropriate patterns of impulsive aggression as they approach puberty. Whether these general effects are more pronounced as a function of an individual's genetic predisposition regarding stress reactivity and propensity for impulsive aggressiveness is currently a topic of intense study in the laboratory.

Other recent studies have examined the consequences of foster-rearing rhesus monkey infants with unrelated multiparous females who differ systematically in their characteristic maternal "styles." Foster-reared infants readily develop attachment relationships with these foster mothers, and in most cases they develop species-normative social and nonsocial behavioral repertoires. However, different matches between infant temperament and characteristic foster-mother maternal "style" yield differential short- and long-term developmental outcomes. Of special interest is the finding that infants with "behaviorally inhibited" or physiologically "reactive" temperaments who are reared by unusually "nurturant" foster mothers develop "secure" early attachments and appear to have optimal long-term developmental outcomes. In contrast, infants of similar temperaments who are reared by more punitive foster mothers develop "insecure" attachments and subsequently are likely to display extreme reactions to environmental novelty and stress throughout development (Suomi, 1987).

These and other findings demonstrate that different patterns of early attachment are associated with predictable long-term developmental outcomes. In most cases the long-term outcomes are best predicted by considering both infant temperament and maternal style factors, as well as how they interact through the attachment process.

## REFERENCES

- Berard, J. (1989). Male life histories. Puerto Rico Health Sciences Journal, 8:47-58.
- Berman, C. M., Rasmussen, K. L. R., & Suomi, S. J. (in press). Responses of free-ranging rhesus monkeys to a natural form of social separation: I. Parallels with mother-infant separation in captivity. Child Development.
- Dittus, W.P.J. (1979). The Evolution of Behaviours Regulating Density and Age Specific Sex Ratios in a Primate Population. Behaviour, 69:265-302.
- Harlow, H. F., & Harlow, M. K. (1962). Social deprivation in monkeys. Scientific American, 107:136-146.
- Higley, J. D., Linnoila, M., & Suomi, S. J. (1993). Ethological contributions. In: Hersen, M., Ammerman, R. T., & Sissin, L. A. (Eds.), Handbook of aggressive and destructive behavior in psychiatric patients. New York: Plenum.
- Higley, J. D., Thompson, W. T., Champoux, M., Goldman, D., Hassert, M. F., Kraemer, G. W., Scanlan, J. M., Suomi, S. J., & Linnoila, M. (1993). Paternal and maternal genetic and environmental contributions of CSF monoamine metabolites in rhesus monkeys (Macaca mulatta). Archives of General Psychiatry, 50:615-623.
- Humphreys, N. K. (1976). The social function of intellect. In: Bateson, P. P. G. & Hinde, R. A. (Eds.), Growing points in ethology. Cambridge: Cambridge University Press.
- Lindburg, D. G. (1971). The rhesus monkey in North India: An ecological and behavioral study. In: Rosenblum, L. A. (Ed.), Primate behavior: Developments in field and laboratory research (Vol. 2). New York: Academic Press.

- Lindburg, D. G. (1991). Ecological requirements of macaques. Laboratory Animal Science, 41:315-322.
- Novak, M. A., & Suomi, S. J. (1991). Social interaction in nonhuman primates: An underlying theme for primate research. Laboratory Animal Science, 41:308-314.
- Sade, D. S. (1967). Determinants of social dominance in a group of free-ranging rhesus monkeys. In Altmann, S. A. (Ed.), Social communication among primates. Chicago: University of Chicago Press.
- Seyfarth, R. M., & Cheney, D. L. (1990). How monkeys see the world. Chicago: University of Chicago Press.
- Suomi, S. J. (1979). Peers, play, and primat prevention in primates. In Kent, M. W., & Rolf, J. E. (Eds.), Primary prevention in psychopathology (Vol. 3): Social competence in children. Hanover, NH: University Press of New England.
- Suomi, S. J. (1982). Abnormal behavior and primate models of psychopathology. In Fobes, J. L., & King, J. E. (Eds.), Primate behavior. New York: Academic Press.
- Suomi, S. J. (1987). Genetic and environmental contributions to individual differences in rhesus monkey biobehavioral development. In: Krasnagor, N., Blass, E., Hofer, M., & Smotherman, W. (Eds.), Perinatal development: A psychobiological perspective. New York: Academic Press.
- Suomi, S. J. (1991a). Up-tight and laid-back monkeys: Individual differences in the response to social challenges. In: Brauth, S., Hall, W., & Dooling, R. (Eds.), Plasticity of development. Cambridge, MA: MIT Press.
- Suomi, S. J. (1991b). Adolescent depression and depressive symptoms: Insights from longitudinal studies with rhesus monkeys. Journal of Youth and Adolescence, 20:273-287.
- Suomi, S. J. (1991c). Early stress and adult emotional reactivity in rhesus monkeys. In: Barker, D. (Ed.), The childhood environment and adult disease (CIBA Foundation Symposium 156). Chichester: Wiley.
- Suomi, S. J., & Harlow, H. F. (1975). The role and reason of peer friendships. In: Lewis, M., & Rosenblum, L. A. (Eds.), Friendship and peer relations. New York: Basic Books.
- Suomi, S. J., & Novak, M. A. (1991). The role of individual differences in psychological well-being in rhesus monkeys. In: Novak, M. A., & Petto, A. (Eds.), Through the looking glass: Psychological well-being in laboratory primates. Washington, DC: American Psychological Association Press.

# Discussion of Papers by Max Cynader, Kanehisa Morimoto and Steve Suomi

---

*Dan Offord*

## Introduction

I am going to begin in a somewhat unusual way by talking first about a childhood mental disorder. Then, I will return to the presentations and address the question: In what ways can the work described in these three presentations be helpful now and in the future in addressing important problems or conditions that lower life quality and reduce the life chances of large groups of children?

The problem I have chosen is conduct disorder or antisocial behavior. It is the leading mental health problem in children and adolescents in Canada and the United States. Conduct disorder refers to children who chronically break rules (eg., steal, run away, are truant), or who are chronically aggressive (eg., engage in physical fights). It has a heavy burden of suffering in terms of its frequency, morbidity in childhood and beyond, and the immense amount of resources spent on assessing and managing children with this disorder. Although there are a good deal of data on the associated features of the disorder in the child, family and community domains, less is known about variables that qualify as true risk and protective factors. Lastly, there is a paucity of effective treatment and prevention programs for this condition.

Now, to return to the central question: How might the findings in these papers help in lowering the burden of suffering from antisocial behavior? Four themes should be emphasized.

### 1) The importance of individual differences

Dr. Suomi's results indicate that Rhesus monkeys differ dramatically from one another in their tendency to engage in impulsive, aggressive behavior. Most monkeys learn to inhibit their antisocial behavior primarily in the play situation with their peers. Some do not, and although this subgroup is no more than 5-10% of their age group, they are involved in a high proportion of the group's aggressive behavior.

Individual differences are of central importance in not only investigating the etiology of antisocial behavior but in monitoring the response to interventions.

### 2) Early social environments can have serious longterm effects

Dr. Cynader's work indicates the pervasive effects of early environments through altered brain functioning. Further, therapy that is begun after a critical period in development may be ineffective in remediating a deficit. Dr. Suomi's work emphasizes that early learning patterns can result in problems appearing throughout development.

The challenge at the human level in advancing our understanding of the genesis of antisocial behavior is to discover how to conceptualize accurately different early environments, and measure those environments with psychometrically sound instruments.

### **3) Observed behavioral differences have biological underpinnings**

Dr. Suomi's aggressive young male monkeys tend to have an unusually low cerebrospinal fluid concentrations of the serotonin metabolite 5-HIAA. Dr. Morimoto reported correlations between chromosomal abnormalities and individual life styles.

In both these works, cause and effect relationships cannot be established, but they are examples of the association of biological and behavioral abnormalities. Etiologic studies of antisocial behavior have focused for the most part on environmental variables. Studies are now needed which combine epidemiologic and biological approaches.

### **4) The importance of pathways**

The data from, Drs. Cynader and Suomi emphasize the central significance of pathways in understanding both normal and abnormal development. It will not be enough to identify risk and protective factors in antisocial behavior, but knowledge is needed about their mechanisms of action not only at one point in time but across the developmental years.

### **Summary**

The results of these three papers emphasize the importance of individual differences, early social environments, the biological correlates of behavioral patterns, and mechanisms of action and pathways through which risk and protective factors operate in understanding development. The application of these themes to the study of antisocial behavior will enrich both etiologic and intervention research.

Lastly, I hope the discussion makes clear the need to continually evaluate the relevance of findings from basic work in development to real world problems of children and adolescents. A corollary is that structures must be in place to promote collaboration between groups carrying out these diverse groups of studies.

# Introductory Comments, Session 4: "Health Care as a Determinant of Health"

---

John Frank

The conceptual framework put forward by the Population Health Program of the Canadian Institute for Advanced Research has, as one of its central themes, the notion that the effects of health care on the health status of human populations are inherently limited. As many of you will be aware, from the proceedings of this meeting thus far, and from documents produced by the Institute's Population Health Program, there is a particular purpose to this sub-theme in the conceptual framework: it is set out in order to counter a "conventional wisdom" which is often met in health policy circles - namely, that further expenditures on health care can be expected to produce commensurate reductions in morbidity and mortality. It should be noted, therefore that the context in which this point is made has very much to do with the current and future state of medical care, often referred to as "flat of the curve" medicine. By this is meant very simply that substantial additional improvements in population health status in developed countries would seem, *a priori*, to be unlikely products of significant increases in current health care "investment" - i.e. expenditure.

Our two speakers in this session will approach this theme from two rather different vantage points. Both seek to test empirically the notion that health care can have significant impacts on the health of populations, albeit in very different time-frames and somewhat different settings. In reexamining these issues, it is important to acknowledge that one need not throw out the baby with the bath water, in critiquing further major investments in health care as a means to improving population health status. Specifically, it does not necessarily follow, and Population Health Program members whom I know would not necessarily suggest, that medical care of certain sorts has not made a significant difference to current health status, particularly since the advent of more efficacious medical and surgical therapies in the last half century or so. Indeed, the issue would seem not to be prudently framed as a polemic between those who champion health care as a determinant of health in general, and those who would suggest that it is never an important determinant of health, but rather as an exploration of which sorts of health care interventions in particular have been and continue to be worthwhile investments, in that they are capable of improving health status at reasonable cost.

In other words, the usual "plumbing diagram" of inter-relationships between population health status, prosperity, wealth and health care and health outcomes (Figure 1) which was initially conceptualized by Professors Greg Stoddart and Robert Evans from the CIAR Population Health Program in their landmark paper "Producing Health, Consuming Health Care" need not be interpreted nihilistically as implying that health care is generally useless, or completely unimportant as a determinant of health. Rather it can be interpreted as indicating that health care has a very variable role, as one of many factors influencing the health of individuals and populations. It is in this context, of attempting to examine exactly how and when various health care interventions make their most important contributions to population health status, that the two presentations this morning should be situated.

The first presenter, Professor Johann Mackenbach from the Department of Public Health and Social Medicine of Erasmus University in Rotterdam, the Netherlands, has already tackled this problem several times in a series of elegant analyses of time-trends and spatial pat-

terns of mortality potentially amenable to medical intervention, published over the last decade. In his paper today Professor Mackenbach will appropriately take as his departure point the seminal work of Thomas McKeown, to which most of us in this room have often referred in our professional careers if we are population health scientists. Specifically, Professor Mackenbach will re-examine the conclusions of McKeown, and the data that he based these upon, concerning the relative importance of medical care in producing the "epidemiological transition" of massive mortality decline in industrialized countries (and particularly the United Kingdom) from the middle of the 19th through to the middle of the 20th century (Figure 2). It is fitting that we re-examine the way that we think about McKeown, because a number of new lines of evidence have recently been adduced concerning the time trends of decreasing mortality that he spent two decades of his life examining. Again, in order not to throw out the baby with the bath water, it may be necessary to re-think McKeown's observations and his conclusions. I would put to you, therefore, one particular question which I think you should keep in your minds as Professor Mackenbach critiques the conventional wisdom concerning McKeown and his conclusions. Specifically, I would ask you whether we have sufficient empirical evidence to truly explain various elements of that epidemiological transition, even in the United Kingdom where a great deal of work on the question has been done. If so, is that evidence of a sort that allows us to draw firm conclusions about the role of various sorts of medical care in contemporary society? - which is exactly the issue that Professor Mackenbach will address.

The second speaker in this session, Professor Noralou P. Roos, will also use empirical evidence, from present-day Manitoba, one of Canada's western provinces, to assess the relationship between health care and particularly its quantified utilization, and the health status of various regions in that province. Professor Roos will therefore be taking us directly from the more historical considerations of Professor Mackenbach's paper to the current health policy dilemma faced not only in Manitoba, but throughout much of the industrialized world: "Should more money be spent on health care to meet the needs of those sub-populations with relatively poor health status? If so, what sorts of health care?"

Professor Roos is eminently qualified to tackle this problem, as one of the founders and now Director of the Manitoba Centre for Health Policy and Evaluation. Both Noralou and her husband Les are Associates of the Population Health Program of the Canadian Institute for Advanced Research. Their establishment in the last few years of the Manitoba Centre has been a landmark development, in terms of expanding a network of inter-connecting "nodes" of inter-disciplinary research expertise in the population health arena throughout Canada and beyond.

The particular strength of the Manitoba Centre is that it has the capacity, thanks to two decades of solid background work by the Roos and others, to make routinely collected health service utilization data, for both in-patient and ambulatory care by a wide-range of health care providers, computer-linkable to each and every Manitoba citizen longitudinally. This unique capacity allows the Centre to examine relationships between detailed patterns of health care utilization, and their associated resource expenditures, and individual or aggregate health status indices and outcomes, in the entire provincial population. Without question, the Manitoba Centre is as a result, the best population laboratory for health services research in this country, and perhaps in much of the western world.

It is in this context that Noralou and her colleagues from the Centre have examined the relationship between health care use and the health of the population among eight regions of Manitoba over recent years. As we shall see, their purpose in assessing the relationship is exactly what their title implies: "Seeking Levers for Policy Makers."

Figure 1 A Conceptual Model of the Determinants of Health

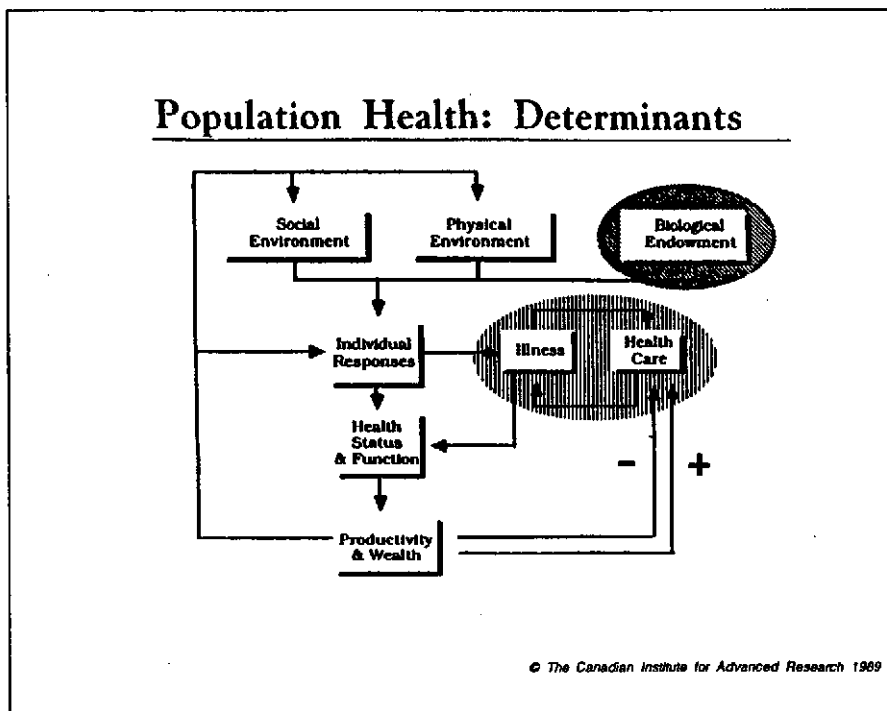
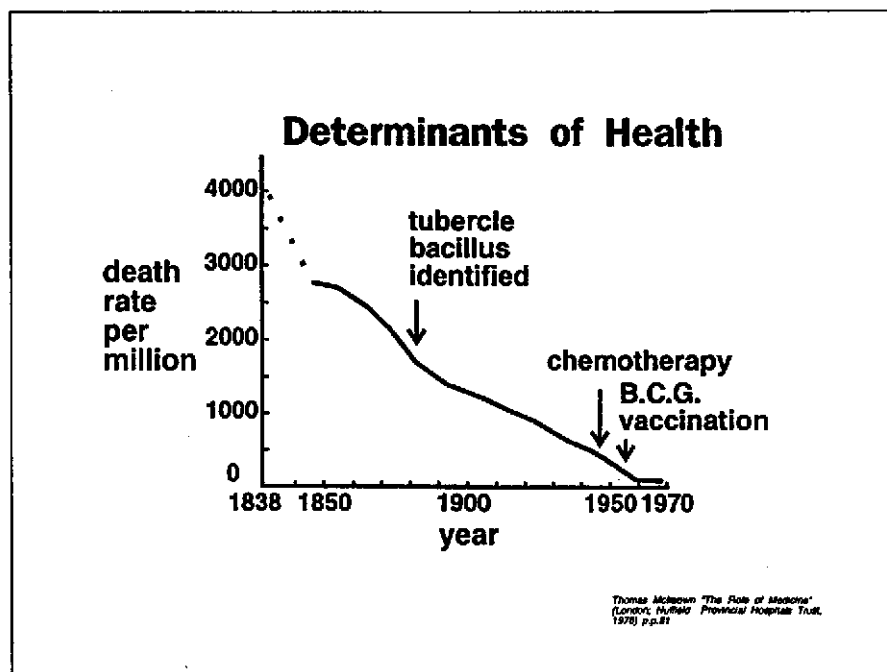


Figure 2 Tuberculosis Mortality in the U.K., 1838 - 1970



# The Contribution of Medical Care to Mortality Decline: McKeown Revisited

---

*Johan P. Mackenbach*

## Introduction

In St. James's Church (Sussex Gardens, London, U.K.) there is a stained glass window, which shows Sir Alexander Fleming at work in his laboratory. Fleming's laboratory was located in the vicinity of this church, and it was there that he discovered the bactericidal effects of penicillin. This church window has always struck me as a powerful illustration of the almost religious role of medicine in our modern society. In any case, it shows a deep respect for the advances of medicine, which in the popular opinion has become a master of life and death.

The contrast between this popular opinion and the current scientific view of the effects of medical care on the health of populations is enormous. This is due to a large extent to the work of the late Thomas McKeown. His famous analyses of the history of mortality in England and Wales, which originally appeared in a number of journal articles, were summarized in 1976 in two books, of which "The role of medicine: dream, mirage or nemesis" has been very influential<sup>1,2</sup>. In this book he showed, that the decline of mortality since of the middle of the nineteenth century was the result mainly of a decline in mortality from infectious diseases. The latter decline for the greatest part antedated the introduction of specific medical therapies, including Fleming's penicillin.

From these analyses McKeown has drawn two conclusions, which both have generally been accepted by the scientific community, at least by public health experts. The first was that medical care has not made a substantial contribution to mortality decline since the middle of the nineteenth century. The second, and more far-reaching conclusion was that today's health problems are also more likely to be controlled by changing the environment than by medical care.

Since the publication of McKeown's findings important criticisms have been voiced, which together with recent evidence suggest that some modifications of these conclusions may be necessary. The objective of this paper is to review McKeown's analysis and the more recent evidence, and to investigate whether the scientific opinion should or should not be reshifted a bit towards the popular opinion. Throughout this paper data from the Netherlands, a country with abundant data of relatively high quality, will be used as illustrations.

## The pattern of mortality decline and McKeown's arguments about the role of medicine.

In all western countries life expectancy at birth has increased tremendously since the middle of the nineteenth century. This increase in life expectancy was caused by an unprecedented decline in mortality, which in the pre-McKeown era was generally ascribed to improvements in medical care<sup>3</sup>. McKeown's arguments against any important effects of medical care on this decline of mortality have the structure of a syllogism:

- the decline in mortality can be attributed to a decline of infectious disease mortality;
- no effective medical interventions were available against infectious diseases before the middle of the twentieth century;
- thus, most of the decline in mortality cannot be attributed to the introduction of effective medical interventions.

The first part of McKeown's syllogism was supported by the results from an impressive analysis of changes in cause-specific mortality in England and Wales between 1848/54 and 1971. In this analysis it was shown that 74% of the reduction in total mortality was due to a decline of mortality from conditions attributable to micro-organisms. Among the latter, airborne infections such as respiratory tuberculosis and pneumonia made the largest contribution<sup>1,2</sup>.

Largely similar patterns of cause-specific mortality decline have been observed in other western countries<sup>4,7</sup>. Table 1 illustrates this for the Netherlands, where the earliest available data on mortality by cause of death apply to the years 1875/79. A number of specific infectious diseases accounted for 47% of mortality decline between 1875/79 and 1970 among men, and 45% of mortality decline among women. Airborne infections form the subgroup with the largest contribution. In addition to these specific infectious diseases, some other causes of death which are distinguished in table 1 are also of an infectious nature. Convulsions is an important example: deaths from convulsions mostly occurred among febrile children. Puerperal fever of course also is an infectious disease. Finally, it is quite likely that a part of the mortality decline for the "other" causes, which included ill-defined conditions, was actually due to infectious diseases. The 45 to 47% of table 1 should therefore be raised by some 20 to 30% in order to arrive at a more realistic estimate - an estimate which is also closer to that of McKeown<sup>6,7</sup>.

The second part of McKeown's syllogism was supported by a detailed analysis of the moment at which effective medical interventions for each infectious disease became available. In McKeown's view, the advent of the sulphonamides and antibiotics in the 1930's and 1940's marked this point in time for a large majority of infectious diseases. McKeown has published a number of impressive graphs, in which it is shown that the decline of mortality from a specific infectious disease had already largely occurred when these drugs became available. There was only a small number of exceptions, for example diphtheria (antitoxin, late nineteenth century) and syphilis (salvarsan, early twentieth century)<sup>1,2</sup>. For some infectious diseases, immunization was a more important breakthrough than antibiotics, but most forms of immunization were only introduced in the 1940's and the 1950's. Smallpox was the only important exception<sup>4,9</sup>.

In view of the importance of the timing of the decline of infectious disease mortality for McKeown's arguments, table 2 presents some quantitative data for the Netherlands. For many infectious diseases mortality decline was concentrated in the period 1903 to 1930. Only for smallpox and malaria, which had already declined as causes of death before 1875, the (remaining part of the) decline occurred almost completely before 1903. Only 19% of the decline of mortality from infectious diseases occurred in the period 1930 to 1950, and 4 to 5% in the period 1950 to 1970<sup>6</sup>.

After his account of the contribution of medical care, McKeown turned to a consideration of other possible explanations of the decline in infectious disease mortality. He first discussed the possibility of a spontaneous change in the relation between infectious agents and human hosts. There is some evidence that a decreasing virulence of the streptococcus has played a role in the decline of scarlet fever mortality, but this mechanism is unlikely to have been important for the decline of infectious disease mortality in general.

By exclusion, McKeown considers a third possibility, a favourable change in the environment, as the most plausible explanation for the decline of infectious disease mortality. McKeown considered two specific contributions to these changes: hygienic measures and rising standards of living. In his choice for the latter as the most important factor in mortality decline, the distinction between airborne and water- and foodborne infectious diseases played an important role. The classical hygienic measures from the second half of the nineteenth century (safe drinking water, sewage disposal) are not likely to have affected the airborne diseases, which made the largest contribution to mortality decline. Furthermore, McKeown argues that these hygienic measures cannot explain the start of the mortality decline, because they were effective only after ca. 1880 when mortality decline was already well underway.

After having discarded all these other possibilities McKeown concludes that a rise of living standards, particularly the associated improvements in nutritional status, have probably been the most important factor in mortality decline. It is important to note that no direct evidence in support of this conclusion is given by McKeown - he only cites some circumstantial evidence suggesting improvements in food supply<sup>1,2</sup>.

### **The direct effects of medical care on the historical development of mortality: other views**

I will now turn to the criticisms of McKeown's conclusions, and start with the evidence regarding the direct effects of medical care on mortality.

Although the first antibiotics, as McKeown showed, only arrived when the largest part of the decline of mortality from infectious diseases had already taken place, a quantitative analysis of the associated changes in mortality does suggest that the impact as such was substantial. In the Netherlands, when one compares the trend in mortality from infectious diseases before the introduction of antibiotics with that in a period after the introduction of these drugs, one sees that for almost all infectious diseases mortality decline accelerated. Table 3 shows that this acceleration is particularly impressive for bacillary dysentery, scarlet fever, tuberculosis, syphilis, erysipelas, rheumatic fever, and puerperal fever. Whereas mortality from all infectious diseases together declined 4% per year before the introduction of antibiotics, it declined 10% per year after the introduction. In the analysis on which table 3 is based, it was shown that for many infectious diseases there was not only a more permanent acceleration of mortality decline, but also an instantaneous, stepwise decrease in mortality coinciding with the introduction of antibiotics<sup>10</sup>.

Of course, although the observed changes in mortality are roughly coinciding with the introduction of antibiotics, a causal relationship is not necessarily implied. At least two other explanations should be considered: changes in the rate of improvement of living standards, and changes in the intensity and effectiveness of public health programmes. However, time series data on the availability of food, on housing conditions and on other correlates of living standards do not show accelerations of improvements after World War II in the Netherlands. The other alternative explanation, public health programmes, can only apply to tuberculosis. Pasteurisation of milk was introduced in 1940 in the Netherlands, reducing the incidence of non-respiratory tuberculosis. Mass radiography was introduced in 1949, and the combined effects of early detection and effective treatment may have caused a considerable acceleration in tuberculosis incidence decline. Nevertheless, in the late 1940's and the early 1950's mortality from tuberculosis declined much faster than incidence in the Netherlands, suggesting that case fatality was also declining considerably, possibly due to the introduction of streptomycin and other antibiotics<sup>10</sup>.

Although antibiotics were certainly a very important factor in the history of the medical treatment of infectious diseases, it is not as certain as McKeown would like us to believe, that they were the first effective medical intervention against infectious diseases. McKeown him-

self mentions a number of quantitatively unimportant infectious diseases for which other types of effective interventions were available before the middle of the twentieth century. Much more important is the question, whether improvements in medical care may have contributed to the decline of tuberculosis mortality<sup>11</sup>. Unfortunately, evidence on the effectiveness of isolation of patients in sanatoria and on the effectiveness of contact tracing and dietary advice is lacking. To assume that all these interventions have had no effect at all, of course is not a rational response to this lack of evidence. It is quite likely that these interventions did do some good - we only do not know how much.

Because in this discussion quantification of effects is so essential, I have tried to perform a sensitivity analysis of the direct contribution of medical care to the decline of mortality in the Netherlands, combining the data which were presented in tables 1 and 2 with a number of assumptions on the effect of medical care on mortality from specific conditions (table 4). 46% of the decline of total mortality was due to a number of specific infectious diseases; only a small part occurred after the introduction of antibiotics, which has arbitrarily been fixed in 1946 (14% - this figure was estimated on the basis of more detailed data than were presented in table 2). Table 3 showed that the rate of mortality decline for infectious diseases more than doubled after the introduction of antibiotics. How much of the decline after introduction of antibiotics was due to introduction of these drugs is difficult to say. It may be 60% (if one assumes that the pre-war trends would have continued unchanged in the absence of antibiotics), but it could also be more (if one assumes that the improvements in living standards had already reached their maximum effect before the war) or less. In table 4 it is assumed that the lower limit of the contribution of medical care to the decline of infectious disease mortality after 1946 is 25%, and the upper limit 75%. If we combine all these figures, it would appear that the lower limit for the contribution of medical care to decline of total mortality between 1875/79 and 1970 in the Netherlands, on the basis of the effects of antibiotics on specific infectious diseases, is 1.6%, and the upper limit 4.8%.

This is however not the only direct contribution of medical care to decline of mortality. Effects of interventions which were already available before the advent of antibiotics may have contributed to the decline of mortality from specific infectious diseases before 1946. Also, some infectious diseases are probably hidden in other causes of death. Furthermore, the effects of medical care were not limited to the infectious diseases. Two other categories of causes of death for which medical care has probably contributed to decline of mortality are diseases of the digestive system (improvements in surgery starting around 1930 have probably contributed to decline of mortality from appendicitis<sup>12</sup>, cholecystitis<sup>13</sup>, intestinal obstruction<sup>14</sup>, etc.) and perinatal conditions (improvements in antenatal and perinatal care have probably contributed to a mortality decline after about 1930<sup>15</sup>). Combining all these effects one arrives at an estimate of between 4.7% (lower limit) and 18.5% (upper limit) for the direct contribution of medical care to decline of mortality between 1875/79 and 1970 in the Netherlands.

### **Indirect effects of medical care on the historical development of mortality**

The second line of criticism against McKeown's conclusions deals with the indirect effects of medical care on mortality decline, through the contributions of doctors to sanitary reform and to behavioural changes in the population.

After McKeown had discarded the possible contributions of medical care to mortality decline, he turned to other explanations for the decline of mortality from infectious diseases. Unfortunately, as he himself already noticed, his argument at this stage resembles that of Sherlock Holmes, who is reported to have said to Watson: "When we have eliminated the impossible, whatever remains, however improbable, must be the truth"<sup>16</sup>.

McKeown's conclusion about the predominant role of the rise in living standards, particularly improvements in nutritional status, in the explanation of mortality decline is drawn by exclusion. The evidence supporting this remaining possibility was judged much less severely than that concerning the explanations which had been rejected before. Furthermore, such a conclusion drawn by exclusion will only be valid if all possible alternatives have been considered. That is not quite the case: sanitary reform may have been more important than McKeown thought, and he has also practically overlooked a completely different class of explanations, those related to cultural or behavioural changes.

McKeown's argument against a substantial contribution of sanitary reform was twofold. He thought it was unlikely that sanitary reform had contributed to a decline of mortality from airborne diseases, and the decline of mortality from food- and waterborne diseases started before hygienic measures had been implemented on a large scale. Both reasonings can be criticised. Safe drinking water supplies and sewage disposal systems were not the only hygienic measures which were implemented by the sanitary reformers. Improvements in working and housing conditions were other important parts of their program. It is not difficult to see, that better ventilation of working places and less crowded houses can have contributed to the decline of mortality from airborne diseases<sup>16</sup>. With regard to the second reasoning it can be argued that McKeown's emphasis on the starting point of the decline of mortality from food- and waterborne diseases distracted from the fact that the largest part of the decline in mortality from these conditions occurred in the period between ca. 1900 and ca. 1930 (table 3). That is precisely the period in which the benefits of safe drinking water supplies and sewage disposal systems are to be expected<sup>18,20</sup>.

Of course, the credit for these reforms mainly belongs to the public health sector. But many of these sanitary reformers were actually medical doctors, both in the Netherlands and in many other Western-European countries. They applied medical knowledge, gained inside the medical care system, in their epidemiological studies and in their proposals for reform. It is as difficult to imagine these reforms without the medical knowledge contributed by these doctors, as it is to imagine them without the expertise and skills of engineers. One can therefore argue that a small part of the mortality decline which was caused by the sanitary reform measures, actually is an indirect effect of medical care on mortality.

Even if one rejects conscious action on the part of medical care or the public health system as a main factor in the decline of mortality, it is not evident that the only autonomous changes were those directly flowing from the rise in living standards. Obviously, the rise in living standards in its turn needs an explanation. How could countries in Western Europe suddenly escape from the old economical order, and start an unprecedented rise in economic output? A discussion of this topic is outside the scope of this paper, but one plausible explanation is that cultural changes, in the broadest sense of the word, preceded the changes in economic activity<sup>21</sup>. In the Netherlands around 1850 a change in mentality occurred which affected many sectors of society: the educational system starts its expansion, all kinds of societies promoting the public good are founded, and also investments in new economic enterprises are made<sup>22</sup>. These cultural changes may act as a confounder of the association which is observed between living standards and decline of mortality.

Two important changes in determinants of mortality which are clearly related to these cultural changes are improvements in personal care, particularly personal hygiene, and the application of birth control. Improvements in personal hygiene may have contributed to the decline of mortality from many infectious diseases, such as typhus<sup>23</sup> and water- and food-borne diseases. Birth control will have contributed to an improvement of the health status of mothers, and also to an improvement of the health of children<sup>24</sup>.

Were these cultural changes completely external factors from the point of view of medical care? No, not entirely. Many private societies propagated modern behaviour among the pop-

ulation. This propaganda was also part of the work of the welfare clinics for infants and toddlers which were founded after the turn of the century. Again, among people from other disciplines doctors have played an outstanding role in these activities. They wrote treatises on infant care and on personal hygiene, and their medical expertise was diffused in this way throughout the population<sup>18,25</sup>.

### **The contemporaneous effects of medical care on mortality**

The third line of criticism deals with McKeown's statement that the historical evidence suggests, that today's health problems are also more likely to be controlled by changing the environment than by improving medical care. However, from the fact that environmental change has made a larger contribution to mortality decline than medical interventions in a certain historical period, it does not necessarily follow that environmental changes are more effective means of intervention under current circumstances too<sup>26</sup>.

In fact, a universal comparison of the effectiveness of environmental change and medical care is not possible with McKeown's historical data. "Environmental measures were tested when mortality was high, antibiotics when the mortality was low; environmental measures were tested alone, antibiotics against the background of environmental change. (...) English archers at Agincourt did better than English riflemen on the Somme, and it does not follow that archers have greater military potential"<sup>27</sup>.

One could further argue that the comparison between the changes induced by the introduction of antibiotics and other modern forms of medical care, and the changes that occurred in a very long period of time before that, does injustice to the advances of medical care. For contemporary patients only contemporary effects count, and it may therefore be more appropriate to compare the mortality effects of medical care with those of other influences operating in the same recent period<sup>28</sup>.

The basis for such a comparison can be found in a number of publications in which mortality from conditions which have become amenable to medical intervention is analyzed. These publications, which have dealt with the experience in many different countries, employ Rutstein's lists of "unnecessary untimely mortality" - selections of conditions in which death can be avoided by adequate preventive or therapeutic intervention<sup>29</sup>. Most of these conditions have only become amenable to medical intervention in the last 40 years. Rapid declines in mortality from the amenable causes were observed in all countries which have been studied<sup>30</sup>.

This is illustrated by table 5, which provides estimates of the effects of these mortality declines on average life expectancy at birth in the Netherlands. Due to these declines in mortality, life expectancy at birth increased by almost 3 years for men, and almost 4 years for women. The difference between men and women can almost entirely be explained by the greater importance of hypertensive and cerebrovascular disease among women. Of course, it is too generous to attribute all the decline in mortality from these conditions (and the associated increase in life expectancy) to improvements in medical care. Probably only a part was due to these improvements, but this may at least in part be compensated for by the fact that medical interventions also made an impact on causes of death which were not included in these selections, such as ischemic heart disease<sup>31</sup>.

In any case, if one compares these gains in life expectancy due to reductions in mortality from amenable conditions, with the observed gains in life expectancy in the same period, it is evident that something remarkable has happened. Among men, the actual gain in life expectancy was less than the gain which could have been expected on the basis of the reductions in mortality from amenable conditions. This implies that mortality from non-amenable conditions has increased over the same period. Among women, the observed gain in life expectancy was almost six years, of which two-thirds appeared to have been due to a decline

of mortality from amenable conditions. Whatever the exact contribution of medical care to the decline of mortality from amenable conditions, the contribution to the development of overall mortality and life expectancy is likely to have been substantial relative to that of other contemporary determinants<sup>32</sup>.

## Conclusions

The evidence presented in this paper suggests that:

- the direct contribution of medical care to historical mortality decline was certainly not negligible, and perhaps even sizable;
- medical care has also made indirect contributions to historical mortality decline, by transfer of knowledge from the medical profession to sanitary reformers and the general public;
- medical care is likely not to have been a less important factor in recent mortality decline than other contemporary determinants.

Where does this leave us with regard to McKeown's two main conclusions and the gap between the two conventional wisdoms (one popular, one scientific) on the effectiveness of medical care? First of all, it is important to emphasize the limitations of the analyses reviewed in this paper. As so many other papers before, including those of McKeown, only the effects of medical care on mortality were considered. Due to an almost complete lack of data on time trends in morbidity, the effects of medical care on morbidity are even more uncertain than those on mortality. Secondly, although the evidence on mortality is much less sparse, it is still far from dense. Patterns of mortality decline have been described in detail for several countries, but estimates of the contribution of medical care to mortality decline are largely based on informed guesses. This also applies to the contribution of other factors, such as hygienic measures and rising living standards - actually, there is even less firm evidence on these competing factors, so that a valid comparison with the effects of medical care is difficult.

Given these uncertainties it is perhaps wise to refrain from any far-reaching conclusions. The role of medicine may not have been, nor now be as modest as McKeown liked us to believe, and it is not clear at all that the conviction among public health experts on the relative inferiority of medical care as a determinant of mortality change has more scientific backing than the popular wisdom on the role of medicine reflected in church windows.

## References

1. Th.F. McKeown, The role of medicine - dream, mirage of nemesis? (London: Nuffield Provincial Hospitals Trust, 1976).
2. Th.F. McKeown, The modern rise of population (London: Edward Arnold, 1976).
3. G.T. Griffiths, Population problems in the age of Malthus (London: Frank Cass, 1967).
4. A.R. Omran, "Epidemiologic transition in the U.S.; the health factor in population change", Population Bulletin 32 (1977): 3-42.

5. G. Caselli, "Health transition and cause-specific mortality", in R. Schofield, D. Reher, A. Bideau (eds.), The decline of mortality in Europe (Oxford: Clarendon Press, 1991).
6. J.P. Mackenbach, "De epidemiologische transitie in Nederland", Nederlands Tijdschrift voor Geneeskunde 137 (1993): 132-138.
7. J.P. Mackenbach, "De achtergronden van de sterftedaling tussen 1875/79 en 1970 in Nederland", Tijdschrift Sociale Gezondheidszorg 71 (1993): 192-197.
8. A.J. Mercer, "Smallpox and epidemiological-demographic change in Europe: the role of vaccination", Population Studies 39 (1985): 287-307.
9. P.E. Razzell, "An interpretation of the modern rise of population in Europe - a critique", Population Studies 28 (1974): 5-17.
10. J.P. Mackenbach, C.W.N. Looman, "Secular trends in infectious disease mortality in the Netherlands: quantitative estimates of changes coinciding with the introduction of antibiotics", International Journal of Epidemiology 17 (1988): 618-624.
11. A. Mercer, Disease, mortality and population in transition (Leicester etc.: Leicester University Press, 1990)
12. B. Berry, R.A. Malt, "Appendicitis near its centenary", Annals of Surgery 200 (1984): 567-575.
13. W.F. Becker, J.L. Powell, R.J. Turner, "A clinical study of 1060 patients with acute cholecystitis", Surgery Gynaecology Obstetrics 104 (1957): 491-496.
14. I. Cohn, "Intestinal obstruction", in: H.L. Bockus (ed.), Gastroenterology (vol. 2, 3rd edition) (Philadelphia: Saunders, 1976).
15. N.R. Butler, E.D. Alberman, Perinatal problems: the second report of the 1958 British Perinatal Mortality Survey, Edinburgh: Livingstone, 1958).
16. S. Szreter, "The importance of social intervention in Britain's mortality decline c. 1850-1914", Journal of the Social History of Medicine 1 (1988): 1-37.
17. E.W. Hofstee, De demografische ontwikkeling van Nederland in de eerste helft van de negentiende eeuw (Deventer: Van Loghum Slaterus, 1979).
18. E.S. Houwaart, De hygiënisten: artsen, staat en volksgezondheid in Nederland 1840-1890 (Groningen: Historische Uitgeverij Groningen, 1991).
19. I. Vogelzang, De drinkwatervoorziening van Nederland voor de aanleg van de drinkwaterleidingen (Academisch Proefschrift Utrecht, 1956)
20. H. van Zon, Een zeer onfrisse geschiedenis. Studies over niet-industriële vervuiling in Nederland 1850-1920 (Academisch Proefschrift Groningen, 1986).

21. W.W. Rostow, The world economy: history and prospect (London: MacMillan, 1978)
22. J.H. van Stuijvenberg, "De economie in de Noordelijke Nederlanden 1770-1970", in Algemene Geschiedenis der Nederlanden, deel 10 (Haarlem: Van Dishoeck, 1981).
23. B. Luckin, "Evaluating the sanitary revolution: typhus and typhoid in London, 1851-1900", In: R. Woods, J. Woodward, Urban disease and mortality in nineteenth century England (London: Batsford Academic and Educational, 1984).
24. R. Reves, "Declining fertility in England and Wales as a major cause of the twentieth century decline in mortality", American Journal of Epidemiology 122 (1985): 112-126.
25. A. de Regt, Arbeidersgezinnen en beschavingsarbeid: ontwikkelingen in Nederland 1870-1940 (Meppel: Van Gorcum, 1984).
26. S. Kunitz, Disease change and the role of medicine - the Navajo experience (Berkeley etc.: University of California Press, 1983).
27. A.F. Lever, "The role of medicine" Lancet (1977): 352-355.
28. P.B. Beeson, "McKeown's 'The role of medicine' - a clinician's reaction", Milbank Memorial Fund Quarterly 55 (1977): 365-371.
29. D.D. Rutstein, W. Berenberg, T.C. Chalmers et al, "Measuring the quality of medical care: second revision of tables of indexes", New England Journal of Medicine 302 (1980): 1146 and New England Journal of Medicine 294 (1976): 582-588.
30. J.P. Mackenbach, M.H. Bouvier-Colle, E. Jougl, "'Avoidable' mortality and health services: a review of aggregate data studies", Journal of Epidemiology and Community Health 44 (1990): 106-111.
31. L. Goldman, E.F. Cook, "The decline in ischaemic heart disease mortality rates", Annals of Internal Medicine 101 (1984): 825-836.
32. J.P. Mackenbach, C.W.N. Looman, A.E. Kunst et al, "Post-1950 mortality trends and medical care: gains in life expectancy due to declines in mortality from conditions amenable to medical intervention", Social Science and Medicine 27 (1988): 889-894.

**Table 1** The contribution of specific causes of death to mortality decline between 1875/79 and 1970 in the Netherlands [a]

Cause of death	Percentage contribution to all-cause mortality decline [b]	
	Men	Women
Infectious diseases	47	45
Diarrheal diseases	11	10
Smallpox	0	0
Scarlet fever and measles	2	2
Pertussis and diphtheria	3	3
Respiratory tuberculosis	13	14
Pneumonia/acute bronchitis/influenza	12	10
Malaria	1	1
Other	6	5
Mental diseases/diseases of nervous system	13	11
Convulsions	7	6
Other	6	5
Diseases of respiratory system	8	8
Diseases of digestive system	3	3
Diseases of pregnancy/childbirth/puerperium	na	2
Puerperal fever	na	1
Other	na	1
Congenital diseases	0	0
Perinatal conditions	4	4
All other diseases	24	27
Total	100	100

[a] Adapted from ref. 7

[b] The percentage contribution of each cause of death to mortality decline was determined by subtracting the age-standardized mortality rate for that cause in 1970 from the age-standardized rate in 1875/79. The difference was divided by the total mortality decline over the same period, and multiplied by 100.

**Table 2** The percentage distribution of the decline in mortality from infectious diseases in the Netherlands over 4 subperiods, 1875/79-1970 [a]

	Sex	Mortality decline 1875/79- 1970 [b]	Percentage distribution				
			1875/79- 1970	1875/79- 1903	1903- 1930	1930- 1950	1950- 1970
Diarrheal diseases	m	1.89	100	20	71	8	1
	f	1.68	100	29	64	6	1
Smallpox	m	0.02	100	100	0	0	0
	f	0.02	100	50	50	0	0
Scarlet fever and measles	m	0.30	100	23	50	23	3
	f	0.28	100	21	54	21	4
Pertussis and diphtheria	m	0.51	100	53	22	18	8
	f	0.53	100	51	23	20	7
Respiratory tuberculosis	m	2.23	100	36	41	16	7
	f	2.34	100	44	31	21	4
Pneumonia/ acute bronchitis/ influenza	m	2.14	100	13	53	32	3
	f	1.66	100	2	60	31	7
Malaria	m	0.17	100	100	0	0	0
	f	0.17	100	94	6	0	0
Other infectious diseases	m	1.00	100	28	41	22	9
	f	0.86	100	33	36	23	9
Subtotal infectious diseases	m	8.27	100	27	49	19	4
	f	7.54	100	31	45	19	5
Total mortality, all causes of death	m	15.14	100	40	45	15	0
	f	15.79	100	40	33	17	11

[a] Adapted from ref. 6

[b] Difference between the age-standardized death rates (in deaths per 1000 person-years) of 1875/79 and 1970.

**Table 3** Trends in age- and sex-adjusted mortality from infectious diseases before the introduction of antibiotics (1921-1939) and after the introduction of antibiotics (1947-1968) [a]

	Annual percentage change in mortality [b]	
	1921-1939	1947-1968
Bacillary dysentery	+5	-19
Typhoid fever	-10	-2
Scarlet fever -3	-13	
Measles	-8	-9
Tuberculosis -6	-16	
Pneumonia	-4	-5
Acute bronchitis	-7	-9
Influenza	-1	-5
Syphilis	-1	-9
Erysipelas	-2	-11
Septicemia	-6	-4
Meningococcal meningitis	-1	-6
Non-meningococcal meningitis	-8	-4
Otitis media +5	-8	
Upper respiratory infections	+3	-5
Cystitis	-4	-6
Skin infections	-3	-7
Osteomyelitis	-1	-5
Rheumatic fever	-0	-16
Puerperal fever	-1	-11
All infectious diseases	-4	-10
All other diseases	-1	0
Total mortality	-2	-1

[a] Adapted from ref. 10

[b] Estimated on the basis of a Poisson regression analysis, in which the mortality rate was modelled as a function of calendar-year and the number of person-years at risk by age and sex. The period 1940-1946 was left out of the analysis because of the effects of World War II.

**Table 4** An estimate of the direct contribution of medical care to the decline of mortality between 1875/79 and 1970 in the Netherlands.

		Contribution of medical care to decline of mortality	
		Lower limit	Upper limit
1. a.	Contribution of specific infectious diseases to decline of total mortality 1875/79-1970:	46%	
b.	Of which after 1946:	14%	
c.	Of which due to medical care:	25-75%	
d.	1a * 1b * 1c	1.6%	4.8%
2. a.	Contribution of specific infectious diseases to decline of total mortality 1875/79-1970:	46%	
b.	Of which between 1900 and 1946:	57%	
c.	Of which due to medical care:	5-20%	
d.	2a * 2b * 2c	1.3%	5.2%
3. a.	1d + 2d	2.9-10.0%	
b.	Contribution of infectious diseases hidden in other cause-of-death groups to decline of total mortality 1875/79-1970	20-30%	
c.	(1a + 3b)/1a	1.4-1.7	
d.	3a * 3c	4.1%	17.0%
4. a.	Contribution of diseases of digestive system to decline of total mortality 1875/79-1970:	3%	
b.	Of which after 1930:	23%	
c.	Of which due to medical care:	30-70%	
d.	4a * 4b * 4c	0.2%	0.5%
5. a.	Contribution of perinatal conditions to decline of total mortality 1875/79-1970:	4%	
b.	Of which after 1930:	35%	
c.	Of which due to medical care:	30-70%	
d.	5a * 5b * 5c	0.4%	1.0%
6.	Total contribution of medical care to decline of total mortality 1875/79-1970 (3d + 4d + 5d)	4.7%	18.5%

**Table 5** The effects of mortality reductions for conditions which have become amenable to medical intervention on average life expectancy at birth in the Netherlands, 1950/54 - 1980/84 [a]

	Life expectancy gains (in years)	
	Men	Women
Gains [b] due to reductions in mortality from:		
- infectious diseases [c]	0.94	1.13
- certain surgical conditions [d]	0.36	0.25
- maternal and perinatal conditions	0.72	0.63
- hypertensive and cerebrovascular disease	0.43	1.32
- other amenable conditions [e]	0.58	0.80
- all amenable conditions [f]	2.96	3.95
Observed gain [g]	1.91	5.86
<p>[a] Adapted from ref. 32</p> <p>[b] Calculated as the difference between actual life expectancy at birth in 1980/84 and the life expectancy which would have been attained if the mortality rates for the selected conditions had remained at the level of 1950/54.</p> <p>[c] Tuberculosis, Pneumonia/influenza, Septicemia, Infections urinary system, Other infectious diseases.</p> <p>[d] Peptic ulcer, Appendicitis, Cholelithiasis/-cystitis, Abdominal hernia, Ileus without hernia, Benign prostatic hyperplasia</p> <p>[e] Diseases of the thyroid, Diabetes mellitus (&lt; 25 years) Pernicious anemia, Other anemias, Congenital digestive anomalies, Congenital cardiovascular anomalies, Rheumatic heart disease, Nephritis and nephrosis, Cancer of lip and skin, Cancer of kidney (&lt; 15 years), Morbus Hodgkin, Cancer of testis, Leukemia (&lt; 15 years), Cancer of cervix uteri.</p> <p>[f] Sum of the above</p> <p>[g] Calculated as the difference between the average life expectancy at birth in 1980/84 and the average life expectancy at birth in 1950/54.</p>		

# Assessing the Relationship between Health Care Use and the Health of a Population: Seeking Levers for Policy Makers

---

*Noralou P. Roos\**

*Charlyn Black\**

*Norm Frohlich\*\**

*Carolyn DeCoster*

*Marsha Cohen\*\*\**

*Doug Tataryn\**

*Cam Mustard\**

*Fred Toll*

*K.C. Carriere\**

*Charles Burchill*

*Leonard MacWilliam*

*Bogdan Bogdanovic*

This work was supported by the Health Services Development Fund via a contract establishing the Manitoba Centre for Health Policy and Evaluation. All authors are part of the Centre's Population Health Information Systems Group; \*Members of the Department of Community Health Sciences, Faculty of Medicine, University of Manitoba; \*\*Member of the Faculty of Management, University of Manitoba; \*\*\*Member of the Institute of Clinical and Evaluative Sciences, Sunnybrook Health Science Centre.

**A**t a time of upheaval in health care systems across North America, there is a pressing need to shift the focus of the health care debate from a preoccupation with health care expenditures towards concern for the health of a population. In Manitoba, Canada, the Manitoba Centre for Health Policy and Evaluation (MCHPE) has developed a population health based information system to help the public understand the need for such a shift, and to help planners identify the levers for making such a shift. We believe that the principles underlying this system are applicable in other jurisdictions. This article describes the approach and illustrates its application in one health care jurisdiction, the province of Manitoba.

The health care system has developed in a haphazard manner. Allocations for hospital services have historically been made in response to demands mediated by population growth, increases in volume of use, technological imperatives and political pressure. The numbers and specialties of physicians and their practice locations have never been linked to health needs of populations. Critical assessments of medical care typically focus on the clinical outcomes of individual treatments and quality of care delivered by institutions, not on the health of populations.

The Population Health Information System (PHIS) is designed to focus on the link between health care utilization and health, to make it possible to examine how effectively and efficiently a health care system produces (or fails to produce) health across defined population groups. It compares health characteristics of regional populations with the use they make of the system. This information system builds on an administrative data system documenting services by hospitals, nursing homes, and physicians. The validity of information contained in the Manitoba system and the utility of the data for addressing important questions in health services research have been demonstrated repeatedly.

## Conceptual Model and Key Concepts

The conceptual model which underpins the Population Health Information System is outlined in Figure 1. It is a modification of the model proposed by Evans and Stoddart (1990), which has been expanded to incorporate individual level effects and to incorporate change over time. It is useful for consideration at both the individual and the population levels.

Health status is recognized as encompassing health states (ie. disease) as well as functional status. Together with health perceptions, health status is a component of well-being.

The PHIS model recognizes that a range of background factors influence health status and well-being. These include socio-economic factors, which characterize the economic, social and physical environments in which individuals live and work, as well as demographic and genetic factors. Health states are influenced by all of these background factors, but are mediated by individual response to socio-economic and biologic circumstances. While health status strongly influences individuals' "need" for health interventions, the interplay between individuals' health status and health perceptions (which are shaped in part by physician practice patterns) influences demand for these activities. Need and demand both, in turn, influence utilization of specific health care services, but utilization patterns are also affected by supply and practice pattern factors operating within the health care system. Utilization of health care services, mediated by individual response, may affect the health status of individuals at a later point in time (Time 2).

Health states (as identified at Time 2) are therefore influenced by socio-economic background factors, initial health status and utilization of the health care system at an earlier point in time (Time 1). The model permits assessment of the relative contribution of each of these factors to health status and to population health. Conversely, the model can be used to understand how health status ultimately feeds back both to influence the socio-economic environment in which individuals operate and to further influence utilization of the health care system.

Not too long ago the world was thought to be very much simpler. It was assumed (and it is still assumed by most of the public) that ill health completely determined use of the system (and associated costs) and that such use was effective in improving health. We have since recognized that factors other than ill health influence use including the supply of physicians and hospital beds, and that there are marked variations in how physicians practice medicine. The role of patient demand - independent of physician supply and practice style is - less well understood. Furthermore, ill health is not a random event. Genetic endowment including predispositions to develop specific diseases, aspects of the physical environment including pollution and key social and economic characteristics, such as poverty and lack of education, have a strong negative impact on the health of the population<sup>1</sup>. The impact of social determinants is thought to be substantially greater than that of environmental pollution (Rose, 1992) and the evidence for social determinants at least as significant as that for genetic impacts (Harris, et al., 1992; Horgan, 1993).

The principles around which the information system is organized are outlined in Figure 2. The PHIS is population based and designed to track the background risk factors, health status, and health care use of populations (regardless of where such usage takes place), as distinct from examining clinical care and associated outcomes for individual patients or treatments. Standardizing the age and sex characteristics of a population across geographic areas adjusts for two of the important determinants of health and use of the health care system.

Data on the supply of services (hospital beds, nursing home beds, and physicians), the usual parameters of health planning, are presented but our system also directly measures access to care, focusing on the proportion of individuals resident in a given area who access a service, regardless of where the service is obtained.

The system is organized around issues relevant to policy makers, for example intensity of use: how much do residents of regions vary in their use of high-tech teaching hospitals versus their use of small rural hospitals for their acute care? What is the relative use across regions of resource-intensive procedures (ie. hysterectomy versus less intensive procedures such as dilatation and curettage)? What proportion of care is delivered by specialist physicians versus generalists? To what extent is health care concentrated on individuals with frequent contact with physicians? As well, to what extent is the use of health services differentially concentrated across regions for individuals who are close to death? We will compare usage patterns across regions whose residents have similar levels of health, allowing policy makers to come closer to answering the "which rate is right?" question.

This system also permits comparing discretionary use of services across geographic areas. We have calculated expected length of stay of nursing home patients based on the age, sex, and level of care at entry to a nursing home using the estimates of Shapiro & Tate (1988). For example, females aged 65-74 years entering at the lowest level of care will stay on average 14.7 years in a nursing home while males 85 years and older entering at the highest level will remain on average 1.8 years. (In Manitoba 95% of the individuals entering a nursing home remain there until death.) Discretionary use of resources is suggested by the degree to which one region institutionalizes younger, healthier individuals (who have a much longer expected length of stay) than another region, as does the extent to which patients are admitted to hospitals for diagnoses whose admission patterns are known to vary markedly across areas (Wennberg, 1984). A similar effort will be made to classify physician visits into conditions classified as "posing a serious threat to health" versus those classified less serious.

Costs incurred by region's residents will be affected by how often residents access a type of care (are admitted to hospital during the year or contact a physician at least once), the average cost of the service incurred (whether the physician was a specialist or whether the hospital day was spent in a high-tech teaching hospital or a small rural institution) and the number of services per user. How each of these factors interact in determining total average cost per resident will also be assessed.

Indicators have also been selected because of their potential value to those managing the system. The hospital indicators distinguish between medical, surgical, pediatric, psychiatric, and obstetrical admissions and focus on use which takes place in the region of residence versus out-of-region. The nursing home indicators distinguish between types of homes (those with an ethno-religious base versus others, proprietary- or non-proprietary ownership, and whether the home is juxtaposed to a hospital or freestanding).

This system permits cross sector analyses, combining usage using a dollar figure where possible and, in the case of nursing home and hospital use, summing total days of institutional care. Finally, regional profiles will be created, showing how each region's health, socio-economic risk and use characteristics differ from the provincial norm.

### Modules of the System

The system will consist of several modules as outlined in Figure 2. Each module will annually produce a set of key findings, tables and graphs reporting data for a set of indicators across regions (or hospital service areas) of Manitoba. Each of these modules is briefly described below and results from the first annual reports presented.

### Population Health/Ill Health

Many instruments to measure health status are currently available and can be broadly classified into those that measure individual health status and those that measure the health of populations or communities. It is with the latter that our Population Health Information

System is concerned. There are several health status indicators currently developed to measure the health of populations (U.S. Health 2000, Statistics Canada Health Indicators Working group and so forth), and our information system draws heavily from these. Using administrative data to assess health status has the major advantage over surveys in that the entire population's health status can be readily assessed repeatedly over time for a fraction of the cost of a comprehensive health survey.

Cohen and MacWilliam (1993) have developed 105 separate indicators of health status from administrative data using Vital Statistics mortality data, hospital discharge diagnoses, and the diagnosis associated with each physician visit. These indicators include various aspects of community health such as mortality/morbidity from cancer, injuries, and chronic diseases (see listing in Appendix). In addition, we focus on the prevalence of conditions such as hypertension and rheumatoid arthritis that are associated with poor functional status (Pope, 1988). We also include a set of indicators which relate to the impact of medical treatment, hospitalizations and deaths which should be avoidable given timely and appropriate medical intervention or public health action (Charlton et al., 1983; Weissman et al., 1992).

While mortality is widely accepted as a useful measure for assessing the health of communities, utilization data have been used less frequently for this purpose. We have demonstrated in several studies that administrative data perform well in assessing health status (Young et al., 1991; Roos NP, 1989; Mossey & Roos 1987). In Manitoba 85% of the population resident in Manitoba over the full year has at least one physician contact over the course of a year, and 94% over a two year period. Even elderly individuals not visiting physicians over the course of the year tend to be well (Shapiro and Roos, 1985). However, since utilization is influenced by factors other than health status (supply of beds and physicians being one example), we have been conservative when calculating utilization-based measures of health status. Indicators are grouped and summarized separately by source of data, that is, those developed from cause of death records are analyzed separately from hospital or physician use variables. As well, all utilization derived health status measures are individual-based, not discharge or contact based; for example, we calculate rates of persons hospitalized at least once per year for diabetes, not admission rates for diabetes.

British researchers (Carstairs and Morris, 1991) as well as a group at McMaster (Eyles et al., 1991) have suggested that the standardized mortality ratio (SMR) for those aged 0-64 is the best single indicator of health status capturing the need for health care<sup>2</sup>. The Scottish health authorities have used the SMR for those 0-64 for allocating funds for health care across the system. Figure 3 demonstrates the marked differences across Manitoba regions in this measure, with the two northern regions (Norman and Thompson - Figure 4) showing very poor health status and residents of Westman particularly, demonstrating quite good health status.

For the most part, across indicators based on every type of data, Norman and Thompson regions show SMRs which are both above the provincial average and much higher than for the other regions of the province. For the rest of the report we will order all graphs identical to Figure 3 - with the least healthy regions on the far right.

### Socio-Economic Status

Systematic relationships have been observed between socio-economic characteristics and health status for roughly 100 years in England and France (Libertos et al., 1988) and they are currently being rediscovered in North America (Pappas et al., 1993; Cassel, 1976; Haan et al., 1987). In Manitoba, Thompson residents are at highest risk for poor socio-economic status,<sup>3</sup> followed by residents of Norman and Parklands (although the latter region's score is not significantly different from that of the province (Figure 5). Residents of the five remaining regions score similarly and are at low risk for poor socio-economic status (Frohlich and

Mustard, 1993). In Manitoba we observe a strong relationship between indicators of health status and regional scores on our socio-economic risk index. When regional health status (measured by the standardized mortality ratio) is regressed against the risk index, 87% of the variation in health status as measured by the test index is explained by differences in the socio-economic risk index. This relationship was observed both at the regional level and less robustly across two hundred and fifty-nine smaller geographic units, the analytic level used for developing the index.

### Use of Hospitals

Table 1 is taken from the hospital use module of the Population Health Information System (Black, et al., 1993). Table 1 focuses on acute care (stays of less than 60 days). The rate of hospitalization for individuals during a given year is distinguished from the number of hospital separations per 1000 residents, the latter being approximately 44% higher than the former. Because patients are sometimes transferred from small rural hospitals to larger centres we have also calculated episodes of hospital care (hospital to hospital transfers are counted as one episode). Rural Manitoba residents (the non-Winnipeg comparison group) are admitted to hospital at much higher rates than Winnipeg residents, ranging as high as 263 separations per 1000 residents in Thompson compared with 101 separations per 1000 Winnipeg residents. Although Winnipeg residents tend to have longer hospital stays than non-Winnipeg residents (averaging 7.3 days versus 6.1 days), non-Winnipeg residents use substantially more acute-care hospital days than Winnipeg residents (1018 days versus 743 days per 1000 residents).

If one compares use across the regions (Figure 6) the high use of hospitals by particularly Norman and Thompson residents parallels these regions' poor score on the socio-economic risk index (Figure 5) and their poor health status as measured by the standardized mortality ratio. Note also that fully 42% of Thompson residents' care and almost 1/3 (28%) of Norman residents' care occurs in teaching or urban hospitals - ones located out of these regions - so bed supply (although high in Norman) does not adequately explain these findings.

Residents of Westman, Winnipeg, and Central regions demonstrate similar and quite good health status and are at relatively low socio-economic risk (Figure 3 and 5). However, there are very different patterns of hospital use across these regions. Winnipeg residents use many fewer days than residents of the other regions, suggesting that hospital use in the other healthy regions may be unnecessarily high.

The hospital use module contains many types of indicators characterizing use across the system (Appendix 1). Figure 6 illustrates the very different mixes of care received by residents of different regions. Winnipeg residents receive a much higher percentage of their care in a teaching hospital (the two teaching hospitals are located in Winnipeg). These are expensive institutions as reflected by the inter-provincial per diems negotiated with other provinces (approximately \$750 per day in 1991-92) in contrast to an average of approximately \$300 per day negotiated for those institutions classified as "other rural".

Despite the marked differences in hospital use across regions, there is no clear relationship between rate of use and proportion of hospital days used by very sick patients (Figure 7). Although Winnipeg residents use half as many short-stay days as do Thompson residents, 22.5% of hospitalized Winnipeg residents had high levels of co-morbidity compared with 19.3% of Thompson residents who were hospitalized. Even this difference may be overstated because many Thompson residents' stays occur in small rural hospitals known to systematically under-report co-morbid diagnoses relative to reporting styles in Winnipeg hospitals. Co-morbidity was measured by Charlson's et al., (1987) index<sup>2</sup> which we have implemented using claims data.

As outlined in Appendix I, the information system will replicate Table 1 for the following sub analyses:

- where individual was hospitalized (in-region/out-of-region, Winnipeg/out-of-region, other)
- type of hospital to which admitted (teaching, community, etc.)
- service to which admitted (surgical, obstetrical, etc.)
- resource intensity of care received (using DRG weights)
- complexity of case (whether individual had major co-morbidities or not)
- discretionary nature of admission (high variation conditions)

### Nursing Home Care

Table 2 presents an overview of regional residents' use of nursing homes. This table concentrates on the elderly (those aged 75 years and older), since 85% of those admitted to Manitoba nursing homes are in this age group.<sup>6</sup> Use of the system by all ages and by age-specific groups will also be presented (DeCoster et al., 1993). As can be seen, Manitoba's two northern regions Norman and Thompson, have very few elderly residents. Despite or perhaps because of their small elderly populations, Norman has a high per capita rate of nursing home use, whereas Thompson has a low per capita rate of nursing home use. These two "outlier" regions aside, there is some variation in availability of nursing home beds per 1000 population across regions. Given that new nursing home beds are opened according to a planning ratio based on 90 beds per 1000 population aged 70 years or more, existing inequities are historical. The differences in nursing home bed availability results in elderly residents of Westman one of the healthiest regions of the province spending approximately 27% more days in nursing home than Parklands' residents; statistical tests confirm there is much more variation in hospital use across regions than in use of nursing homes.

Manitoba's centralized assessment system which controls placement in all nursing homes also operates to minimize the amount of discretionary use in the system; regions vary relatively little in the expected length of stay, regardless of the type of home examined (secular/ethno-cultural, proprietary/non). There is also little variation in the estimated cost per day of nursing home care (Table 2), ranging from \$75.06 for Westman nursing home residents to \$79.01 for Winnipeg nursing home residents. (The range in costs per day of nursing home care is largely influenced by the proportion of residents at each care level.) The information system will replicate Table 2 for:

- Level of care to which patient is admitted
- Type of home

### Use of Physician Services

Table 3 illustrates information identifying access to and use of physician services. We have included all physician visits except those to hospitalized patients: physician office visits, ambulatory clinic care occurring in hospitals, emergency room visits, physician visits to a patient's home or to nursing home residents, and all consultations which occur in any of these settings (Tataryn et al., 1993). Despite the remarkably different availability of physicians across the various regions, ranging from 14.9 physicians per 10,000 population in Winnipeg to 5.0<sup>7</sup> physicians in Eastman, access to physicians is remarkably uniform across Manitoba regions, as reflected in the percentage of individuals contacting a physician at least once over the course of the year (almost 80% of the residents in every region make at least 1 contact)<sup>8</sup>. Given that physician supply varies 300 percent across the regions, the 30% range in rate of physician contact (from 4.1 visits per resident in Central region to a high of 5.3 per Winnipeg resident) is small.

In addressing the "which rate is right" issue, preliminary comparisons across the healthiest regions can be made. Residents of Central, Westman, and Winnipeg score similarly on the measures of health status and socio-economic risk factors and a similarly high percentage of residents has contact with physicians throughout the course of the year ranging from 81% of Central residents to 85% of Winnipeg residents. However, 41% more per capita was spent on physician contacts for Winnipeg residents in 1991/92 than for Central residents (\$123 per resident vs. \$87 per resident) because both visits per user is higher for Winnipeg residents and more Winnipeg care is delivered by medical specialists - hence a higher cost per visit. Although not presented here, when contact patterns of Winnipeg residents are compared to those of Central residents, the three biggest contributors to Winnipeg's high expenditure patterns were visits to psychiatrists, pediatricians and internists, with over a third of the additional expenditures relating to psychiatric contact alone.

The information system replicates Table 3 for the following subanalyses:

- physician specialty
- where visit occurred (in-region, out-of-region Winnipeg/out-of-region others)
- frequency of contacts made by individual in year (1-7/8-14/15+)

### Demographic Patterns

Using the population registry, it is possible to track movement to and from regions of the province by key characteristics such as age and socio-economic risk status. It has been argued that population growth, particularly increases in the numbers of the very young or very old, should be used for planning health services. Carstairs (1991) has argued, however, that changes in the number of individuals in a population who are at high socio-economic risk is at least as important in estimating health care needs; that is, elderly who are well off will have fewer health care needs than those who are at risk. In future versions of the system we will report demographic changes across geographic areas (migration, births and deaths) according to the age of individuals and also according to the socio-economic characteristics of the neighbourhood of residence.

### Cross-Sector Summaries of Use

Focusing on the number of days which the elderly population of a region spends in a nursing home or in non-acute hospital stays permits examining the degree to which the two services substitute for one another across regions. Somewhat unexpectedly, areas whose residents are high users of hospital beds for non-acute stays (stays of 60 days or longer) also tend to be high users of nursing home resources (Figure 8); regions which are low users of one also tend to be low users of the others. However, elderly residents of Winnipeg region and Parklands make relatively low use of nursing home resources and relatively high use of hospitals for chronic care.

In the second phase of the development of this information system, per capita costs for health care across the regions will be compared by combining expenditures on hospitals, nursing home and physician services.

### Tracking Indicators Over Time

We plan a limited circulation of the 1991/92 reports of the health information system to those who can provide feedback on the general usefulness and interpretation of our indicators. We will then update the reports annually, always including the two previous years of data tracking relevant indicators.

By tracking indicators over time we hope to monitor the direction in which the system is moving, changes in the intensity with which care is delivered, and changes in health status

over time. Computerized data on the provincial delivery of home care services and drug utilization should be available to the information system within the next two years. This will permit an even more comprehensive examination of usage patterns.

### Investigating the Determinants of Health and Utilization

The information system will be used to understand better the determinants of population health and utilization patterns. An example of the insights one can gain about the relationship of various measures of need and utilization can be offered by a comparison of two candidate measures of need. As noted above, Eyles et al. (1991) have proposed the use of Standardized Mortality Rates as an indirect measure of need for health care and hence as a basis for adjusting capitation rates. Their approach can be compared with one that explicitly relies on the conceptual model presented here and uses a socio-economic risk index to assess need.

A preliminary test of the two approaches was run by segmenting the population into 26 age and sex categories and using the relevant mortality rates and the socio-economic risk index to "explain" variation in hospital separation rates of the eight regions - a key indicator of health care utilization. In doing this, it should be noted that separate mortality rates are available for each region for each of the 26 categories, while the socio-economic risk index, being an ecological measure, relies on the same regional score for each of the 26 groupings. While this would appear to give an overwhelming advantage to mortality rates as an explanatory factor, the results do not bear that out. Table 4 shows the relationships between the two candidate explanatory variables and separations. From a scan of the data it is clear that the socio-economic risk index is a much more consistent predictor: it is significantly related to the hospital separation rate in 24 of the 26 age/sex categories. The value of the adjusted  $R^2$ , the proportion of variation explained, vary from a low of 44.1% to 94.6%, with a median of 75%. On the other hand, mortality rates are significantly related to separation rates in only 9 of the 26 categories. The adjusted proportion of the explained variation ranges from 40.1% to 68.9%, with a median of 49.4%. Thus use of the integrated data base allows us to reach a tentative conclusion that an ecological factor, socio-economic risk, as measured by our index, has a much more powerful explanatory role in explaining a measure of usage/need than does an age/sex specific measure of health status. Future work will be directed at developing adjustment factors based on socio-economic risk factors as a means of adjusting regional utilization patterns for need.

### Discussion

Our commitment to develop this information system was made as part of a three-year contract establishing the Manitoba Centre for Health Policy and Evaluation. The evolution of the system has been influenced by close interaction of Centre staff with administrators and policy makers.

The development of the system has taken longer than expected. Developing data for use by policy makers requires great attention to detail. In particular, stakeholders will challenge aspects which do not support preconceptions. The tendency to produce volumes of indicators and cross tabulations (because the data are in fact fascinating and in the minutia lie many of the key relationships), must be balanced by the commitment to produce a limited set of key indicators on a regular basis. A happy solution to this dilemma has been Manitoba Health's appreciation that detailed analysis can most appropriately be applied as special projects. We have also agreed to use the information system over time to assess the impact of major health reform initiatives. In addition we are developing a separate module focusing on use of mental health services across the province.

The PHIS assesses the health of the population as the cornerstone for challenging and improving the health care system. While the indicators used for measuring health status may not be perfect, they are clearly sensitive enough to pick up marked differences in the health of residents across Manitoba's regions. Across a wide variety of measures, we find big and consistent differences documenting the fact that residents of Thompson and Norman regions have poor health status. The strong (statistically significant) differences in mortality rates we have documented across Manitoba regions can be compared with the failure of many popular and expensive medical therapies to show a positive impact on reducing all cause mortality rates. For instance, neither cholesterol lowering nor breast screening mammography have ever been shown to reduce overall mortality in a population (Rossouw et al., 1990; Holme, 1990; Schmidt, 1990)<sup>9</sup>. Future analyses will systematically attempt to estimate the strength of the relationships outlined in Figure 1 including the relationship between background factors, health status, supply of services, practice patterns and health care use over time.

What conclusions can be drawn to date? First of all, the Canadian system which provides first dollar coverage of health use for hospitals, nursing homes, home care and physician services appears to work well. Individuals across the province regularly make themselves available to the health care system for the detection and treatment of disease. At the regional level, the Manitoba system appears to be in part needs-driven. The two regions whose residents have the highest levels of hospital use (the northern, geographically isolated Thompson and Norman regions) are also those whose residents have the poorest health and the highest scores on the socio-economic risk index.

In the case of Thompson, residents make heavy use of the system despite having only an average availability of hospital beds in the region and a most limited supply of physicians. The supply picture for Norman residents is quite different. While much of their high use may be related to need, this region also has more hospital beds available per capita than any other region, and a rich supply of physicians (9.6 physicians per 10,000 population versus 7.0 in rural Manitoba more generally). In future analyses we will assess what proportion of use appears need driven and what proportion associated with supply of services and/or practice patterns.

One must not push the argument presented here too far. Clearly the health care system is supposed to provide care for those who are ill and Manitoba's system provides a remarkably high level of care for residents of disadvantaged regions. However, given the strong relationship between socio-economic risk factors and health status and usage, it is important to ask whether high use of the health care system is the most effective way to ensure access to health? Clearly high use does not guarantee health. The data presented here are cross-sectional and a planned longitudinal analysis will help to sort out whether high health expenditures over time are associated with improved population health. However, current health care expenditure patterns may drive out the ability to fund other programs and appear unlikely to resolve the underlying problems creating poor health. Figure 9 outlining the Wise Men of Chelm's solution to kids falling off ledges - "Build a Hospital", has an important message. While it would be quite wrong to suggest that the health problems of Norman and Thompson residents be ignored, the data raise fundamental questions about the role of the health care system in improving the health of the population. The challenge now is to convince ourselves it is safe to decrease our investments in hospitals and to figure out how to build effective "guardrails".

Much of the effort in the U.S. and Canada to date has been on fine-tuning the existing system, for example, developing practice guidelines to make sure physicians use more effective treatments and improving the efficiency with which hospitals treat patients. The analyses presented here suggest that such an approach will have limited impact because it is based on the implicit assumption that the health care system itself is the most powerful determinant of

health. Our analyses suggest this is simply not true. By measuring socio-economic risk status and demonstrating its strong link to population health, an independent set of social levers outside the health care system is identified for policy makers. Such levers and the information base to support the use of such levers are important for breaking the cycle of the "medicalization of social ills" (Hurowitz, 1993).

The implications of our analyses are straightforward: to improve the health of the population, resources must be reallocated from health care to activities which more directly affect the determinants of health. But where to direct these funds? Into social policy aimed at improving the standard of living for individuals (sewage treatment and better housing on the Reserves is a long unanswered call)? Into early childhood intervention programs in an attempt to enhance development (Grantham-McGregor et al., 1991). Into the private sector to create meaningful jobs so that families can become more functionally viable? Now is clearly the time to direct funds away from basic bio-medical and drug-based research towards both fundamental and applied research on the determinants of health from a broad social policy perspective. While the PHIS cannot specify where these funds should go - it will make it possible to track, long term, their impact on the health of the population.

There are many ways in which the analyses presented here could be used by policy makers. For example, provinces are now making difficult decisions about closing hospital beds. Should one close beds equally across the system? Or should one look at regions according to the health of the residents and according to how efficiently, particularly healthy regions, are operating with regards to hospital use? Closing beds in high use healthy regions would be the type of policy option suggested by PHIS.

An information system which looks both at specific health care sectors and across health care sectors is an invaluable tool for decision-making. It not only permits a look at one sector but uncovers inter-sectional relationships which can help to identify alternatives. It also permits the identification of trends which can be encouraged or discouraged early in their development.

Finally, this information system offers policy makers and the public the ability to separate issues of health from other factors. That is, whenever cutbacks are proposed in the health care sector or there are purchases which are denied, the sceptre of threats to health of the population always looms heavy. The proposed approach can target measurement of health across populations in the areas for which reductions are proposed - are current health care expenditure patterns tightly linked to good health outcomes, particularly in otherwise healthy areas? If not, it should be possible to separate health sector employees' concerns about income reduction, layoffs and job disruption (and treat them legitimately as should concerns by employees laid off in any sector be treated) from issues as to whether such actions threaten the health of the population.

## REFERENCES

- Black C, Roos NP, Burchill C: Utilization of hospital resources. 1993. Manitoba Centre for Health Policy and Evaluation.
- Carstairs V and Morris R: Deprivation and Health in Scotland. Aberdeen University Press, 1991.
- Cassel J: The contribution of the social environment to host resistance. *Am J Epidemiol* 1976;104:107-123.

Charlson ME, Pompei P, Ales KL, MacKenzie CR: A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis* 1987;40:373-383.

Charlton JRH, Hartley RM, Silver R, Holland WW: Geographical variation in mortality from conditions amenable to medical intervention in England and Wales. *Lancet* 1983;696-699.

Cohen M, MacWilliam L: Population health: Health status indicators. 1993. Manitoba Centre for Health Policy and Evaluation.

DeCoster C, Roos NP, Bogdanovic B: Utilization of personal care home resources, 1993. Manitoba Centre for Health Policy and Evaluation.

Evans, RG and Stoddart GL: Producing Health, Consuming Health Care. *Soc Sci Med* 1990;31:1347-1363.

Eyles J, Birch S, Chambers S, Hurley J, Hutchison B: A needs-based methodology for allocating health care resources in Ontario, Canada: development and an application. *Soc Sci Med*, 1991;33:489-500.

Frohlich N, Mustard C: Socio-economic status and Health: a preliminary regional analysis. 1993. Manitoba Centre for Health Policy and Evaluation.

Grantham-McGregor SM, Powell CA, Walker SP, Himes JH: Nutritional supplementation, psychosocial stimulation and mental development of stunted children: the Jamaican Study. *Lancet*, 1991;338:1-5.

Haan M, Kaplan GA, Camacho T: Poverty and health: Prospective evidence from the Alameda County study. *Am J Epidemiol* 1987;125:989-998.

Harris JR, Pedersen NL, McClearn GE, Plomin R, Nesselroade JR: Age differences in genetic and environmental influences for health from the Swedish adoption/twin study of aging. *J Gerontology*, 1992;47:213-220.

Holme I: An analysis of randomized trials evaluating the effect of cholesterol reduction on total mortality and coronary heart disease incidence. *Circulation* 1990;82:1916-1924.

Horgan J: Trends in Behavioral Genetics. Eugenics Revisited. *Scientific American*, June 1993;123:131.

Hurowitz JC: Toward a social policy for health. *New Engl J Med* 1993;329:129-133.

Libertos P, BG Link, and JL Kelsey: The Measurement of Social Class in Epidemiology. *Epidemiological Reviews* 1988;10:87-121.

Mossey JM, Roos LL: Using insurance claims to measure health status: the illness scale. *J Chron Dis* 1987;40:45-50.

- Pappas G, Queen S, Hadden W, Fisher G: The increasing disparity in mortality between socio-economic groups in the United States, 1960 and 1986. *New Engl J Med* 1993;329:103-109.
- Pope GC: Medical conditions, health status, and health services utilization. *Health Services Research* 1988;22:857-877.
- Roos NP: Predicting hospital utilization by the elderly. The importance of patient, physician, and hospital characteristics. *Med Care* 1989;27:905-919.
- Rose G: *The Strategy of Preventive Medicine*. New York, Oxford University Press. 1992.
- Rossouw JE, Lewis B, Rifkind BM: The value of lowering cholesterol after myocardial infarction. *New Engl J Med* 1990;323:1112-1119.
- Schmidt JG: The epidemiology of mass breast cancer screening - a plea for a valid measure of benefit. *J Clin Epidemiol* 1990;43:215-225.
- Shapiro E and Tate R: Survival patterns of nursing home admissions and their policy implications. *Can J Public Health* 1988;79:268-274.
- Shapiro E and Roos NP: Elderly nonusers of health care services. Their characteristics and their health outcomes. *Med Care* 1985;23:247-257.
- Tataryn DJ, Roos NP, Black C: *Utilization of Physician Resources. Volume I: Key Findings*. 1993. Manitoba Centre for Health Policy and Evaluation.
- Weissman JS, Gatsonis C, Epstein AM: Rates of avoidable hospitalization by insurance status in Massachusetts and Maryland. *JAMA* 1992;268:2388-2394.
- Wennberg JE, McPherson K, Caper P: Will payment based on diagnosis-related groups control hospital costs? *N Engl J med* 1984;311:295-300.
- Young TK, Roos NP and Hammarstrand KM: Estimated burden of diabetes mellitus in Manitoba according to health insurance claims: a pilot study. *Can Med Assoc J* 1991;144:318-324.

## ENDNOTES

<sup>1</sup> While we anticipate eventually incorporating a number of key physical environmental indicators into the information system, in the first phase we are focusing primarily on the socio-economic determinants of health and will be relating these to utilization, supply, and health status indicators.

<sup>2</sup> We will demonstrate that a socio-economic risk index may have a better claim than SMR as a measure of need for care.

<sup>3</sup> Using Census data gathered and published by Statistics Canada, we obtained a variety of measures of socio-economic status for residents in each of the eight regions. A set of five poor health and usage indicators which were deemed to be particularly sensitive to differences in socio-economic status were chosen as a test index against which to measure the explanatory power of candidate socio-economic indicators. The measures most highly associated with our poor health status index were: the percentage of the population between the ages of 25 and 34 having graduated high school, the percentage of the labour force between the ages of 15 and 24 unemployed as well as the percentage between ages 45 and 54 who were unemployed, the percentage of single parent female headed households, the percentage of females participating in the labour force, and the average dwelling value. The test health status index included admission rate to hospitals of females for injuries, admission to hospital of male for injuries, admission to hospital of children age 0-4 for respiratory infections, admission to hospitals of persons greater than age 65 years for respiratory infection, and fertility rates. When fertility was removed from the test index - in response to observations that it was not a measure of "poor health" - essentially similar results were obtained.

<sup>4</sup> We are currently calculating accurate inpatient costs for every Manitoba hospital which will allow us to assess the average cost per day of care received by regional residents (taking into account the mix of institutions used by residents as well as type of care received in each institution).

<sup>5</sup> Comorbidity refers to medical conditions existing in addition to the main reason for hospitalization, usually recorded as the "most responsible diagnosis" on hospital discharge abstracts. The type and number of specific co-morbid conditions provide an indication of the health status (and risk of death) of patients (Charlson et al. 1986).

<sup>6</sup> The federal government funds seven nursing homes for Status Indians in Manitoba, comprising 206 beds. Information from six of these homes is not available to us. Therefore, the nursing home rates will be understated by approximately 2%.

<sup>7</sup> These figures are based on those practitioners receiving payments of \$40,000 per year or more and excludes radiologists, pathologists and anaesthesiologists most of whom practice in Winnipeg but do not deliver ambulatory care.

<sup>8</sup> Thompson, the region with the lowest proportion of the population in contact with a physician (77.7%), receives a certain amount of primary care delivered by nurses in nursing stations. These contacts (and costs) are not reflected in our data.

<sup>9</sup> Thus screened women who are "saved" from dying of breast cancer appear to die just as early as unscreened women - but of another cause. At least there are no data to date demonstrating an impact of screening on overall mortality.

TABLE 1 - REGIONAL USE OF HOSPITAL RESOURCES<sup>1</sup>:USE OF SHORT STAY<sup>2</sup> INPATIENT CARE

1991/92

	Central	Eastman	Interlake	Norman	Parkland	Thompson	Westman	Winnipeg	Non-Winnipeg Comparison	Manitoba
Number of residents	94,484	85,180	71,936	24,952	46,056	45,019	117,724	655,055	485,351	1,140,406
Active treatment beds per 1000 population located in region	4.3	2.4	2.8	7.7	6.8	4.8	6.4	4.9	4.7	4.8
Number of persons hospitalized per 1000 population	101	98	99	142	123	159	106	75	110	90
Number of episodes of hospital care per 1000 population <sup>3</sup>	140	139	139	210	184	231	149	99	157	124
Number of hospital separations per 1000 population	149	148	148	224	195	263	160	101	168	130
Average length of stay per hospital separation	6.3	6.1	6.4	5.4	6.3	4.4	6.6	7.3	6.1	6.6
Number of days of hospital care per 1000 population	920	946	938	1351	1108	1532	985	743	1018	861

<sup>1</sup>All rates (except for average length of stay per hospital separation and bed ratios) are age- and sex-adjusted to the Manitoba population using an indirect method of standardization.

<sup>2</sup>Stays of 1-59 days are defined as short stays.

<sup>3</sup>An episode of hospital care is defined as a continuous period of hospital care which may involve one or more transfers between facilities.

TABLE 2 - UTILIZATION OF PERSONAL CARE HOME RESOURCES, 1991/92:

AGE 75 +<sup>1</sup>

	Central	Eastman	Interlake	Norman	Parklands	Thompson	Westman	Winnipeg	Non- Winnipeg Comparison	Manitoba
Population (Age 75+)	6,223	3,826	4,038	811	4,108	497	9,761	36,488	29,264	65,752
PCH beds per 1000 population	121	118	125	160	122	52	144	128	130	128
Residents of PCH per 1000 population (number)	130.6 (834)	136.2 (489)	137.3 (526)	161.3 (115)	120.0 (496)	87.2 (38)	142.0 (1,471)	130.8 (4,747)	134.8 (3,969)	132.6 (8,716)
Admissions to PCH per 1000 population (number)	28.8 (183)	24.4 (90)	31.8 (123)	26.6 (20)	32.2 (134)	19.3 (9)	27.1 (277)	27.1 (983)	28.3 (836)	27.7 (1,819)
Days of PCH care per resident of region	37.9	39.6	39.0	48.8	33.0	23.3	41.9	38.2	39.0	38.6
Estimated costs of PCH care per resident of region (\$)	2,903	3,081	3,060	3,762	2,495	1,766	3,144	3,018	2,977	3,000
Estimated costs per PCH day (\$)	77	78	78	77	76	76	75	79	76	78

<sup>1</sup>Rates are age- and sex-adjusted to the Manitoba population using an indirect method of standardization. All are based solely on the population aged 75 years and older.

TABLE 3 - REGIONAL USE OF PHYSICIAN SERVICES  
ALL VISITS 1991/92

	Central	Eastman	Interlake	Norman	Parklands	Thompson	Westman	Winnipeg	Non- Winnipeg Comparison	Manitoba
Physicians per 10,000 Population	5.1	5.0	6.0	9.6	8.7	5.1	11.0	14.9	7.0	12.7
Percentage of individuals with contact (%)	81.0	81.3	83.8	81.0	80.9	77.7	82.4	85.2	81.4	83.6
Visits per resident	4.1	4.5	4.9	5.2	4.4	4.7	4.7	5.3	4.6	5.0
Visits per user	5.1	5.5	5.8	6.4	5.5	6.0	5.7	6.2	5.6	6.0
Dollars per visit (\$) <sup>1</sup>	21.0	22.1	21.9	20.8	21.0	22.4	21.4	23.3	21.5	22.6
Costs of physician visits per resident (\$)	87.0	98.8	106.8	108.0	93.4	105.0	100.1	123.0	97.9	112.5

<sup>1</sup>For this measure as well as others in this table, only visits per se are included not the cost of laboratory services or any other procedures performed in office.

TABLE 4 - COMPARISON OF THE SOCIO-ECONOMIC RISK INDEX AND MORTALITY RATES  
AS EXPLANATIONS FOR HOSPITAL USE RATES

CATEGORY	SOCIO-ECONOMIC RISK INDEX			MORTALITY RATES		
	ADJUSTED R <sup>2</sup>	P-VALUE	SLOPE ß	ADJUSTED R <sup>2</sup>	P-VALUE	SLOPE ß
0-4 Female	0.292	0.096	34.11	-0.033	0.411	49.20
0-4 Male	0.441	0.043	47.68	0.489	0.032	127.04
5-9 Female	0.602	0.014	7.88	-0.053	0.451	15.47
5-9 Male	0.475	0.035	7.36	-0.038	0.421	15.56
10-14 Female	0.537	0.024	12.80	0.400	0.055	84.25
10-14 Male	0.642	0.010	7.37	-0.119	0.632	-16.58
15-19 Female	0.946	0.000	41.45	0.685	0.007	143.53
15-19 Male	0.768	0.003	8.80	0.476	0.035	13.85
20-24 Female	0.898	0.000	49.85	-0.098	0.563	-119.84
20-24 Male	0.827	0.001	14.38	0.610	0.014	22.87
25-34 Female	0.734	0.004	27.11	0.131	0.202	76.43
25-34 Male	0.782	0.002	12.37	0.573	0.018	18.92
35-44 Female	0.887	0.000	21.01	-0.072	0.495	39.11
35-44 Male	0.914	0.000	14.51	0.104	0.226	18.80
45-54 Female	0.856	0.000	25.32	0.254	0.116	35.85
45-54 Male	0.785	0.002	15.34	-0.097	0.561	12.98
55-64 Female	0.911	0.000	52.92	0.423	0.048	28.60
55-64 Male	0.445	0.042	26.11	0.494	0.031	18.38
65-74 Female	0.898	0.000	68.35	0.630	0.011	23.06
65-74 Male	0.527	0.025	41.74	0.496	0.031	10.36
75-79 Female	0.840	0.001	103.68	0.200	0.149	20.39
75-79 Male	0.756	0.003	71.43	0.314	0.086	8.51
80-84 Female	0.672	0.008	114.44	-0.068	0.485	8.55
80-84 Male	0.754	0.003	91.37	-0.003	0.360	11.62
85+ Female	0.396	0.056	71.19	-0.131	0.677	1.38
85+ Male	0.776	0.002	133.44	-0.149	0.772	-1.24

Figure 1. THE CONCEPTUAL MODEL UNDERLYING THE  
POPULATION HEALTH INFORMATION SYSTEM

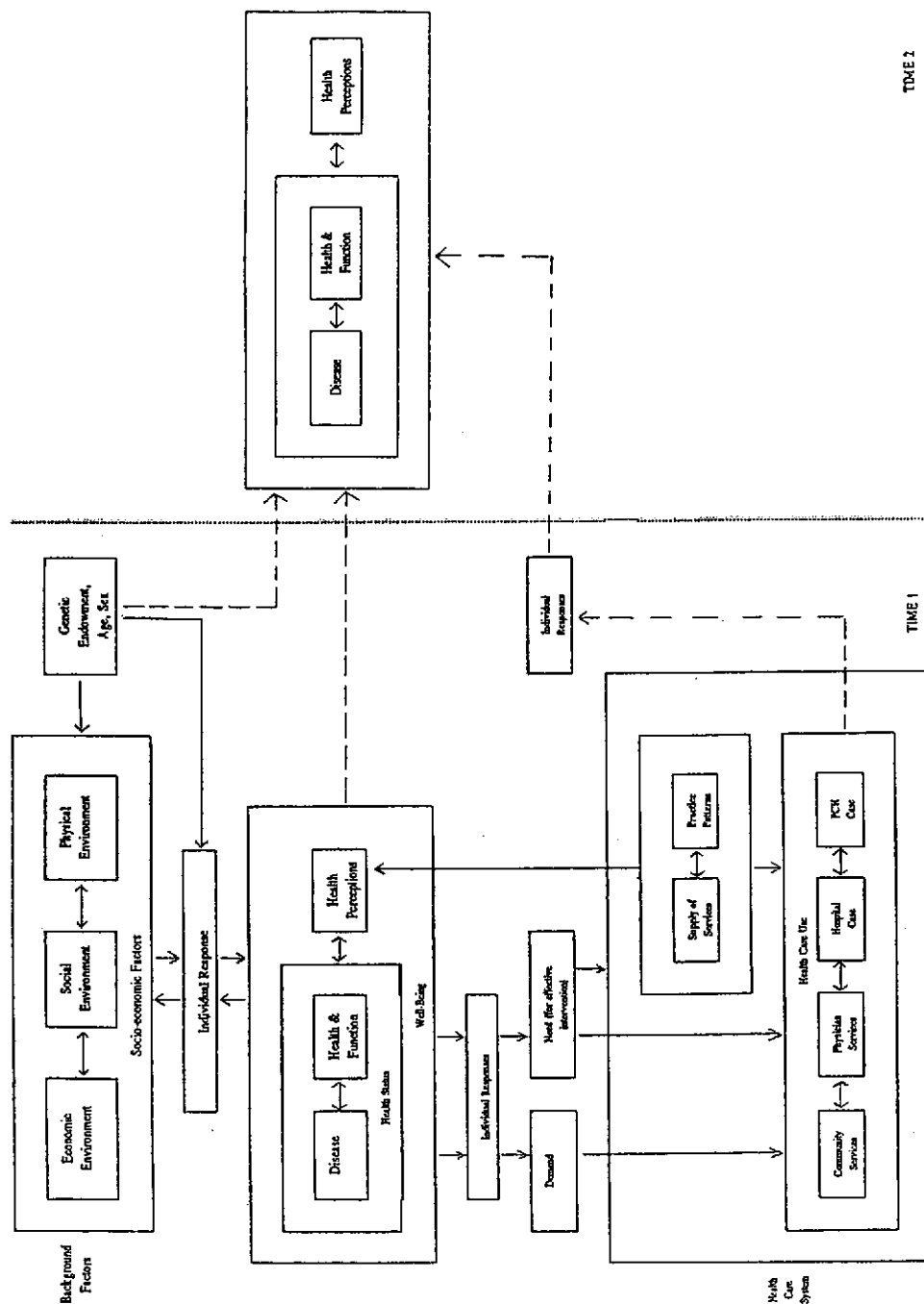


Figure 2

### THE MANITOBA POPULATION HEALTH INFORMATION SYSTEM

#### Organizing Principles

- Population based
- Describe:
  - supply
  - access to care
  - intensity of use
  - discretionary use
- Assess contributions to costs per capita of:
  - differential access
  - \$ per service (visit, day of care)
  - services per user
- Relevant to managers of system
- Sum use across sectors
- Create regional profiles

#### Modules of System

- Population/Health/Health Status Indicators
- Socio-economic Characteristics
- Utilization of Hospital Resources
- Utilization of Personal Care Home Resources
- Utilization of Physician Resources
- Demographic Patterns

Figure 3 - STANDARDIZED MORTALITY RATIO FOR PERSONS  
AGED 0-64 YEARS

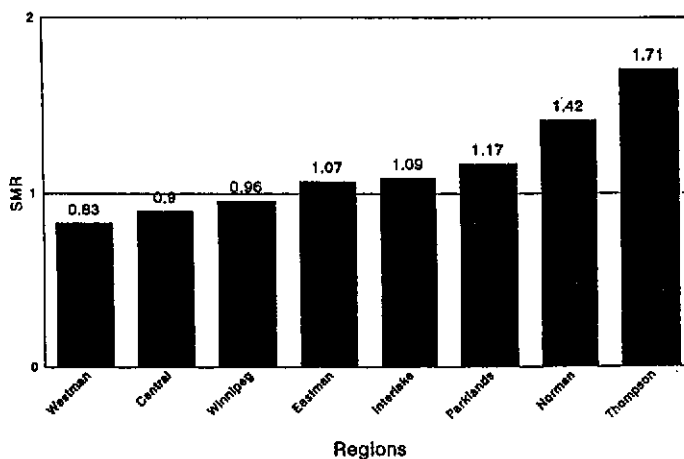


Figure 4

Standardized Mortality Ratio  
Across Regions of Manitoba  
Age 0 – 64

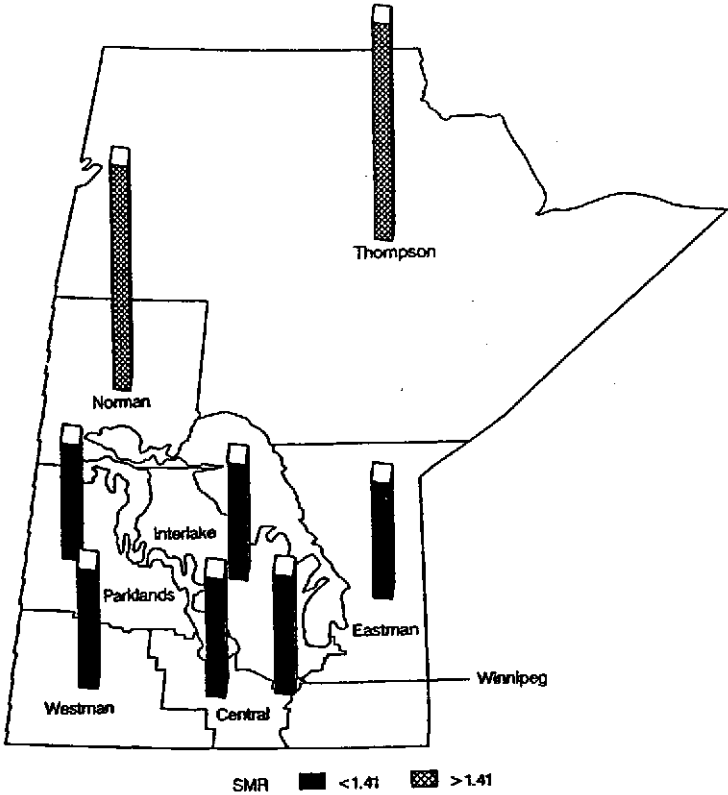
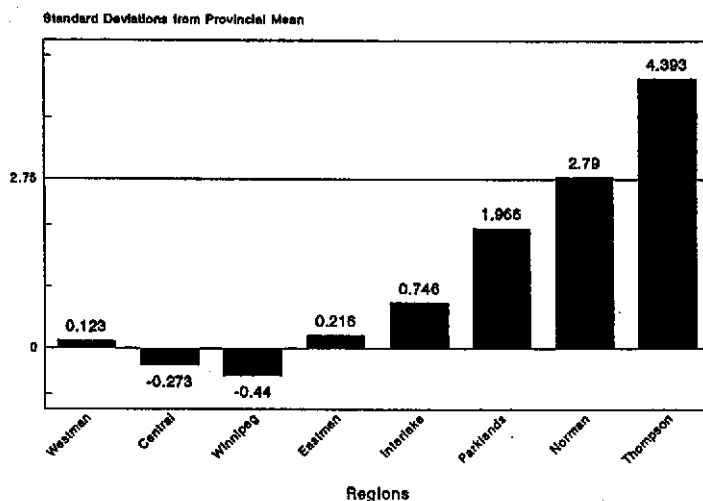


Figure 5 - REGIONAL SCORES ON THE SOCIO-ECONOMIC RISK INDEX\*



A high score on the index means poor socio-economic status (SES) relative to the province. Regions with scores of 2.75 or greater have significantly poorer SES than that of province (alpha = .05)

Figure 6 - ACUTE HOSPITAL DAYS PER 1,000 RESIDENTS BY LEVEL OF CARE

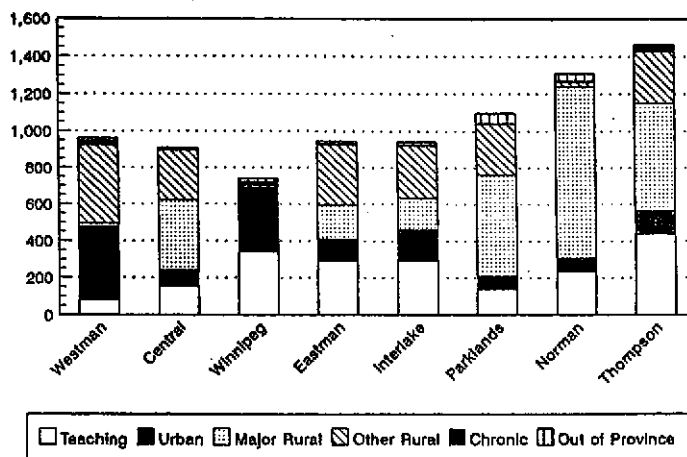


Figure 7 - PROPORTION OF ACUTE HOSPITAL DAYS USE  
According to Level of Co-Morbidity

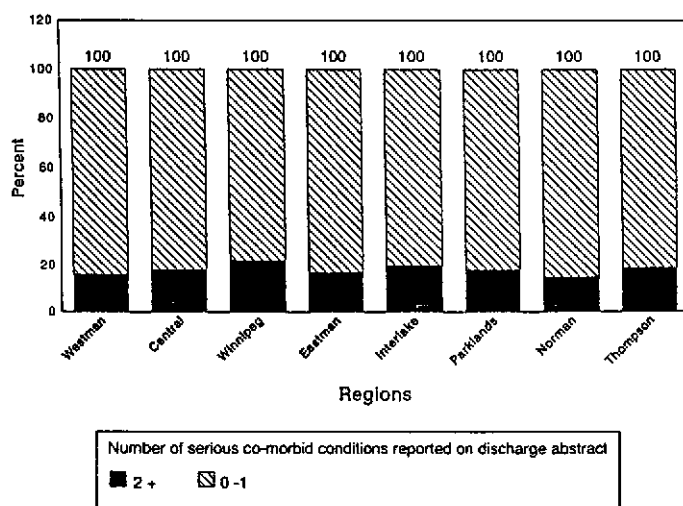
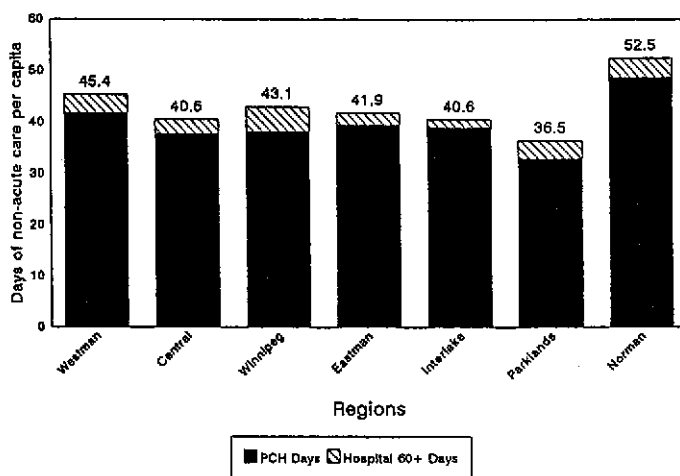


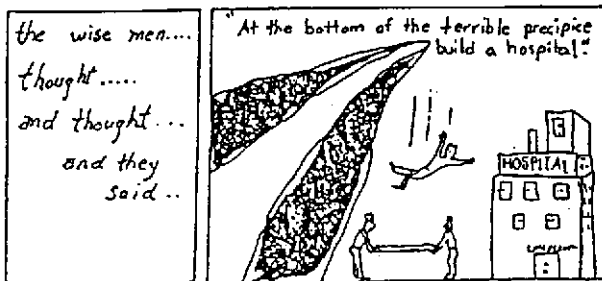
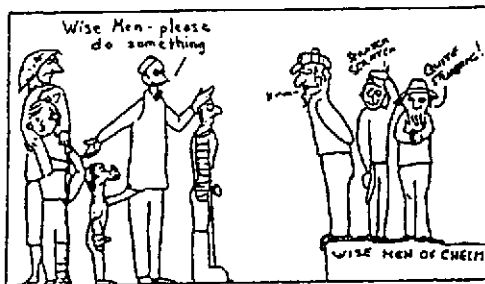
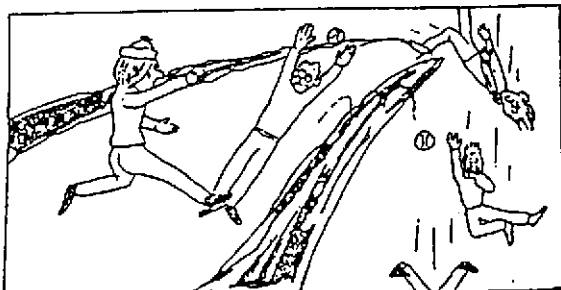
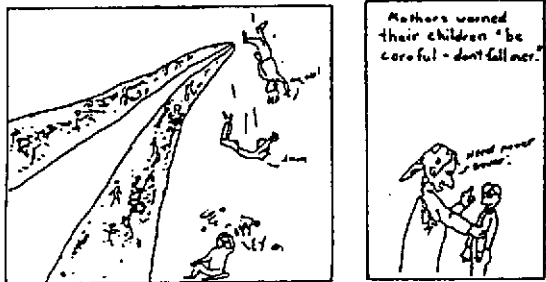
Figure 8 - Use of Nursing Homes and Hospitals (60+ Day Stays\*)  
(Age 75+)



\*Excluding Day 1 to Day 59. Due to the small number of elderly (497) and missing data, Thompson is excluded from analyses.

Figure 9 The Wise Men of Chelm

In Chelm, there was a very good road.  
It came to a sudden stop and beyond  
was a precipice.



# **APPENDIX I** **INFORMATION SYSTEM COMPONENTS**

## **HEALTH STATUS**

### **I. Overall Mortality Rates**

- age and sex adjusted standardized mortality ratio for the population aged 0-64 years
- age specific mortality rates for females and age adjusted rates
- age specific mortality rates for males and age adjusted rates
- age specific mortality rates total and age and sex adjusted

### **II. Infectious Diseases**

#### Mortality Rates for:

Pneumonia  
AIDS

#### Individuals Hospitalized per 1000 Residents for:

Pneumonia  
Influenza  
Tuberculosis  
Sexually Transmitted Diseases  
Hepatitis  
AIDS

### **III. Injuries**

#### Mortality Rates for:

Motor Vehicle Accidents  
Falls  
Drownings and Suffocation  
Poisonings  
Fire & Flames  
Suicide  
Homicide  
Other Injuries  
All Injuries

#### Individuals Hospitalized per 1000 Residents for:

Motor Vehicle Accidents  
Falls  
Drownings and Suffocation  
Poisonings  
Fire & Flames  
Vehicular Non-traffic  
Attempted Suicide  
Attempted Homicide  
All Injuries

### **IV. Cancer**

#### Mortality Rates for:

Cancer lung  
Cancer Bladder  
Cancer Kidney  
Cancer Colon

#### Individuals Hospitalized per 1000 Residents for:

Cancer Lung  
Cancer Bladder  
Cancer Colon  
Non-Melanoma Skin Cancer

Cancer Breast  
All Cancers

Cancer Breast  
Cancer Kidney  
Other Cancer  
All Cancers

V. Reproductive Health of Women

- Rate of low birthweight infants

VI. Chronic Disease

Mortality Rates for:

Asthma  
Vascular Disease Complications  
Diabetes  
Hypertension  
Ischemic Heart Disease  
Emphysema

Individuals Hospitalized per 1000 Residents for:

Asthma  
Vascular Disease Complications  
Diabetes  
Hypertension  
Ischemic Heart Disease  
Emphysema

VII. Indicators Which Reflect on Health Care System

Mortality Rates for:

Conditions amenable to medical treatment  
Events for which single occurrence  
is cause for concern  
Events for which high rates are  
cause for concern

Individuals Hospitalized per 1000 Residents for:

Conditions amenable to medical treatment  
Events for which single occurrence  
is cause for concern  
Events for which high rates are cause  
for concern  
Ambulatory care sensitive conditions  
Conditions for which hospitalization is avoidable

VIII. Disability and Functional Limitations - Indicators based on Physician Contacts

Measures of Disability Among Youth

Rate of individuals aged 24 years or less who had 2 or more visits to a physician over the course of one year with the following types of conditions:

- emotional disturbances
- hydrocephalus
- spina bifida
- hearing loss
- cerebral palsy
- cystic fibrosis

- developmental delay
- hearing loss

Medical conditions most reflective of functional limitations. Rates calculated separately for total and elderly population. Based on individuals with one or more visits to a physician with the following types of conditions.

- musculoskeletal
- other respiratory
- other heart condition
- other gastrointestinal condition
- nervous conditions
- hypertension
- endocrine conditions
- ischemic heart disease
- ill-defined conditions
- other circulatory conditions
- mental disorders
- cerebrovascular conditions
- urinary tract conditions
- upper respiratory conditions
- fractures
- sprains

#### INDICATORS OF SOCIAL WELL-BEING

Mean household income  
 Average dwelling value  
 Household housing cost  
 Owner occupied dwellings  
 Unemployment rate (by age)  
 Level of education (by age)  
 Single parent household  
 Rate of single elderly female headed households (high risk nursing home entry)  
 Mean age dependency ratio  
 Native mother tongue  
 Treaty Indians (group with high health needs)  
 Rate of births to females aged 15-19  
 Region in-migration  
 Labour force employed in farming  
 Labour force employed in manufacturing, construction and transportation  
 Female labour force participation

## RESOURCE USE

### I. Hospitals

Beds/1000 Population

Rates of Use

- Individuals
- Separations
- Hospital days
- Length of stay

Characteristics of Use

- Acute vs. chronic (60 days or longer)
- Age and sex
- Medical/surgical/pediatric/obstetrical/psychiatric
- Complexity of Care (presence/absence of co-morbidity and complications)
- Intensity of resource use (diagnostic group weights)
- Type of hospital (teaching - small rural)
- In region/out of region/Winnipeg
- Discretionary admissions (high variation medical/surgical, low variation, obstetric)

### II. Nursing Homes

Beds/1000 elderly aged 75+ years

Rates of Use

- Individuals resident during year
- Admissions during year
- Days
- Cost (mix of levels of care)

Characteristics of Use

- Age
- Expected length of stay (influenced by age, sex, level of care at entrance)
- Type of home (proprietary - non; ethnic/religious-non; freestanding - juxtaposed)
- From where admitted (hospital/community)
- Level of care

### III. Physicians - Ambulatory Care

Physicians/10,000 population

Rates of Use

- Individuals with contact
- Visits
- Visits per user
- Costs (mix of consults and specialists)

#### Characteristics of Use

- Type of visit (unreferred/consult/outpatient or emergency room)
- Type of physician seen (specialist - medical/surgical/generalist, and by specific specialties)
- Age and sex
- In-region/out of region/Winnipeg
- Severity of diagnosis
- Rate of contact in year (0 visits/1-6/7-14/15 visits or more)

#### Other Care

- Chiropractic Contact
- Optometrist Contact

### IV. Cross Sector Summaries of Use

#### Total Days

- hospitals (acute, chronic) and nursing home

#### Total \$

- Hospital/nursing home/physician services

# Discussion of Papers by Johan Mackenbach and Noralou Roos: Health Care as a Determinant of Health

---

*Patricia A. Baird*

**T**his session is entitled "Health Care as a Determinant of Health". So as a sensible starting approach, we need to define what we mean by the words "health care" as well as "health". I won't touch the latter with a barge pole - it is a well known black hole - but I want to note what we all here know - that health care and medical care are not synonyms, even though they are quite often used interchangeably.

Health care could be interpreted to permeate all aspects of daily life and include the prevention of illness. It would include such things as good housing and sanitation, a safe work environment, good interpersonal relationships, sufficient income and education. Medical care is only one aspect of health care broadly defined this way, and it centres on the diagnosis and treatment of disease after it has developed. The topic that the two speakers in fact addressed was not health care as a determinant of health but "medical care as a determinant of health".

It is not easy to make a clean separation of medical care and health care, even if we use the term medical care to refer to the illness care system provided by physicians, nurses and other specially trained personnel. Some of what they do is preventive and not simply responsive to illness. But in trying to isolate medical care as a parameter to assess, it may work as a first approximation to include all such services.

But perhaps there is a lesson in here - that in fact you should not be considering one (that is medical care) without the other (that is health care). Only if you simultaneously take into account the length and the breadth of a playing field can you know its area - and perhaps you must simultaneously know about health care in a population to get any meaningful assessment of the role of medical care. They are likely to be interactive and in fact related to each other, so perhaps we could benefit from framing a different question - how does health care (as I broadly defined it) relate to medical care?

It may be expected that as health care is better, medical care will be appropriately needed to produce health for individuals less often. But paradoxically, medical care may be much more important for those individuals who have become ill in these good conditions, than for those who become ill when social conditions are abysmal.

The relationship is likely to be complex and iterative with the level of health care in a population importantly influencing the place of medical care.

It is rather like the old chestnut - does nature or nurture cause a particular condition in a person? The answer is both are always necessary - and you can only know the role of one if you take into consideration the other.

To return to our theme, I think that our two speakers have shown very clearly how difficult it is to arrive at well founded, specific data to answer the question rigorously - and part of the difficulty may be because we need simultaneously to know more about both aspects - health care and medical care.

For example, Dr. Mackenbach talked about time trends in health in a population related to the available data bearing on the contribution of environmental change or medical care.

In discussing McKeown's insights he made the important point that the effect of environmental changes were evaluated by McKeown over a time period when mortality was high, whereas antibiotic and medical measures were evaluated when mortality was low, so that it was not an equal comparison of their potential effect on population health. The data he gave us seemed to show that in terms of contemporary determinants, medical care has a substantial contribution to the rapid gains of life expectancy observed in the Netherlands in recent years. His conclusion was that it is wise to refrain from any far reaching conclusions, that we are not on very firm ground in largely discounting medical care as a contributor to health.

His paper also demonstrates once again the need for data on morbidity. We have all lamented at times how mortality is a crude measure of health, and as he noted, mortality as a measure of health is an inadequate gauge. It reinforces the importance of building data bases and information systems that can also provide measures of morbidity. His paper led nicely into Dr. Roos's talk as she exemplifies the foremost of several groups in Canada who are trying to build such data bases. Dr. Wolfson, for example, will be discussing other approaches as well.

It is the imaginative and insightful use of such complete data sets on a population that may help us eventually to answer the question of how important medical care is as a determinant of health. Dr. Roos said that the data showing the relationship of low SES and poor health gives an independent set of social levers outside the health care system for policy makers. I'd like to hear how much people agree with that, because it is not new information that low SES is linked to poorer health, and I wonder whether that knowledge, in and of itself, would be terribly useful to bring about policy change. A difference is, here the data is more specific and perhaps it can be used as ammunition, especially if it is made widely available and if there is resultant political will to change.

But she has described a potentially very valuable data base that they plan to use and to track over time to answer the question we are addressing. She has shown how they are going to take into account both variables, that is - medical care and health care - using various indicators of medical care use but also measures of vulnerability of the population such as education level, unemployment, single parent households and average dwelling value. For example, would health status in two regions similar with regard to health care broadly defined (and as measured by her parameters) remain the same if medical care expenditures are decreased substantially in one region? This very impressive data base has the potential to ask and answer those kinds of questions. We are all going to watch it with great interest.

In summary both papers show quite clearly that medical care is only part of the picture but that exactly how big a part is very difficult to pin down, and that data bases will need to be built and used to do so.

# Toward a System of Health Statistics

---

*Michael Wolfson*

**I**t is a commonplace that we live in the "Information age". Yet we have inherited a hodge-podge of data collection systems in the health area that are seriously limited with respect to current thinking about health – particularly for the range of determinants of health we have heard discussed today. I am pleased to have the opportunity this afternoon to sketch a vision of health information – a System of Health Statistics. The hope is that we can develop health information in line with emerging views, and thereby contribute much more effectively to our understanding, and ultimately the population's health.

I'd like to start with a quick review of conventional health information, and the main areas where new kinds of information uses are emerging – particularly in terms of a broadened appreciation of determinants of health. This review sets the stage for the next step in the argument – a diagnosis of major problems with our current health data. Of course, any diagnosis is usually followed by some proposed treatment; in this case it is a conceptual framework for a "system of health statistics".

As a preliminary, one semantic point. I shall be using the words "data", "statistics" and "information" frequently. They do not have quite the same meaning. Data are best thought of as the raw material for making statistics and information, by means of analytical effort. In turn, statistics are typically quantitative, and thus a subset of kinds of information.

## **Conventional Health Information**

Health data, particularly vital statistics, are amongst the oldest organized systems of data collection in Canada and in many other countries. Also, Canada has one of the best sets of health data in the world. But we can do much better. Let me review a few major areas of current health information activity briefly, and critically.

The first and perhaps the most common kind of health information is broad macro indicators relating to the health of large population groups. Two of the most widely used macro indicators are health care spending as a percentage of GDP, and life expectancy. These indicators form the basis for virtually all international comparisons. This is not because better indicators haven't been proposed, but rather because the historical importance of the underlying National Accounts, population census, and death registration data make them readily available.

But these are curious indicators of population health – one is a measure of resource inputs, where if we were really healthy as a population, *less* would be better. The other, life expectancy, is essentially an average age at death; it takes no account of the health status of the living.

At a more micro level, institutions like hospitals are major creators of health-related data. As part of their routine operations, they keep track of their numbers of beds and procedures performed, their financial costs, and they produce computerized abstracts for each patient encounter.

With increasing pressures to control costs, hospital funding is starting to be tied to the severity of the case mix of patients, and to adherence to some sort of guidelines for doing surgery only with the appropriate indications. However, data on patients rarely include any systematic socio-economic or psycho-social items, and patient outcomes are rarely followed

after the patient leaves the hospital. Thus, for example, the chief of surgery in a large specialized hospital has no systematic way of knowing if out-of-town patients are experiencing unusually high rates of post-operative complications.

In medicine proper, there is an industry of researchers, substantially funded by governments and pharmaceutical firms, who create and collect detailed information on the effects of myriad interventions. This information is based mainly on the paradigm of randomized experimentation.

However, most of these intervention experiments are very narrowly focussed. They concentrate on specialized end-points in small non-representative populations – for example does drug X affect blood fraction Y by at least P percent in people suffering from disease Z. The essence of randomization is to relieve the experimenter of the need to understand the myriad complex determinants of health. This sort of research is drawn toward “magic bullet” interventions – precisely because the randomization research paradigm is unable to handle a more complex system view that includes, for example, many contending causal pathways, and diverse feedback processes.

Data are voluminous but thin at the population level. For example, we have millions of records each year from hospital and physician visits and from birth and death certificates. But each record tends not to contain much breadth of information, and each record is an isolated event. Many countries also have periodic population health surveys, for example to track self-reported chronic disease burdens and health-related behaviours like smoking. However, these are two data solitudes – with no connections between clinical and vernacular views of health.

Finally, there has been a major growth in the post war period in basic scientific research. One of the most important areas is biochemistry – the mapping of DNA and protein sequences, and efforts to understand the links between these biochemical structures and behaviours of organisms.

### **Emerging Health Information**

The examples just cited illustrate the diversity of health information currently being produced. In contrast, a number of new health information needs are emerging.

One such area is comparative reporting – for example small area geographic variations in surgical procedure rates, and hospital by hospital variations in length-of-stay, appropriately adjusted by severity and other characteristics of the patient case mix.

These information needs require more than isolated visit or encounter data. Data on discrete patient encounters must be strung together to form person-oriented data in order to support even the most basic kinds of analysis. They also require common concepts and definitions, and a broad population perspective. The need for this new kind of information is driven by the completely reasonable (and by no means new) questions, “what actually works in practice, and why”.

One of the central policy questions in the health area at present is how to allocate resources in the most cost-effective manner. Such questions are generating new more probing information requirements. Within a health care organization like a hospital, the question may be the allocation between two different groups of acute care providers – for example cardiology and orthopedics, which in turn may translate into relative numbers of coronary bypass operations and hip replacements. More importantly in my view, this resource allocation issue is starting to be played out at the Cabinet level in provincial governments, for example in the tradeoffs between funding acute care generally and various forms of health promotion programming like support for early childhood development (or reducing the deficit).

At the macro health indicator level, there is a broadening consensus that we have to do better than life expectancy. There are major concerns about chronic disease burdens, and the growing population of frail elderly. There is also debate regarding Fries' "compression of morbidity" hypothesis – essentially whether our increasing longevity is made up of healthy or sickly years of life. Of course, given the evidence you have already heard today on health inequalities, there is growing interest in being able to chart these disparities more reliably.

Health inequalities are not only something to be observed at a point in time. They are also associated with much longer term effects. For example workers with higher incomes turn into retirees decades later with greater longevity (Wolfson et al, 1993 based on Canada Pension Plan data), and children born to parents with lower social status have more frequent behaviour problems at school age, and subsequently as young adults have poorer health (Chris Power, *Class and Health*, from the 1958 U.K. birth cohort sample).

Such long term and significant latent effects in turn raise fundamental questions about causal pathways. What are the roles, for example, of homeostasis / resilience / coping skills / and ideas like Antonovsky's "sense of coherence" as basic explanations? In turn, how can we balance our current preoccupation with disease on the one hand, with information on *positive* health on the other? Perhaps we should be refocussing on health as a dynamic concept, based on repeated longitudinal observations. Then the question would not be "why is Jones sick today?" – rather we could ask "why is Smith always healthy, year after year?" or more generally, what are the determinants of "chronic good health".

Finally, I sense a broadening consensus on the need for reorientation in basic health science research. It is easy to wonder about the relevance to population health of some of the more arcane "rats in mazes" or "yet another pain reliever" strands of health science research – though I hasten to add that these phrases are too simplistic.

In contrast, the significant and long lasting correlations just mentioned between socio-economic status and health problems imply that we must view diseases as long duration processes with a range of determinants (multifactorial, interacting, dynamic) going well beyond medicine and biochemistry.

An intriguing example is psycho-neuro-immunology (PNI) which, from a lay perspective, gives credence to the premise of strong but subtle connections between mind and body. Major advances in understanding may flow from connecting this emerging knowledge on biochemical pathways to population-based surveys, where we ask people for subjective perceptions of their health, and about the psycho-social aspects of their home and workplace milieux.

Personally, I would gladly trade some "rats in mazes" and "pain reliever" research for new kinds of linked biochemical-population survey-based-research. Such reallocation will require changes in our basic health research institutions, and a strengthened role for population-based health information.

### **Diagnoses of Data / Conceptual Problems**

Given this brief review of conventional and emerging health information, there are, in my view, two fundamental problems — over and above the usual complaints that there aren't enough data or that their quality is inadequate.

The first is a pervasive *imbalance*. We know far more in Canada about how much the health care system costs than we do about how healthy the population is. We have regular data on health care spending, life expectancies, and proportions of the population dying of various medically-defined diseases. But do you know if the population is healthier now than it was a few years ago?

Another imbalance is the overly narrow implicit view of the determinants of health. The preponderance of data currently collected on health care derives from the "medical model" of health, and is out of balance with our knowledge of the much broader range of determinants of health.

One important corollary relates to health goals. Ideally, we should define our goals first, and then develop the necessary information systems to monitor progress toward those goals – for example not only adding years to life, but also adding life to years. Unfortunately, I think we sometimes fail to express or even think about such basic health goals simply because we do not have the required data.

The second major problem is a bit more abstract, but equally fundamental. We have sets of health statistics, not a system. Our health statistics suffer from incoherence.

An indication of this incoherence can be gained by perusing any reasonable compendium of health statistics, and for convenience I could pick on Health United States (NCHS, various years) or Statistics Canada's set of health indicators (Statistics Canada, various years). In both cases, the numbers are a hodge-podge. Nothing adds up; there is no integrating framework. One page of Health U.S. has bed-days, another has infant mortality, another numbers of health professionals, and yet another dollars spent. The only thing holding these statistics together is a general pertinence to health, and the binding of the publication itself.

Health clearly rivals the economy in importance, yet in comparison the statistical base is confused, fragmentary, and incoherent. National Accounts numbers for the economy obey a series of arithmetic identities – for example incomes add up to expenditures. Other series like unemployment rates and interest rates are related by equations in various macroeconomic models. The same kind of mathematical structure certainly does not exist for health data.

At a more elementary level, we lack even the basic coherence of common concepts and definitions. In the National Accounts, wages and profits have the same agreed meaning throughout the economy, for example. I am afraid that the same cannot be said for "obstetric visit", "doctor", or "bed-day".

### **Further Desiderata**

Imbalance and lack of coherence are the two most basic problems in our health information system. In addition, there are several more practical premises for a renewed system of health statistics.

Health and health-related activities and phenomena should be measured in their own most natural units. In my view, the central concept should be population health status, based on something like "health status-adjusted" life-years. Such a central concept can then be complimented by elements like health care encounters, economic costs as in the National Accounts, and diseases as classified by the World Health Organization.

A second major desideratum is the need to build on explicit microdata foundations. This means, for example, that data on groups of individuals should be derived by aggregating data on the individuals in each group (or a representative sample thereof). Such microdata foundations are critical to analyses of health inequalities.

This approach may appear self-evident, but it is not the case with the National Accounts, nor with demographic estimates and projections. In each of these existing statistical systems, the smallest level of disaggregation is a group, like an industrial sector or population in a given age interval. The natural units of observation, firms or individuals respectively, are not accessible. The semi-aggregate character of these economic and demographic statistical systems is understandable, given that they pre-date the revolution in computing. However, from the perspective of contemporary computing power and the analytical benefits of explicit

microdata foundations, it is an unfortunate limitation that should be avoided in the development of new statistical systems.

A high quality information system requires both high quality data input, and energetic sophisticated information users. Demanding users may be a source of discomfort, but they are vital to long term stable funding (e.g. Pommier, 1981, discussing French experience with Satellite Accounts). Data providers who are also users of the resulting information, physicians for example, will be much more sensitive to questions of data quality. Thus, an effective health information system should connect data providers and information users.

Let me note one final major desideratum – taking full advantage of modern informatics. A new system of statistics should be much more than a series of print publications. Its full incarnation should be electronic – a database, combined with tailored retrieval and analytical software (subject always to constraints imposed by considerations of privacy and confidentiality). The ideal should be one of offering the user an interface whereby s/he can ask questions and get answers. Of course, many users will still want printed results, but these should be derivative.

### **Theories and Concepts of Health**

In thinking of a new or substantially redesigned health information system, we must ask what is important to measure, and how it ought to be measured. Answers to these fundamental questions derive from a theoretical or conceptual framework, even if implicitly.

For example, theories in astrophysics explicitly provide the basis for the development and construction of specific kinds of radio-telescopes. Similarly, in the social science domain, Keynesian macroeconomic theory played a central role in the development of the System of National Accounts.

Of course, there is no uni-directional causality here – from concepts and theory to a measurement system. Rather, the history is one of iteration back and forth between theory and concepts on the one hand, and observation on the other.

In the case of human health, we are at a point where there is enough accumulated evidence on the broad range of determinants of health to signal the need for major changes in theoretical and conceptual perspective, and consequential changes in approaches to observation and measurement.

To give some flavour to these empirical results and findings, we know for example that cholesterol is connected to CHD – but also to many other important metabolic pathways; that social support and incomes (Wolfson et al, 1993) are connected to longevity – with decades-long latencies; and that there are specific “windows” in early childhood when key opportunities for physiological development are present. These kinds of diverse empirical nuggets all have some superficial connection with one another – for example they are all pertinent to health.

However, we have no grand theories that knit these diverse empirical findings together, or that have driven observational developments in a way similar to those in national accounting, or astrophysics.

A system of health statistics should be based at least upon a conceptual framework (if not a grand theory) that is broad enough to encompass all these kinds of observations. It should also include an associated “inferential apparatus” that can support drawing out quantitative implications. I’ll come back to the notion of inferential apparatus later.

### **The Health Information Template**

There is, of course, a rich and significant history of conceptual frameworks for health. An excellent historical review and new, broader conceptual framework based on the determi-

nants of health has been developed by my CIAR colleagues Bob Evans and Greg Stoddart (Evans and Stoddart, 1990).

Given the ferment of new ideas in Canada, represented in part by the work of members of the CIAR, and a broadening appreciation of the malaise in our current health information system – the imbalance and incoherence mentioned earlier, a National Task Force on Health Information was organized about three years ago, under the direction of the former Chief Statistician of Canada, Martin Wilk (Wilk, 1991). Very early in the work of the Task Force, it was agreed that a “roadmap” or conceptual framework setting out the territory to be covered was essential.

Peter Glynn, then at Health and Welfare, and I embarked on this task. We convened an ad hoc working group under the aegis of the Task Force, and started from the Evans and Stoddart analysis. The result is something called the Health Information Template (Wolfson, 1992). I’ll be showing you some of the Template images in a moment.

The Template has served a variety of purposes. One is simply as a pedagogical device, to assist the Task Force in its cross-country consultations to explore and hopefully reach consensus on what was to be considered part of the field of health information. You can appreciate the diversity of views we encountered.

A second purpose of the Template is to provide the beginnings of a classification structure for the myriad kinds of data, statistics, and information under discussion.

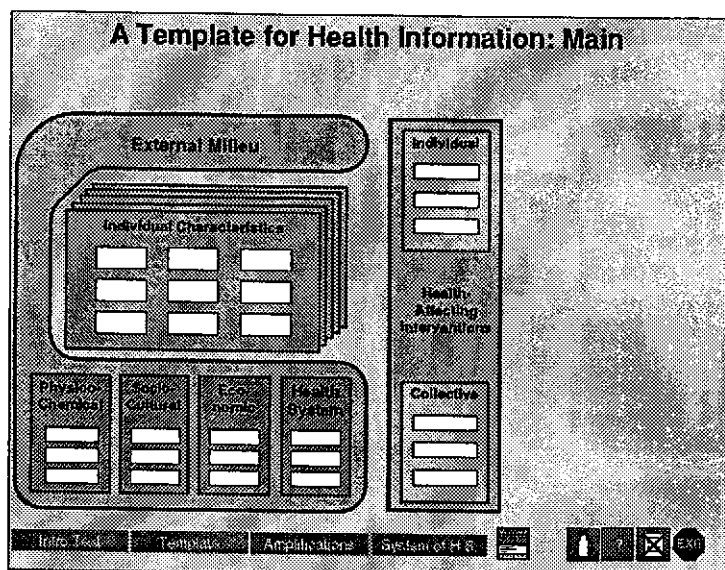
Finally, the Template provides the core of a conceptual framework – not only a way of indexing various kinds of information; but also a structure for indicating how different pieces are connected to one another.

Given the variety and complexity of the classification structure and conceptual framework we developed, the Health Information Template follows one of the desiderata mentioned earlier – to use modern informatics. The full Template is realized as interactive graphical software using a series of visual metaphors – in a McLuhanesque way, the medium is a key part of the message. (A freely copyable diskette with the Template software can be found in Wolfson, 1992.) However, I only have time to show you a few “stills” from the Template software.

### **The Template Image**

Figure 1 shows a monochrome version of the basic Template image. It is the “roadmap” used by the National Task Force on Health Information. In its roles delimiting the field of health information, and as classification structure, the Template divides the health field into three broad domains. At the center of the image (deliberately so) is a population of individuals and their characteristics. The visual nuance of a set of files is also deliberate — it indicates that the data for the population must be at the individual level, i.e. take the form of microdata. Health outcomes are also central, though not immediately apparent in this image; they will come in a moment.

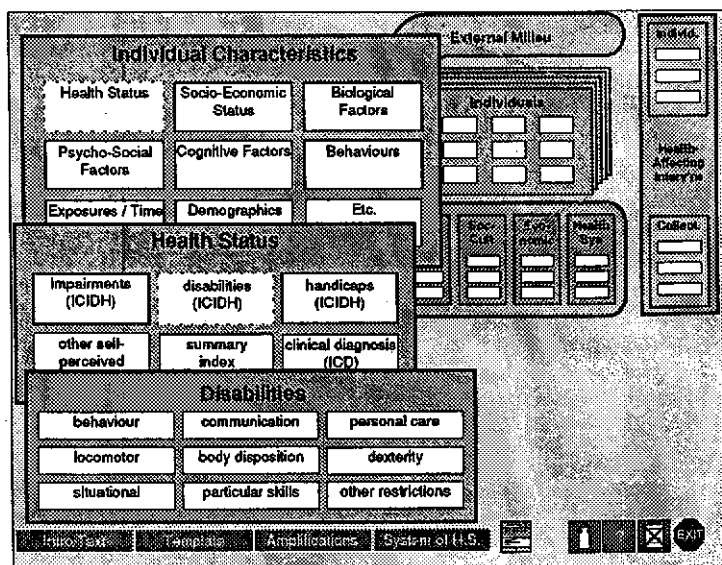
Figure 1



Surrounding, but not completely enveloping the population (both visually and metaphorically), is the “External Milieu”. It in turn has been structured into a series of environments – the most popularly understood environment of physico-chemical exposures, the socio-cultural environment that includes TV and social support (and emphasized this morning by Syme in connection with the Alameda county study and the failure of the MRFIT trial), the economic environment of income and work, and inherited aspects of the health system viewed as part of the environment – like the bricks and CAT scanners of hospitals.

Finally, as the product of our conscious intended actions, the Template defines a third domain of “health-affecting interventions”. This phrase was deliberately chosen to be broader than “health care”. These interventions are of two general kinds. The first, “individual” health-affecting interventions, act on us as individuals one-on-one – for example encounters with providers of various health care services. The second is “collective” health-affecting interventions. These take the forms of government programs and regulations that act on us collectively, though indirectly (and often inadvertently), through the external milieu.

Template: Main (exploded)  
Figure 2



The Template software also allows the user to “drill down” into the various domains and sub-domains – so they can be “exploded” to show finer levels of underlying classification structure. Figure 2 shows this for the domain of Individual Characteristics using a conventional hierarchical classification structure – first expanding “Health Status” and then “Disabilities” building on the World Health Organization’s International Classification of Impairments, Disabilities and Handicaps (ICIDH).

### Inferential Apparatus

The images given from the Template so far are static. But clearly, health is a dynamic process. Expression of genetic predispositions, risk factor exposures, and diseases, for example, evolve through time and the underlying bio-medical phenomena often exhibit very long latencies. Salient health information should include descriptions of these dynamic processes.

Moreover, a major use of an information system is generating projections, and relatedly, posing and answering “what if” questions. This implies a tight coupling between the information system and some sort of “inferential apparatus” – by which I mean a systematic, rigorous and replicable method for drawing out the implications of data under hypothesized scenarios.

Examples of inferential apparatus include demographic projection and macro-econometric models in the social sciences; and use of CHD risk functions estimated from the Framingham data to predict the impact of the interventions in the MRFIT trials in health science. My point is not to comment on the quality of the resulting inferences, or the underlying methodology. Rather, it is to show how important it is for an information system to have an asso-

ciated inferential apparatus (perhaps several). The Health Information Template includes a sketch of such apparatus, and I'll show you a graphical image in a moment.

But first, how might we conceive an inferential apparatus for the health of human populations? One approach is to borrow the core concepts used in descriptions of dynamic systems, whether in astrophysics or evolutionary economics. These are the ideas of "state variables" which evolve over time according to well-defined "laws of motion". In the case of health information, the state variables can be any number of attributes of individuals in a population, such as their age, cholesterol levels, occupational milieu, and coronary heart disease status.

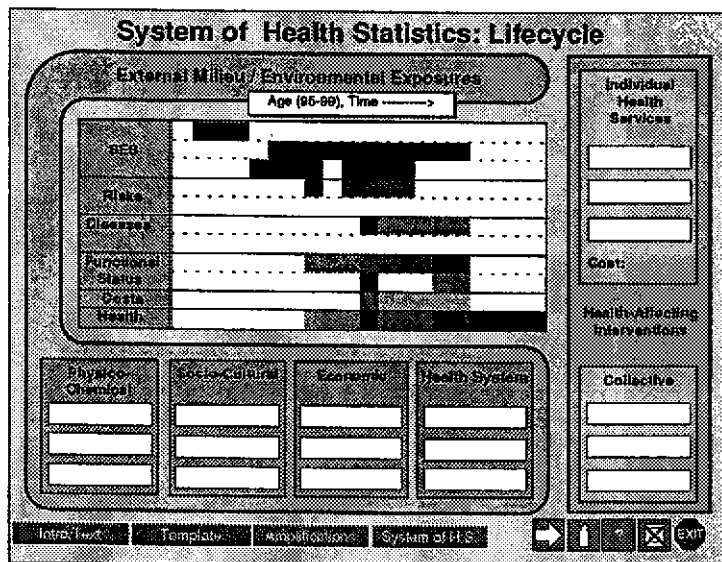
The laws of motion would then range from the trivial — age next year is one plus age this year, to much more complex — the chance that an individual will suffer a coronary event conditional on his or her cholesterol and smoking history, i.e. a CHD risk function for which many estimates have been published in the epidemiologic literature.

Of course, most of the dynamics of human health is not well-understood, so it is a huge leap to assume the existence of well-defined laws of motion for the dynamics of various attributes of human health. But that is not the point. The intent, rather, is to have a framework that can absorb and utilize such information as it is generated — and equally importantly, to have a framework that can guide the development of new health-related data and information. A key implication is that the core of our health information system should be individual population members considered in a life cycle context.

### Template: SHS Lifecycle

It is not easy to represent the notions of state space and laws of motion in a single graphical image. The Template illustrates these notions by adjusting the central portion of the main image as in Figure 3. It now shows a matrix of cubbyholes for information on individuals structured as in a biography, as well as some coloured shading to illustrate a hypothetical individual's biography of events and states. These cubbyholes in an individual's biography are similar to the record layout of a longitudinal microdata set. Each row corresponds to a group of variables, and each column refers to a time interval with age along the horizontal axis. The main groups of variables are:

Figure 3



- socio-economic status (SES) including education and income;
- various risk factors like smoking, genetic predispositions, and availability of social support;
- clinically defined diseases like CHD, lung cancer, arthritis, and Alzheimer's;
- vernacular health problems like being able to get around without a wheel chair, and chronic pain;
- direct costs of health services used, as well as economic costs like foregone earnings; and
- finally the bottom row, as well as the "bottom line" – a summary health status value (i.e. a number between zero and one) summarizing each individual's overall health status for the year.

The "external milieu" as a source of environmental exposures, and "health-affecting interventions" are the same as described earlier, and remain explicit in the framework.

The dynamics or laws of motion for these state variables in an individual's biography are not explicit in this image. All that is shown is the top levels of a classification structure. However, descriptions of the dynamics of various health-related processes are an essential part of the health information that should be encompassed by this framework.

The prototype approach we are taking at Statistics Canada is to collect and store diverse nuggets of information on health dynamics in a computer simulation model. This population health model, POHEM, is the inferential apparatus mentioned earlier.

### Concluding Comments

In closing, I have had time only to sketch briefly some of the major problems with health information currently, and to show you a few highlights of the Health Information Template. The Template provides a very ambitious image of a health information system. It has been developed through a process of consultation, and is designed to meet the concerns outlined earlier about our existing health data. It is possible to imagine a coherent and balanced health information system that can encompass a broad view of the determinants of health.

Moreover, I am hopeful that this image can be realized. Our POHEM model is moving beyond prototype to support applied analysis; we have a new Canadian Institute for Health Information being born; and Statistics Canada (notwithstanding the difficult fiscal climate) has new resources to support innovations in health information, including a longitudinal national population health survey. I am optimistic that significant advances are at hand. At the same time, today's discussions – especially regarding the importance of such diverse factors as neighbourhood and peer group influences, and psycho-neuro-immunology and other emerging bio-chemical pathways – which help deepen our understanding of the health of our population, suggest exciting, new, and as yet unfunded health information challenges. Thank you.

### References

Evans, R. and G. Stoddart (1990), "Producing Health, Consuming Health Care", Social Science and Medicine, Vol. 31 No. 12 (1347-1363).

National Center for Health Statistics (various years), Health United States, DHHS Pub. No. (PHS) 86-1232, Public Health Service, Washington, U.S. Government Printing Office.

Pommier, P. (1981), "Social Expenditure, Socialization Expenditure? The French Experience with Satellite Accounts", Review of Income and Wealth, December.

Power, Chris (199X), Class and Health,...

Statistics Canada (various years), Health Indicators, Ottawa

Syme, L. (1993), "The Social Environment and Health", Honda Foundation Symposium (elsewhere in this volume ...)

Wilk, M. (1991) Report of the National Task Force on Health Information, Statistics Canada, R.H.Coats Building, 24A, Ottawa, K1A 0T6.

Wolfson, M.C. (1992), "A Template for Health Information", World Health Organization Statistical Quarterly, March, WHO, Geneva (includes software diskette).

Wolfson, M.C. G. Rowe, J. Gentleman, M. Tomiak (1993), "Career Earnings and Death – A Longitudinal Analysis of Older Canadian Men", Journal of Gerontology, Vol. 48 No. 4, S167-S179.

# How the Developed Countries Became Rich

---

*Nathan Rosenberg*

In addressing a topic as ambitious as the one assigned to me, I have no alternative except to confine myself to focusing upon essentials. In doing so I am going to have to sidestep some important aspects of the story. For example, if I were proceeding at a more leisurely pace, I would want to assert that the first proper response to the question of "how the developed countries became rich" is that they became rich in different ways. In particular, the countries that developed later did so in some distinctly different ways than the countries that were the first to develop. Follower countries confronted a very different set of opportunities from those countries that developed earlier. The earlier developers were, in a very serious sense, exploring terrain that was distinctly terra incognita. Follower countries, by contrast, were moving along a path at least some features of which had already been marked.<sup>1</sup>

Thus, when the United States began to industrialize early in the 19th century, it did not need to invent railroads, power looms, or elaborate metallurgical processing technologies. These already existed in western Europe. I certainly do not want to suggest that follower countries necessarily had an easy time of it. In fact, in some respects follower countries encountered difficulties of a kind that did not confront the early developers. Ldcs today need to deal with the fact that economic coexistence with advanced industrial countries entails the continual threat that the sophisticated technologies in the hands of rich countries will generate innovations with very deleterious consequences to countries that are just aspiring to industrialize.

The twentieth century provides numerous examples of the substitution of new products for old ones upon which some less developed countries have been heavily dependent - synthetic fibers for cotton and wool, plastics for leather and for some non-ferrous metals, fixing nitrogen from the atmosphere rather than relying on deposits of guano, synthetic for natural rubber, synthetic detergents for vegetable oils in the manufacture of soap, optical fibers for copper wires, etc. An economy without some command over advanced technologies may be highly vulnerable to sudden changes in demand generated by more sophisticated technologies abroad. If the domestic capability to adjust is limited, the economic consequences may well be devastating. Of course, it needs also to be remembered that it was technical change and economic growth in the industrial world that created the initial demand and profitable market opportunities for many resource-intensive products in the first place. The extraordinarily high income levels of some middle-eastern oil-exporting countries would have been inconceivable in the absence of the intense demand for petroleum that was created by western industrial technology in the first place.

## II

What factors, then, accounted for the accelerated rate of economic growth that the presently-developed countries achieved at earlier stages in their history? If, for the moment, one puts aside what I would regard as critical facilitating institutional changes, I believe it is

possible to identify certain strategic and apparently common components in the experiences of each of the successful industrializers.

A useful way to begin an examination of these components is to recall an illuminating exercise undertaken by Moses Abramovitz and other economists back in the mid-1950s.<sup>2</sup> Neoclassical economics had, for many years, suggested that the long-term rise in per capita incomes could be understood as a process of mobilizing more inputs into productive activity, primarily by a high rate of capital accumulation that raised the capital/labor ratio. But when Abramovitz compared the rough quadrupling of U.S. per capita incomes between 1869 and 1953 with the increase in conventionally-measured inputs of capital and labor, he found that the rise in such inputs was in fact quite small. The rise in inputs could account for only about 15% of the measured growth in output. Other economists, such as Solow, arrived at a similar result, even though they employed different methodologies and somewhat different measurements. It appeared that 85% or so of the observed growth in per capita incomes - labelled The Residual - could not be accounted for by the growth in conventionally-measured inputs.

Although one might ask why the main neoclassical tradition had remained immersed for so long in an apparently "dogmatic slumber", there was, in fact, an excellent reason that has not been fully appreciated. An economic approach to economic growth was not feasible until certain magnitudes could be reduced to some credible form of quantitative expression. The measurement of economic growth itself required a conceptual framework and a methodology for aggregating the myriad of activities that make up the daily life of a large economy. A way of measuring economic output and its changes over time was, in fact, a critically-important "invention" of the interwar years. It was this prior conceptual thinking, and the painstaking empirical research to which it gave rise, that eventually made the discovery of The Residual possible.

Once it became clear that only a small portion of the measured growth in output could be accounted for by measured growth in inputs, numerous attempts were made to sort out the respective contributions of technological change and other factors to long-term productivity growth. The most systematic of these attempts were the growth-accounting exercise of Edward Denison.<sup>3</sup> There is now broad agreement that long-term economic growth is much more a matter of using resources more productively than merely employing more resources. Beyond that, however, there is still no consensus on the quantitative importance of the separate variables.

The reason for this failure to achieve a greater consensus goes to the heart of our present concerns about the industrialization process. Successful industrialization, where it has occurred, has involved the creation of what might be called a growth system, i.e., a highly interactive set of forces and institutions that mutually influence one another. Within such a system, it is not possible to partition off the contribution of each of a set of variables, separate from the others, and then to add up each of these separate contributions. Rather, within this system, changes in any one variable need to be recognized as significantly influencing the "payoff" from other variables.

Thus, central to this growth system have been the interactions between technological change and capital formation. The large size of the residual pointed forcefully toward technological change - an unmeasured input - as the critical variable. Yet technological change, by itself, can probably exercise only a very limited impact upon productivity improvement. This is so because most technological changes - not all - exercise their impact only when they have been embodied in the form of new and more productive capital equipment. Thus, capital formation is typically the necessary vehicle that has determined the speed with which, and the range over which, new technologies generate economic benefits. *Ceteris paribus*, societies that generate high rates of capital formation will derive the benefits of technological change more rapidly, and to a greater extent, than societies with lower rates.

But even this formulation is obviously too simple, because the incentive to invest is not independent of the availability of new technologies that hold out the prospect of raising the rate of return on capital formation. Indeed, in a world of unchanging technology, the returns to capital formation might be expected to fall quickly to very low levels, as Malthus and Ricardo once feared. In such a world, even high levels of investment will eventually make relatively small contributions to economic growth. A construction industry that relied exclusively upon shovels for earth-moving equipment would soon discover, when all workers were adequately equipped, that additional shovels contributed very little to further growth in output. On the other hand, in a regime of technological change, the availability of wheel-barrow would be expected both to raise labor productivity and the rate of return on capital. Similarly, the introduction of bulldozers would provide yet a new set of opportunities for growth in output and the prospect of a higher rate of return on investment.

Thus, the economic payoff to technological change and capital formation have been closely intertwined, with the behavior of each simultaneously determining the value of the other. Similar observations hold, within an industrial system, between the commitment of resources to education and training of the labor force, and the rate of technological change. An economy experiencing rapid technological change requires a more highly educated labor force because such a labor force can learn, can acquire new skills, and can more readily make the kinds of adaptations and adjustments that are likely to be required of it under conditions of rapid technological change. Conversely, in a world of absolutely unchanging technology, the purely economic payoff to further education is likely to be low.

But here again, causality has to be seen as running in both directions. One cannot appropriately think only in terms of what kind of education is required by a technological system of some given level and rate of change. For it is grossly erroneous to think of technology as some exogenous variable to which the labor force merely adjusts. On the contrary, an educated labor force will have a greater capacity for generating technological change in the first place. Indeed, one powerful insight that has emerged in the wake of the discovery of The Residual is that economists have, in the past, defined capital formation far too narrowly. There can be no doubt that certain kinds of expenditures, not previously regarded as capital formation, have been crucial to the growth process in industrializing economies. If capital formation is conceived of in a generic sense to include any kind of present resource use that demonstrably raised future productivity, it would incorporate expenditures upon R&D as well as some fraction of expenditures upon education, nutrition and health.

Thus, although a rise in the rate of capital formation has, historically, been associated with the industrialization process, and although the timing and the extent of the increase has shown considerable variation among countries, it is also true that the national income accounts of successful economies have systematically understated the proportion of their resources that have been devoted to investment. A redefinition is necessary, one that expands beyond the narrow notion of capital as consisting exclusively of tangible reproducible goods to include expenditures on R&D and some portion of expenditures on education, training and health. One estimate by Simon Kuznets attempted to provide some adjustment for the discrepancy between the generic and the conventional definitions of capital formation. Kuznets' estimate, admittedly crude, suggests that one expanded definition of the share of GNP devoted to gross capital formation would be twice as large as the conventional definition [See Table I].

Table I

	Private and Public Consumption	Gross Capital Formation
National accounts definition	77	23
Omitting intermediate products	70	30
Allowing for investment in man but excluding income foregone from capital formation and GNP	58	42
Adding income foregone in formal education and on-the-job training	53	47

Source: Moses Abramovitz, "Manpower, Capital and Technology," in Ivar Berg (ed.), *Human Resources and Economic Welfare, Essays in Honor of Eli Ginzberg*, Columbia University Press, 1972, pp. 68-69, footnote 1. Abramovitz' calculations draw upon Simon Kuznets, *Modern Economic Growth*, Yale University Press, 1966, Table 5.2.

Kuznets' estimate indirectly raises another significant issue concerning the interaction of the relevant variables. It is common to point to net capital formation as the appropriate magnitude in relating present resource use to future growth in productivity, since net capital formation indicates the share of an economy's output that is devoted to increasing the stock of capital. However, if technological change is proceeding at a rapid pace, the mere "replacement process" will involve the introduction of capital of superior productivity to the older vintage of capital that has been replaced. Under these circumstances, gross capital formation may well be a more accurate measure than net capital formation of the extent to which resources are being applied to productivity-increasing purposes. In industrial societies that already possess large stocks of capital, the proportion of GNP devoted to replacement is also large and therefore net capital formation figures will substantially understate the share of the economy's current output that is being devoted to uses that will raise future output. Thus, Kuznets has estimated that, for the period 1919-1948, the United States devoted a modest 6.8% of Net National Product to net capital formation, but at the same time devoted a very substantial 18.2% of Gross National Product to gross capital formation.<sup>4</sup>

Some final significant interactions within the growth system of industrialization involve the demographic variable. The growth in per capita output and the new uses to which industrial societies devote their resources - including improved nutrition and housing and better public sanitation facilities - have regularly brought with them reduced levels of mortality and greater life expectancy. This improvement in life expectancy has, by itself, had a significant impact upon the willingness to commit resources to education. Insofar as education is regarded as an investment that will yield a return in the form of a higher future income stream, an improvement in life expectancy will raise the prospective return on such investments by extending the period of time over which these higher incomes will be experienced. And, conversely, a country with a low life expectancy is also one in which, *ceteris paribus*, the returns to investment in education will be low. As Abramovitz has pointed out:

The high level of mortality rates a century and more ago meant not only that infant mortality was high, but also that the chance of survival through adolescence was small and life expectancy thereafter much shorter than it is today. Crippling morbidity was also more common. All these conditions effectively reduced the prospective rewards to investment in man. To spend eight or ten years in school beyond the age when earnings might otherwise begin, to forego these earnings, and to bear the other expenses of schooling was obviously less attractive when the remaining span of working life was, say, twenty-five years, on the average than it is today when it is forty or forty-five years. In the same way, the lengthening span of working life must have made people more ready to accept the risks and costs of seeking their fortunes in distant places and in new occupations.<sup>5</sup>

Thus, industrialization and rapid economic growth were not simply consequences of a slow but sure accumulation of capital and labor augmented by exogenous technological change. Rather, they were the product of a "growth system" that was far more highly interactive than is commonly recognized. Positive feedbacks ran from technological change to investment in human and physical capital and vice versa. The loosening of constraints in one area was likely subsequently to allow the loosening of constraints in other areas. Improvements in one place triggered greater benefits elsewhere. The cumulative effects of this interactive process, continued over a period of many decades, has been a truly astonishing rise in the productive capacities of the industrial societies that underwent it.

These improvements in productivity, moreover, have also been pervasive, and may even be applied to areas of human life and social organization where we normally think of such concepts as "productivity" or "efficiency" as foreign and inappropriate. Considered from a narrowly economic point of view, industrial societies operate at a much higher level of "demographic efficiency" than Idcs. Obviously, reducing the death rate among the young is a goal of overwhelming importance for its own sake. At the same time, it can be recognized that such reductions also have major efficiency consequences in terms of economic performance. As a result of the historic decline in mortality in industrial societies, fewer resources are absorbed in the reproductive process by infants and children who do not survive to adulthood as replacements for the stock of people who die each year. Poor countries, with high birth rates to offset high death rates, "waste" large amounts of resources on people who do not live long enough to enter the labor force and thus make a contribution to the output of the economy. Such poor countries must commit a significant fraction of the energies of women in their reproductive years to child bearing and child care - energies that culminate in the heartbreak as well as the "waste" that is the consequence of high levels of mortality.

The notion of improved "demographic efficiency" raises issues where, of course, an economist is well advised to tread lightly. But since it was economists who unleashed Malthusianism upon the world, it needs at least to be observed that there is no clear historical correlation, among the industrial countries, between their rates of population growth and their rates of growth of per capita income. What can be stated positively, however, is that over the past 200 years, all successfully industrializing countries experienced not only an acceleration in their rate of growth of per capita income, but in their rate of growth of population as well.<sup>6</sup> Obviously, the nature of the emerging growth system in industrial countries was such as to reconcile accelerated population growth with an acceleration in the growth of living standards. Put another way, it appears as if the same forces that were at work raising material productivity were also at work reducing premature mortality.

### III

Within the framework of interacting forces that have made up the industrial growth system, one, so far unmentioned, was pre-eminent. This was the application of scientific knowledge and scientific methodology to a progressively widening range of productivity activities. If one takes the truly long-run view measured in decades or even centuries, Kuznets was surely right in asserting that "The epochal innovation that distinguishes the modern economic epoch is the extended application of science to problems of economic production."

But in precisely what way has science played this role? Which aspects of the scientific enterprise have played the role, and under what circumstances? It is commonly argued or implied that the greatest economic benefits of science flow from research going on at the scientific frontier, and the usual formulations focus on causal forces that run exclusively in a "downstream" direction, from science to technology.

In fact, the economic benefits of frontier scientific research have always been highly uncertain. Some fundamental scientific breakthroughs have provided "only" intellectual or aesthetic satisfaction - as, e.g., in astronomy. Others have yielded benefits only many decades later. It is quite possible, although the proposition cannot be readily tested, that the lag between the growth of scientific knowledge at the frontier, and useful applications of that knowledge, has declined in recent decades, partly as a result of institutional innovations that facilitate those applications. The huge growth of industrial research laboratories in the twentieth century, that monitor frontier research within the university community and elsewhere, and that increasingly participate in frontier research themselves, has certainly tended to reduce the length of the lags. The Directory of American Research and Technology, 1992, listed 11,934 organizations, i.e., "all known non-government facilities currently active in any commercially-applicable basic and applied research, including development of products and processes."<sup>8</sup> [p. vii. For data on the formation of industrial research laboratories up till 1946, see Table II].

**Table II**

Establishment of Research Labs in American Industry

	<u>Manufacturing</u>	<u>Resource-based</u>	<u>Utilities, etc.</u>	<u>Total</u>
Before 1899	112	0	27	139
1899-1908	182	1	41	224
1909-1918	371	2	83	456
1919-1928	660	12	141	813
1929-1936	590	8	128	726
1937-1946	388	6	99	493
<b>Total</b>	<b>2,303</b>	<b>29</b>	<b>519</b>	<b>2,851</b>

Source: David Mowery, *The Emergence and Growth of Industrial Research in American Manufacturing, 1899-1945*, Stanford University doctoral dissertation, 1981, p. 51.

A major function of such industrial labs is to identify significant innovations elsewhere and to provide the scientific and technological expertise that makes rapid imitation possible. In this sense, it has been a major goal and achievement of industrial labs to bring about a more rapid diffusion of innovations that have already been made elsewhere.

In fact, from the point of view of deriving the economic benefits of science, the critical issues revolve not around creativity and originality in basic science, but upon the effectiveness of the institutional framework for the application of scientific knowledge, wherever and whenever that knowledge may have been generated. For there has been in place for some time a communications network that is capable of transmitting the findings of scientific research all over the world. Developing successful commercial applications of this knowledge is, however, far from simple and straightforward. Some of America's most distinctive accomplishments in the past have been in developing institutional couplings that have been highly successful in achieving those commercial applications. This has included not only the increasingly numerous industrial research laboratories but, in agriculture a land-grant college system along with agricultural experiment stations and county agents; in medicine the National Institutes of health; and at the level of more fundamental research, an extensive system of federal financial support for research carried out within a highly diversified university structure.

In addition to the institutional requisites, the successful exploitation of scientific knowledge has flourished best in industrial countries that have offered potential innovators ready access to capital and strong financial incentives, and have nourished and educated effective managerial and engineering cadres. Nineteenth century Czarist Russia produced many brilliant scientists and inventors, but their presence exercised a negligible impact in a society that lacked an appropriate managerial, engineering and financial infrastructure. On the other hand, America's emergence to a position of technological leadership before the First World War occurred in a period when its achievements in basic science were minimal and, with few exceptions, of no great international consequence. In this respect nineteenth century America bears some interesting resemblances to twentieth century Japan. Both countries managed to achieve rapid industrialization with no more than a modest scientific base because of great aptitude for borrowing and exploiting foreign technologies. On the other hand, the relative stagnation of the British economy in the twentieth century has occurred in spite of continued brilliant performance at the scientific frontier. Until quite recently the British scientific community continued to receive more Nobel Prizes per capita than the United States, but it has at the same time failed to maintain competitiveness even in many inventions that had originated in Britain - radar, the jet engine, penicillin, and the Pilkington plate glassmaking process. Moreover, the revolution in molecular biology that began with the discovery of the double-helix DNA molecule was, to a remarkable degree, a British (indeed, a Cambridge) achievement, and yet British firms currently play a role of negligible importance in the emerging biotechnology industry.

The conclusion seems inescapable that a first-rate basic science community has, in the past, been neither a necessary nor a sufficient condition for success in the commercial exploitation of scientific knowledge. Other elements of the industrial growth system can compensate or substitute for its absence domestically - at least so long as a country possesses the requisite social capabilities for borrowing from abroad.

What appears to matter more than the quality of a country's basic science, as judged by the usual academic or Nobel Prize Committee criteria, is the extent to which the activities of the scientific community can be made to be responsive to the needs of the larger society. Much of the rhetoric of academic science, with its overwhelming emphasis on the autonomy of science, seems to conspire against a proper understanding of this role. But the fact of the matter is that, in the course of the twentieth century, and with varying degrees of success, industrial societies have created increasingly dense networks of connection between the conduct of scientific research and the articulated needs of the larger social system. Indeed, the industrial research laboratory is essentially an institutional innovation in which the scientific research agenda is largely shaped by the needs of industrial technology. The role of industri-

al scientists is to improve the performance and reliability of technology. In doing this, they not only make use of the existing stock of scientific knowledge; they frequently add to that stock, occasionally making additions that are judged to be of Nobel Prize quality. This should not be surprising. Industrial activity provides unique observational platforms from which to observe classes of physical phenomena that are unlikely to be observed elsewhere: premature corrosion of an underwater cable, unidentifiable sources of interference in electromagnetic communications systems, or extreme heat generated on the surface of an aircraft as it attains supersonic speeds.<sup>9</sup>

The industrial research laboratory, in short, has rendered science more and more an endogenous activity, whose directions are increasingly shaped by economic forces. But the desire to produce socially useful knowledge has not been confined to industrial research. It also underlies government decisions to commit resources to various fields of science, including scientific research of a fundamental character. This is apparent in the sources of funding of American academic research in recent decades.

The National Science Foundation is indeed primarily committed to the support of scientific research with no concern over eventual utility. But the NSF has accounted for only about one sixth of federal support for university research in recent decades [See Table III]. The Department of Defense and two other government agencies that are allied with Defense in many ways - NASA and the Department of Energy - have accounted for much more, roughly a third. Since 1960 NIH funding of university research has increased drastically. NIH is presently by far the largest federal supporter of academic research, accounting for almost one half of the total.

**Table III**

Agency Funding of Academic Research							
Percent of Federal Research Funds Originating Within Particular Agencies							
Year	NIH	NSF	DOD	NASA	DOE	USDA	Other
1971	36.7	16.2	12.8	8.2	5.7	4.4	16.0
1976	46.4	17.1	9.4	4.7	5.7	4.7	12.0
1981	44.4	15.7	12.8	3.8	6.7	5.4	11.0
1986	46.4	15.1	16.7	3.9	5.3	4.2	8.4
1991 (est.)	47.2	16.1	11.6	5.8	4.7	4.0	10.7

Source: National Science Foundation, *Science and Engineering Indicators* 1991, p. 360.

The mission orientation of the biggest funders of academic research, and their particular fields of interest, are reflected in the distribution of research funding by field. In 1989, funded research in the engineering disciplines exceeded funded research in all the physical sciences (\$2,387,593,000 as compared to \$1,643,377,000.<sup>10</sup> The interests of the NIH and, to a lesser extent the Department of Agriculture, can be seen in the fact that more than one half of

academic research is now in the life sciences," and most of that is in the medical and agricultural science areas. While it is officially called "basic research" and indeed is basic in the sense that it constitutes a search for understanding at a fundamental level, the research is motivated by a concern with very practical problems, such as the helplessness of doctors and hospitals in dealing with various kinds of cancers, or AIDS, and is aimed at providing a better understanding and conceptual framework for arriving at solutions to very urgent social concerns.

#### IV

Thus, the burden of the discussion of the role of science in industrial societies is that science itself has become gradually incorporated, in the course of the twentieth century, into the growth system that has propelled industrial societies along the growth path that they have followed in the course of the 19th and 20th centuries. That growth system, in which technological change has always played a central role, is now reinforced by a powerful scientific research capability which has strong public and private components: (1) public subsidy of science, especially of a more basic nature, (2) higher educational institutions that perform much of society's basic research in addition to their older, established training functions, and (3) a research capability in private industry, where scientific resources are specifically directed toward strengthening and expanding a firm's technological capabilities. The growth system has worked well, however, only to the extent that it has generated strong investment incentives, including the incentive to invest in the search for new technologies.

The inseparability of technological change and investment cannot be overemphasized. New and improved technologies have become economically relevant only when it has appeared to be profitable to undertake the investments necessary for their exploitation. From the early years of the industrial revolution this has meant an enlarged commitment of resources to augmenting the stock of physical capital. In the course of the 20th century the growth system of advanced industrial societies has required a huge expansion of investment in the stock of human capital.

There is, finally, much reason to be dissatisfied with the economic performance of industrial societies these days. In the early years of the 1990s they have been experiencing extended periods of unemployment at levels unknown since the Great Depression of the 1930s. For most of these countries, rates of productivity growth have remained stubbornly below their long-term trend levels for two decades or so. Improving economic performance may require a degree of institutional innovation as well as political discipline that will be difficult to achieve. But a better understanding of the components of the growth system that brought us to our present juncture is an essential condition for future improvement.

## ENDNOTES AND REFERENCES

- 1 See the illuminating examination of these issues in Moses Abramovitz, "Catching up, forging ahead, and falling behind," Journal of Economic History, June 1986, pp. 385-406.
- 2 Moses Abramovitz, "Resource and output Trends in the United States since 1870," American Economic Review Papers and Proceedings, May 1956, and Robert Solow, "Technical Change and the Aggregate Production Function," Review of Economics and Statistics, August 1957.
- 3 Edward Denison, Accounting for United States Economic Growth, 1929-69, Brookings, Washington, DC, 1974; Accounting for Slower Economic Growth: The United States in the 1970s, Brookings, Washington, DC, 1970; and Trends in American Economic Growth, 1929-1982, Brookings, Washington, DC, 1985.
- 4 Simon Kuznets, "Long-term changes in national income of U.S.A. since 1870," Income and Wealth series, volume II, Bowes and Bowes, Cambridge, 1952.
- 5 Moses Abramovitz, "Manpower, Capital, and Technology," op. cit., pp. 54-55.
- 6 Simon Kuznets, Modern Economic Growth, chapter 2. Of course, the beginning of win-  
dom in so many matters of historical demography is to remember the exceptional status  
of France!
- 7 Simon Kuznets, Modern Economic Growth, op. cit., p. 9.
- 8 R.R. Bowker, New Providence, N.J., 1991, p. vii. For data on the formation of industrial  
research laboratories up till 1946, see Table II.
- 9 For further development of this point, see "How Exogenous is Science?", chapter 7 of  
Nathan Rosenberg, Inside the Black Box, Cambridge University Press, 1982.
- 10 See Nathan Rosenberg and Richard R. Nelson, "American Universities and Technical  
Advance in Industry," unpublished manuscript, Stanford University, August 1993, Table 5.
- 11 Ibid.

# Historical Perspective on How Prosperity and the Social and Physical Environments Have Affected Health and Well-Being

---

*J. Fraser Mustard and John Frank*

One of the striking observations in human history is the dramatic increase in population that has occurred in very recent times. McKeown, in his book "The Rise of Populations", shows the changes in mortality and increase in population in England and Wales in association with the Industrial Revolution. There is evidence from other Western countries of dramatic changes in the health of the populations during this period (Fogel, 1994). It is generally assumed that this change in the health status of the populations was primarily driven by improved public health measures, particularly in better sanitation and drinking water. These measures were crucial in reducing the risk of infection from waterborne diseases, but were not the primary factors influencing death from airborne infections. When these analyses were done, the broader factors influencing human health and well-being such as the quality of early life and mind body relationships in relation to host defence systems, were not well understood. There are today a number of hypotheses to explain the sudden improvement in health status in most Western nations since 1800 - hypotheses which concern the quality of social and physical environments, individual development, competence and coping skills of individuals, exposure to hazards, and how these interact to determine the health status and well-being of individuals and the population. This improvement in the health and well-being of a large number of humans is very recent in the history of *Homo sapiens*. Our knowledge of earlier periods is scanty, but it does illustrate some of the differences between today's societies and early communities.

## THE HUNTER-GATHERER PERIOD (OR: PRE-AGRICULTURAL SOCIETIES)

In trying to explore a historical perspective on the determinants of health, it is difficult to probe back into earlier periods, due to a lack of high-quality, morbidity, mortality and socio-economic data. However, it is worth reflecting on the retrospective observations that have been made and the general conclusions that come from them. The observations that have been made about primate evolution have a bearing on understanding our own species. It is now clear that the evolution of primates has been heavily driven by climatic changes and their sources of food. Recent work shows that our brain size, as well as the design of our gastrointestinal tract, are evolutionary related to the sources of food we are able to acquire. One of the important things about this evidence is that the more complicated the sourcing of food, the greater the brain capacity had to be to work out the patterns and techniques for sourcing that food supply. For example, in the tropical forest dwelling primates, fruit eating meant many specific seasonal fruit bearing trees had to be reliably relocated each fruiting season (Milton, 1993). In addition, for most primate species, including humans (particularly during our tribal or troop period) how well the group worked together, was very important in source-

ing and supplying the food. For most primates the long period of development of children requires a strongly supportive nurturing environment around offsprings when they are young to optimize their development.

The newer knowledge, about the relationship among development, coping skills, and host defense systems, has an important bearing on our understanding the determinants of health. The observation from non-human primates - that female rearing has a huge influence on competence, coping skills, health and well-being in adult life and how offsprings rear the next generation (Suomi, 1991) - probably has a direct bearing on ourselves. Boserup (1981), in her book, describes hunter and gatherer societies as being heavily constrained by their need to source their supplies of food, and the fact that they had to move around a great deal. This had considerable constraint on the number of young children that could be coped with under such conditions, and there probably was resultant birth spacing, as well as infanticide, driven by the kind of existence these people had. It is quite likely that in some periods these populations were decimated by lack of food sources. They were also exposed to physical hazards, predators, and presumably, in certain environments, to pathogens. What we do know is that those communities had to function as collective societies to survive, even though they probably did not have a long life expectancy, as far as one can estimate from archaeological findings.

The balance between fertility and mortality constrained population growth during this period. The basic social unit was the troop or tribe and it is unlikely most units embraced more than 100 individuals. During this period which is about 90% of our existence, the social environment was probably good for developing basic human skills for survival while the inability to produce the food needed by the community and the physical environment were the source of the major hazards to health and well-being.

In this period, our ancestors did not work long hours and saved nothing. There was no occupational specialization, no structure of leadership, no concept of property rights, state laws, taxation nor permanent settlements, a different socio-economic environment from the one we live in today.

## **AGRICULTURAL REVOLUTION**

Our knowledge about the technological, economic and social evolution of *Homo sapiens*, in relation to the Agricultural Revolution, is far from complete. When we moved to the agricultural evolutionary stage, about 8,000 years ago, we added a different dimension to the existence of the human species. Farming has to be one of the most powerful technological innovations in our history as a species. Societies could now provide sufficient food stuffs so that a major portion of the population no longer had to work in sourcing food, but could be fed by the work of others. It led to institutional innovations that have created governments, laws, religion and war. The beginning of our experiment with civilization profoundly altered the social environments in which individuals work, and created new conditions that influenced the health and well-being of individuals and populations.

There were several outcomes from this revolution. New social environments were created through the process of urbanization, mechanisms for the control of land for food production, arrangements to organize labour for the production and distribution of foods and other needs of the urban centres and the institutions for governance and control to ensure these more complex systems could work. This new social order was large and no longer had the collective, communal social-economic order of the old hunter-gatherer societies. There were serfs and slaves in the new order, military forces evolved, and governing structures and religious and other institutions emerged. As Boserup (1981) points out, many hunter and gatherer societies stayed away from the new order in order to protect their communal form of existence.

The new urban centres created the added complexity of a density of human beings living close together with effects on the health and well-being of individuals. Here, the quality of their water supply and sanitation systems become powerful factors influencing health, particularly for diseases that are waterborne, (cholera, typhoid) or spread by direct contact, airborne transmission (influenza, tuberculosis) and arthropod vectors (e.g. plague, typhus). The social organization of these communities was complex. A high proportion of the population often lived in some form of slavery or serfdom. These populations appear to have been vulnerable again to a variety of hazards, with the new addition of intensively transmitted infectious diseases. Such diseases not only endemically decimated urban populations, but also created extreme vulnerability to any new infectious agent that swept into the population, causing an extensive epidemic spread. In addition, a new factor was introduced in these societies in terms of the effects of war. Wars were largely fought over controlling the sources of food supply and their distribution. In this period, as today, war was often a key factor contributing to famine and epidemics. During this time, there was not, as far as one can tell from the data available in the archaeological and historical records, a sharp rise in the population of *Homo sapiens* on the planet.

Once you begin to develop a society which has one group of people producing food, and other people living in urban centres who are dependent upon that food, social institutions, economic systems and political systems are created to, not only get the food produced, but to transport and distribute it to the populations not producing the food (Sen. 1993). This creates a substantial problem because if the social and political institutional structures fail, the distribution function can fail, which can obviously lead to political instability, as well as adverse effects on the population through starvation.

In summary, the determinants of health and well-being during this phase of what could be considered *Homo sapiens* early experiments with civilization, are more complex than during the hunter and gatherer period. We now had populations living in close proximity to each other with all the implications this has for the evolution and spread of infectious diseases, (airborne and waterborne). Additional forces affecting health and well-being were wars, political control, and new forms of social order including slavery. The quality of the social environment in urban centres was probably not good for optimum human development. Many children probably grew up poorly nourished in poor nurturing environments. A poor early childhood can lead to adverse health in adult life (Fogel, 1994). Thus, although we had created the capacity to feed ourselves, both the physical and social environments during this period probably tended, from time to time, to be negative in terms of their effects on the health and well-being of large parts of the population. Fogel (1994), has concluded that chronic malnutrition probably contributed to the high mortality rates during this period.

## **INDUSTRIAL REVOLUTION**

Although there were many technological innovations that had effects on our socio-economic development following the Agricultural Revolution, a major change was when we learned how to harness fossil fuels as a source of energy. Prior to the harnessing of fossil fuels, many humans had to be a source of energy to do the physical tasks in the societies in which they lived. As already noted, most of these people, who were necessary as a major source of physical energy in their society, lived in some form of slavery or serfdom. Unfortunately, we do not have data to give us some idea of the differences in the health and well-being of the people who were in serfdom and slavery, versus the more privileged groups in society before the Industrial Revolution. As Rosenberg and Birdzell (1986) have emphasized in "How the West Grew Rich", the Industrial Revolution was associated with vastly enhanced prosperity for Western societies, leading in turn to better health and to the disap-

pearance of slavery and serfdom, and the development of democracy and universal suffrage. Thus, in Western countries we switch, in this very recent period of our history, from 80-90% of the population living in what could be considered poverty, to 80-90% living in what could be considered relatively prosperous circumstances. This was also associated with substantial improvement in the health of populations.

Associated with this sequence of events is a stunning rise in population (generated largely, at least initially in Western Europe, through fertility increase rather than the mortality decline). Boserup (1981) makes the point that it was in the interest of the new industrial societies to be able to sustain larger populations, particularly, to meet labour force needs.

**Massive Mortality Decline:** One of the striking features of the health status of today's developed countries is their low infant mortality rates and the much longer life expectancy of the population. There have been many attempts to understand the factors behind the change in mortality, but until recently the hypotheses explored have been incomplete. McKeown looked at the changes associated with, and subsequent to, the Industrial Revolution in Great Britain, but was unable to frame the changes he was studying in relation to the earlier period.

From an economic point of view, one of the principle changes that was occurring in European society during the last thousand years, was related, in part, to the form of governance. Societies that did not have autocratic rule were the societies that seemed to become more prosperous, in which larger cities and trade developed, whereas societies with autocratic rule had problems. (De Long and Shleifer, 1993). Britain, we know, was a society that got rid of princely rule around 1600. One then sees a staggering growth in the population of the city of London and of its economy in association with that. Of course, in Rosenberg and Birdzell's (1986) argument, it was the breaking down of old institutional structures which freed up innovative capacities for Britain to build its modern economy. These capacities included the capturing of the technology of fossil fuels as an energy source, which gave rise to what we classically call the Industrial Revolution, new institutions, the creation of the British Empire, and the enhanced prosperity of the United Kingdom. It is in this period that we begin, through the stimulation of Adam Smith and others, to try to understand how nations create and distribute wealth.

McKeown's (1976) observations deal with the major downturn in mortality in the U.K. after 1840. From his analysis of the Registrar-Generals' death records, certain things stand out. The bulk of the change in mortality could not be explained through medical intervention because there were no effective interventions during most of this period. Furthermore, there was a large change in mortality from airborne diseases which could not be easily explained simply by improved water and sanitation systems towards the end of the 1800's. McKeown's conclusion, that the bulk of this change was due to improved prosperity and nutrition, is provocative, but it does not take into account other social changes that were taking place that could also have influenced health and well-being.

The factors considered to have influenced mortality declines during these periods include: 1) improved nutrition; 2) improved water systems and sanitation; 3) improved female control over births and child rearing; and 4) improved control of individuals over their lives associated with the new kind of social economic orders that were produced.

McKeown's argument about improved nutrition was obviously part of the story and it fits into Boserup's observations, but does not account for why the U.K. infant and under-five mortality did not begin to fall until 1900, in contrast to the changes in tuberculosis deaths. Reves (1985) provides an important insight into why infant mortality changed much later, after the Industrial Revolution. He showed that the increase in child spacing, that occurred around the turn of the century, was sufficient to change the risk of exposure to infectious diseases in young children. The increasing median age of infection was, in all likelihood, a consequence of increased child spacing, but the fundamental question is, why did the increased

child spacing occur? We know from Boserup's work that in earlier societies, child spacing was, in part, driven by culture, access to resources, and by demands placed on the societies to do a variety of things. Birth spacing was already controlled in earlier societies by female-based rituals, taboos, decisions and controls. Forces operating in the U.K. around 1900 were probably different, but some set of forces must have operated through women, for it is largely women who determine the frequency of child bearing, through breastfeeding practices, coital frequency and nature (e.g. urging men to practice coitus interruptus) and abortion. In understanding the dynamics between the economy and such practices, one suspects it is probably not a simple relationship. For example, in periods of economic downturn, there may be economic pressures to increase birth spacing through which, as long as the economic decline is not so severe, there may be a positive effect on health, because of associated changes in infant mortality and life expectancy.

As McKeown concluded, there is no doubt that, for the waterborne diseases, improved water systems and sanitation would have eventually had a major effect on the changing patterns of these diseases as a cause of mortality. In the case of waterborne diseases, the major decline occurred in the period after 1900, which coincides with the improved water and sanitation systems that came into British communities in association with the hygiene movement and its public health accomplishments. This movement succeeded in reversing the deterioration of the British urban physical environments that had followed the early period after the Industrial Revolution. Of course, there is a powerful lesson here concerning the dynamic interaction between social and economic forces that brought about these changes. It was the socially conscious citizens of an increasingly prosperous society that drove the broader economic structures, and political instruments, to produce hygienic changes to improve the quality of life and society. This is an example of the fact that economics should not be considered in isolation from the basic social environment in which people live and work and the forces that interact within them (Szreter 1988).

Fogel (1994) has concluded that it was escape from chronic malnutrition that had a major effect on health, and life expectancy during this period. He concludes, that at the beginning of this period, the food supplies were not adequate to sustain the lower classes of the population as active members of the labour forces. He shows from the historical record the relationship between stature or development in early life and chronic diseases at middle and late stages. This direct evidence supports McKeown's conclusion that much of the improved health association with the Industrial Revolution was a consequence of improved nutrition.

The one thing that is missing, from all the published analysis of this period, is the relationship between mind and body and disease expression. The weight of the new evidence is that emotional well-being powerfully influences almost all disease expression, including such things as accidents and suicides. It is interesting, in terms of Reves' work, that if you think about the post 1900 improvement in U.K. infant mortality, the now well-know link between breastfeeding and immune system responses, including the mother's early response to antigens in the infant's saliva, could be an important part of this loop. A mother, who is in a deprived circumstance and whose mind-body dynamics is suppressing her immune capability, probably will not be adequately able to support the defenses of the child she is nourishing.

One of the key observations in McKeown's work is the steady decline throughout the nineteenth and twentieth centuries in mortality from tuberculosis, which is primarily a youth young-adult disease problem. Its decline is not easily explained by any of the hypotheses we have just discussed, except for McKeown's general argument of improved nutrition resulting from the prosperity of the Industrial Revolution. However, the improved nutrition argument, given the temporal features of this disease, seems to be an incomplete explanation for the change.

What we do know is that tuberculosis is influenced, not only by the factors involved in its spread, family age - structure, crowding, etc., but also by host response capability. It is well known, as William Boyd taught to young medical students at the University of Toronto a half century ago, that lots of people have the TB bacillus dormant in their body, but only certain factors lead to its expression, as symptomatic disease, in a minority of those infected - factors which may well be related to the conditions of the social environment. "TB" is an opportunistic infection. Thus, one could hypothesize that the increased prosperity and improved control people had over their lives after the Industrial Revolution, through associated effects on the dynamics of the neurological and immune systems, along with improved nutrition, increased their host defense capabilities, despite continuing infection of all birth cohorts with the TB bacillus early in life. Now, with our improved understanding of the dynamic relationship between the nervous system and the immune system, it is clear that such mechanisms will also have to be taken into account in explaining historical changes in patterns of health, associated with changing social economic conditions. Thus, where people are in the hierarchy of a society, and the degree of control they enjoy over what they have to do and the adequacy of their nourishment, could be critical in determining vulnerability to a wide range of diseases. In addition, such mechanisms operating within the mother and child relationship, may exert effects with delayed or latent expression, as disease and premature death fifty or more years later (Barker et al, 1992; Elford, Whincup and Shaper, 1991).

## HEALTH STATUS TODAY

Where relevant data is available, one is struck by the strong and pervasive relationship between where you are in the social structure of society and your health status, as measured by such things as sickness absence from work and mortality. Although life expectancy has improved for all social classes in the United Kingdom for the last 60 years, the relative gradient in mortality across social classes has been widening throughout the period (Davey Smith et al, 1990). And it is important to remember that when Britain introduced public financing of the health care system, Beveridge argued that removing the financial barriers to access to health care would decrease the gradient. When the Morrison Royal Commission did its report in the 1970s, it was stunned to find that the gradient had widened. This led to the Black Report on inequalities of health bringing out how the social environment is a major determinant of the health and well-being of individuals (Black et al, 1982).

The fact that most developed societies show strong risk gradients for most major causes of death, when assessed against income, education, social class and other socio-economic measures, is strong evidence that some basic "host" factors affect disease expression, particularly for the chronic diseases of the Western world (Cassels, 1976). These gradients show that the factor or factors are not only associated with poverty, but have a major influence on individuals in the middle classes.

How nations create and distribute wealth, and how that influences the social structure of society, are obviously factors affecting the health and well-being of individuals. It is important to understand the dynamics between what you have heard at this conference about human development, the development of essential coping skills and competence and the linkage between the nervous system and the host defense system of the human body. Countries that have had major improvements in health status tend to be countries that are prosperous and have a high degree of social equity. This may be part of the explanation for the remarkable improvement in the health and well-being of the Japanese population since World War II (Marmot and Davey Smith, 1989). We need to better understand how economic growth and prosperity affect the social environments of societies and the health and well-being of individuals and populations.

## ECONOMICS, SOCIAL ENVIRONMENT AND HEALTH AND WELL-BEING

One of the least understood fields in economics is the determinants of economic growth. We also poorly understand the relationship between economic growth, or decline, and changes in our socio-economic environment. It would appear that societies that cannot sustain their prosperity have difficulty sustaining high-quality social environments. As Sen (1993) has pointed out, how societies create and distribute their wealth, determines the health and well-being of the population. In periods of profound change, individuals will be at risk, in terms of their health and well-being, unless their societies adequately understand the relationships between society and health, and can design systems to cope with the changes. Changes in two regions of the world, Eastern Europe and Japan, bring out this relationship. The decline in the economies of Eastern Europe has been associated with a decline in health status of the populations while the improved prosperity of Japan has been associated with a marked improvement in their health status. In a recent analysis for the World Bank of the decline in health status in Eastern Europe, Hertzman concluded that the strongest factor seems to be the deterioration of the quality of the social environment (Hertzman, 1993). In contrast, the extraordinary improvement in the health of the Japanese, as measured by life expectancy changes, is associated with enhanced prosperity, through with what appears to have been a remarkable ability to sustain the quality of their social environments and reasonable income equity throughout the society (Marmot and Davey Smith, 1989).

Wilkinson (1992) has found in his analysis of a number of Western countries that life expectancy is correlated with the degree of income equity in the society. Societies with good income equity tend to have good social environments.

A concept that emerges from some of the new understanding of economic growth is that the income produced, through tradable goods and derived from ideas, has a major effect on the prosperity of a society. This sector of the economy could be considered the primary wealth-creating base of a society. It is perhaps analogous to Adam Smith's concept of productive and non-productive labour. Smith captured this in his chapter entitled "Of the Accumulation of Capital, or of Productive and Un-Productive Labour" (Smith, 1991). In describing unproductive labour he states,

*"The labour of some of the most respectable orders in the society is, like that of menial servants, unproductive of any value, and does not fix or realise itself in any permanent subject, or vendible commodity, which endures after that labour is past, and for which an equal quantity of labour could afterwards be procured. The sovereign, for example, with all the officers both of justice and war who serve under him, the whole army and navy, are unproductive labourers. They are the servants of the public, and are maintained by a part of the annual produce of the industry of other people. Their service, how honourable, how useful, or how necessary soever, produces nothing for which an equal quantity of service can afterwards be procured. The protection, security and defence of the commonwealth, the effect of their labour this year will not purchase its protection, security, and defence for the year to come. In the same class must be ranked, some both of the gravest and most important, and some of the most frivolous professions: churchmen, lawyers, physicians, men of letters of all kinds; players, buffoons, musicians, opera-singers, opera-dancers etc."*

It is important that those who work in the parts of society that are not the engines of economic growth understand their relationship to the primary wealth creating sectors of society.

Our weak understanding of how economic forces determine the social environment is brought out when we attempt to address the question, "How does a society sustain and enhance high-quality social environments for people to live and work in"? One of the debates that is taking place at the present moment is that social equality is more important than income equality (Kaus, 1992). Our understanding of human development makes it very clear that our cognitive capacity and behaviour is influenced by the conditions in which we exist when we are young, and when our brain is most plastic. We know that a substantial part of the anti-social criminal behaviour that we see in teenagers and young adults is set in motion by the social environments in which these individuals existed as young children before they entered the school system. There is also emerging evidence that some of the health risks in adult life could be set in early childhood (Barker et al, 1992). We know that there are institutional instruments societies can create which inhibit or prevent the development of poor social environments. For example, suburbs and shopping malls create some advantages, but they can also dramatically change the social environment and the interaction of human beings across social classes and races. Jane Jacobs would argue that that is not a healthy development for cities (Jacobs, 1961). From the standpoint of our understanding of the biology of humans and the determinants of our health and well-being, we may have to give serious consideration to concepts set out by Jane Jacobs. In our attempts to build pluralistic tolerant democratic societies, we will have to find ways to understand the dynamics among social environments, human development and the health and well-being of individuals. The tensions that are being created by our present failures, to build and sustain high quality social environments, are exemplified by the continuous civil strife that exists in the inner core of a number of major cities of the world, where many people are killed each month, and infant mortality rates remain at third-world levels for large sectors of the population.

## SUMMARY

The understanding of the determinants of health that comes from a historical perspective reinforces the Greek Goddess Hygeia. Hygeia was not involved in the treatment of the sick but was the guardian of health; she symbolized the belief that people would remain in good health if they lived according to reason.

We now understand how our existence in the social environment can influence host defence systems, and disease expressions, through the biological pathways in the mind-body system. We know that our skills that influence how we cope with our social environment are largely set in early life, where nurturing plays a key role in the cognitive and behavioural development. In periods of major socio-economic change in modern societies, the group that are most at risk, for long-term effects, tend to be mothers and children. Societies that can adjust to economic changes and sustain their support for mothers and children will probably be the most stable, with the best overall health of their populations.

Our hunter and gatherer ancestors probably had, for the demands on their "coping skills", a supportive social environment. Children were probably reasonably well nurtured for the conditions of these societies. However, people in these communities had an uncertain source of food and were at risk to predators and endemic infectious diseases.

The advent of the Agricultural Revolution changed dramatically how we lived and worked. For many, who were caught as serfs and slaves, these new communities may not have been as attractive as the hunter and gatherer societies. At this stage, health was strongly influenced by the living conditions in cities and epidemic infectious diseases. Although there were now better food supplies, distribution systems could break down with famine affecting

whole populations from time to time. War first becomes a major disruptive force at this stage in our history. It is probable that these communities were variable in terms of the factors influencing human development and coping skills. Some cultures and religions will have tended to create good environments for mothers and children.

The advent of the Industrial Revolution, and the associated new wealth, dramatically changed our species continuing experiment with civilization. This change led to increased prosperity in Western countries, with the bulk of the population living in prosperous and democratic societies enjoying a great improvement in their health and well-being. The improvements in health were related to several factors. Among them, better supplies of more diversified nutritious foods, better water and sanitation systems leading to a decline in water-borne infectious diseases, and a reduced childhood mortality to infectious diseases, probably largely as a result of increased birth spacing. It also seems likely that improved social environments led to a higher quality of human development, with positive effects on mind-body systems, and their influence on disease expression. The social gradients in health, in today's developed societies, appear to be strongly influenced by such biological pathways. We need to explore more deeply the relationship among economic growth, prosperity and high-quality social environments.

Rosenberg (1986), Romer (Ph.D Dissertation, 1983) and others have brought out that in today's world, science and ideas are the driving force in economic growth. A prosperous society needs a well-educated labour force that can cope with change. A major challenge today is how to sustain high-quality environments for individuals to live in while we cope with the techno-economic paradigm shift that is occurring.

## ENDNOTES

**Barker DJP et al.**, The Fetal and Infant Origins of Adult Disease (London: British Medical Journal, 1992).

**Black, D., Morris, J., Smith, C. and Townsend, P.** Inequalities in Health: The Black Report. (New York: Penguin Books, 1982).

**Boserup, E.**, Population and Technological Change. (The University of Chicago Press, 1981).

**Cassels, J.**, The Contribution of the Social Environment to Host Resistance (American Journal of Epidemiology 104:107-23, 1976).

**Davey Smith, G.**, The Black Report on Socio Economic Inequalities in Health 10 years on. (British Medical Journal, Vol. 301, pp 373-377, 1990)

**De Long, J.B. & Shleifer, A.** Princes and Merchants: European City Growth Before the Industrial Revolution. (Cambridge, MA: National Bureau of Economic Research, Working Paper No. 4274, February 1993).

**Elford, J., Whincup, P. & Shaper, A.G.** Early Life Experience and Adult Cardiovascular Disease: Longitudinal and Case-Control Studies. (International Journal of Epidemiology, V20, No. 4: pp 833-44, Dec. 1991).

**Fogel, R.W.**, Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Process in the Making of Economic Policy (Cambridge, MA: National Bureau of Economic Research, Working Paper No. 4638, April 1994).

**Great Britain Royal Commission on National Health Service.** Report of the Royal Commission on the National Health Service (The Merrison Report). (London: HMSO, 1979).

**Hertzman, C.**, The Determinants of Health in Central and Eastern Europe. Working Paper No. 28. (Toronto: CIAR, September 1993).

**Jacobs, J.**, The Death and Life of Great American Cities. (New York: Random House Inc., 1961).

**Kaus, M.**, The End of Equality. (New York: Harper Collins, 1992).

**Marmot, M.; Davey Smith, G.**, Why Are The Japanese Living Longer (British Medical Journal, 299(3):1547-1551, December 23, 1989).

**McKeown, T.**, The Modern Rise of Population. (New York: Academic Press, 1976).

**Milton, K.**, Diet and Primate Evolution. (Scientific American, pp 86-93 August 1993).

**Premier McKenna, Frank**, Lifelong Learning at Work – The New Brunswick Model. (62 Annual Couchiching Conference: August 5, 1993).

**Reves, R.**, Declining Fertility in England and Wales as a Major Cause of the Twentieth Century Decline in Mortality. (American Journal of Epidemiology, 122:112-26, 1985).

**Romer, P.**, Dynamic Competitive Equilibria with Externalities, Increasing Returns and Unbounded Growth (Ph.D. Dissertation, University of Chicago, 1983)

**Rosenberg, N. & Birdzell, L.E.**, How the West Grew Rich. (New York: Basic Books, Inc., 1986).

**Sen, A.**, The Economics of Life and Death (Scientific American, pp 40-47 May 1993).

**Smith, A.**, The Wealth of Nations. Book II, Chapter III. (New York: Everyman's Library, 1991).

**Suomi, S.**, Early Stress and Adult Emotional Reactivity in Rhesus Monkeys (The Childhood Environment and Adult Health, pp 171-188, 1991).

**Szreter, S.**, The Importance of Social Intervention in Britain's Mortality Decline c.1850-1914: a Re-interpretation of the Role of Public Health. (The Society for the Social History of Medicine, 1:1-17, 1988).

**Wilkinson, R.G.**, Income Distribution and Life Expectancy. (British Medical Journal 304:165-8, 1992).

# The Epidemiological Transition: From Material Scarcity To Social Disadvantage?

---

*Richard G. Wilkinson*

## Introduction

Earlier in this volume Rosenberg described the "growth system" which powered industrialisation and the historical rise in prosperity. At its heart were "positive feedbacks (which) ran from technical change to investment in human capital and vice versa". Mustard and Frank (also in this volume), recognising that "high quality social environments" are likely to be an increasingly important part of the growth system, drew attention to our lack of understanding of how economic growth affects the socio-economic environment. A key characteristic of the growth system was that its components formed a virtuous cycle, with positive feedbacks between them. Mustard and Frank's concern for the social environment raises the possibility that if its quality is allowed to deteriorate, then the virtuous cycle which produces economic growth could be weakened or broken.

A historical perspective on these issues is important not simply for making us aware of the continuities between the present and the past. An understanding of the past also increases our awareness of fractures and discontinuities. I want to deal here first, with an important discontinuity in the way that health fits into the growth system, and then move on to discuss evidence - again from health - of a shift in the nature of the most crucial determinants of the quality of the social environment. I will argue that mortality rates in the developed world are now no longer related to per capita economic growth, but are related to the scale of income inequality in each society. I believe that this change was marked by the turn round in the social distribution of heart disease when the so-called "diseases of affluence" became the diseases of the relatively poor in affluent societies.

If it is right to say that health is now influenced more by income distribution than by economic growth, then the same is likely to be true of other important aspects of human welfare. Indeed, the relationship seems to go to the heart of the human condition: if the historically predominant position of material factors as determinants of health has given way to social factors, then the possibility arises that the determinants of the real subjective quality of life may have undergone a similar transformation. But as well as adding to our understanding of how to produce the kind of creative and innovative society on which economic development increasingly depends, the data also raised questions about the benefits of further *undifferentiated* economic growth. If increases in the quality of life now depend primarily on improving the social fabric of society then we must ask whether further undifferentiated economic growth is worth the environmental risks.

## Life expectancy and GNPpc.

Before acknowledging the role of rising living standards in the historical decline of mortality in the developed world, it would be as well to recognise the role of the only important contribution that can be said to be entirely unrelated to economic growth. That is the almost

universally overlooked role of the natural selection of a population with greater genetic immunity to the major infections. Because the infectious scourges of the past typically took their highest toll in childhood, they acted as a powerful selective mechanism, killing the most vulnerable before they had passed on their genes to the next generation. If the population's gene pool contained significant genetic variations in resistance, then the very high rates of childhood mortality would have rapidly increased the proportion of subsequent generations which carried that immunity. Macfarlane Burnet gives important evidence (including some from twin studies) that human populations do contain substantial genetic differences in immunity to some infections, and points out examples (such as Mauritius) of non-European societies without economic development in which infections like TB have run their course from high to low mortality in three or four generations or "something over one hundred years" from first contact.<sup>1</sup>

If, in Victorian England, Archbishop Tait could lose five children (who would not have been short of the material necessities of life) from scarlet fever in a month, it is difficult not to give some weight to natural selection.<sup>2</sup>

Following that cautionary note we can now turn to the widespread association between economic development and mortality decline. While it is possible to argue about the sizes of the historical contributions which better nutrition, sewers, clean water supplies, improved housing and, eventually, immunization made to the long decline in mortality rates in the developed world, there can be no doubt that underlying the appearance of these and many other contributors was the enabling and sustaining power of economic growth.

Evidence of the broad relationship between economic growth and increasing life expectancy can be found not only in the history of the developed countries, but also in the third world today. Figure 1 shows the cross-sectional relationship between Gross National Product per capita (GNPpc) and life expectancy at birth in rich and poor countries in 1990.

Although it is clear that life expectancy rises steeply with increased GNPpc among the poorer countries, the data in figure 1 also suggests that the relationship between GNPpc and mortality peters out in the developed world. It looks as if there is some minimum threshold level of income (around \$ 5,000 per capita in 1990) above which the absolute standard of living ceases to have much impact on health.

Putting income on a log scale (as in figure 2) makes it look as if, rather than a well defined threshold level of income, perhaps it would be more accurate to think in terms of radically diminishing health returns to increasing income. However, as the curve (drawn free-hand) on figure 2 indicates, increases in life expectancy cease to be even a log linear function of income among the richest countries.

Turning from cross-sectional evidence to look at changes over a period of time, figure 3 shows that there is very little relationship in the developed countries between changes in GNPpc and life expectancy during the 20 years 1970-90. This time GNPpc is converted at purchasing power parities to provide a more accurate comparisons of the real material standard of living.

Whether one looks at OECD countries at a point in time or at changes over several decades, there is no strong relationship between income and health among the developed countries.<sup>3</sup> People in one country can be twice as well-off on average as those in another without benefit to their mortality rates. One country's economy can grow twice as fast as another for 20 years (say at 6 percent instead of 3 percent a year) and, as figure 3 shows, it may make no difference to the improvement in life expectancy.

## Approaching limits of human life expectancy?

The most obvious explanation for the virtual disappearance of a relationship between GNPpc and life expectancy might be that it reflects the approaching limits of human life expectancy. But this is implausible because there is no tendency for the rate of improvement in life expectancy to slow down where longevity is already high. Life expectancy continues to increase at 2-2.5 years per decade among the rich countries which is much the same as the average performance during peace-time this century. In addition, it appears that in several developed countries (where health "inequalities" are widening), the fastest improvements are still coming from those sections of the population where life expectancy is already highest. Lastly, mortality rates have recently begun to fall significantly among older people.

For these reasons, we must assume that the levelling out of the curve of rising life expectancy with GNPpc - shown in figure 1 - indicates the declining health benefits of further increases in the standard of living in the developed world.

## The epidemiological transition.

This point in economic development when societies reach the so-called "epidemiological transition" marks a fundamental change not only in the nature of the predominant health problems of human societies but also in the main determinants of health. It signifies the point at which the vast majority of the population gained reliable access to the basic material necessities of life. It is the point at which the so-called diseases of affluence became the diseases of the poor in affluent societies. Coronary heart disease, stroke and obesity, have all changed their social class distribution to become more common among the least well off.<sup>4,5,6</sup> That this transition was associated with attainment of a level of material sufficiency can be seen from the associated reversal in the social gradient in smoking and in the consumption of refined foods, such as white bread and sugar - which started to be eaten more by the less well-off for the first time in history - while the better-off ate more fresh fruit and vegetables. In all but a few countries the evidence on heart disease and stroke is fragmentary, but it seems likely that different developed countries have gone through the same transition and experienced the same turn round in the class distribution of these diseases at different dates.

The change in the social class distribution of obesity is however more widely recognised. Throughout most of human history obesity has been restricted to the privileged and in many cultures it was a status symbol. The rich were fat and the poor were thin. This remained true well into the present century. Although the change was slow, it looks as if it may have been accelerated in Britain by nutritional changes during the Second World War.<sup>7</sup> Certainly the Chief Medical Officer of the Ministry of Education commented in his annual reports of 1956-7 and 1960-1 on the increasing incidence of obesity in British school children.<sup>8</sup> In Germany the changeover probably came during the 1960s.<sup>9</sup>

As if to confirm Bordieu's thesis of the importance of the desire to express social distinction in aesthetic judgements, as soon as the poor became able to be fat, obesity ceased to be a status symbol. The ideal shape - especially for women - became progressively thinner and the slimming industry was born<sup>10</sup>.

Simultaneous with these changes was the decline in infectious diseases and their replacement as the main causes of death by cancers and degenerative diseases - the "epidemiological transition". Infections have always been the diseases of poverty against which the diseases of affluence were contrasted, and they remain the main causes of death in poorer countries today. Their passing is itself a strong indication of the weakening constraints of the absolute standard of living on health.

The last indication of this process comes from trends in the proportion of all babies with birthweights below 2,500 grams. Although perinatal and infant mortality rates have contin-

ued to decline rapidly, since the 1950s in England and Wales there has been no further marked decline in the proportion of low birthweight babies: between six and seven percent of babies have weighed less than 2,500 grams ever since.<sup>11,12</sup>

There is then evidence that in the later stages of industrial development countries go through a health climacteric after which the health of the vast majority of the population is no longer substantially limited by the absolute material standard of living. That such a phenomenon exists is indicated both by the evidence of the changing relationship between national mortality rates and GNP per capita, and by three changes in a society's health profile. The first is the decline of infections and their replacement by cancers, degenerative and other diseases as the main causes of death. The second is that the diseases of affluence become the diseases of the poor in affluent societies as heart disease and obesity become more common in lower social classes. The third is that the proportion of low birthweight babies approaches a lower limit.

### **Income distribution**

After the epidemiological transition, or health climacteric, health remains closely associated with deprivation, but the relationship is now with relative rather than absolute deprivation. Health within developed countries continues to show a clear gradient with measures of socioeconomic status.<sup>13</sup> Figure 4, for example, shows how death rates vary according to income group in the United States. As well as a cross-sectional relationship, evidence from Britain covering most of the present century suggests that the size of social class differences in death rates has got larger or smaller as the proportion of the population living in relative poverty has increased or decreased.<sup>14</sup>

There is however a marked contrast between the well-ordered relationship between socioeconomic status and health within a country (as shown in figure 4) and the lack of any clear relationship shown in the international data in figures 1 and 3. At first sight a close association between income and mortality within countries, and yet so little sign of one between them, looks like a paradox. It is however resolved by the distinction between relative and absolute income.

The importance of relative income is demonstrated by the surprisingly strong relationship between income distribution and national mortality rates which I and Irene Wennemo have found in developed countries<sup>15,16</sup>. Whether you look at the cross-sectional relationship or at changes over time, there is clearly a strong correlation between income distribution and national mortality rates.

Instead of being the richest, the countries with the longest life expectancy are the ones with the smallest spread of incomes and the smallest proportion of the population in relative poverty. Using a standard measure of income distribution, figure 5 shows that in 1970 (the earliest comparable data available) the countries with the smallest income differences were also the countries where average life expectancy was longest. In the years since then the rate of growth of average life expectancy has been closely related to changes in income distribution. Figure 6 shows that in the 12 member states of the European Community average life expectancy has grown fastest in those countries where relative poverty decreased fastest (or increased slowest) between 1975 and 1985. Data from a different source - including a number of countries outside the EC - is illustrated in figure 7 which again shows that the fastest increases in life expectancy have occurred in countries where income differences have narrowed. Lastly, figure 8 shows Wennemo's cross-sectional relationship between the prevalence of relative poverty in different countries and infant mortality rates. (Germany has been excluded because its data set held by the Luxembourg Income Study became corrupted.<sup>17</sup>)

Despite the small number of countries for which there is comparable data, all the relationships with measures of income distribution shown in figures 5-8 are statistically significant. The relationship implies that between a half and three-quarters of the differences in average life expectancy from one developed country to another may be attributable to differences in income distribution alone. Controlling for GNPpc, public expenditure on medical care, or total public expenditure, does not suggest that the relationship with income distribution is a proxy for an effect of better public services.

It might be suggested that because this relationship is established on national data it may prove to be an ecological fallacy. However, there is substantial evidence from other sources not only that individual health and income are related (as shown in figure 4), but also that health is responsive to changes in income.<sup>18,19</sup> The function of the international evidence shown here is primarily to distinguish between relative and absolute income levels. As relative income is inherently a social concept it cannot be dealt with at an individual level: it is societies - not individuals - which have income distributions.

A telling example of the influence of income distribution (or relative income) on health is the contrast between health trends in Britain and Japan. In 1970 life expectancy and income distribution were quite similar in both countries as figure 3 shows. Since then Japan's income distribution has narrowed and is now the narrowest of any country reporting to the World Bank. At the same time its life expectancy has increased by 6.9 years in the period 1970-90 and is now the highest in the world. In contrast Britain's income distribution has widened and Britain's position in the life expectancy league table of OECD countries has fallen from 10th in 1970 to 17th in 1990 with a gain of only 3.9 years in life expectancy. Although Britain has enjoyed some of the general increase in life expectancy, the widening of income differences up to 1986 appears to have slowed down the increase enough to deprive the average citizen of an additional year's life expectancy which the data suggests would have been gained had income distribution not widened.<sup>20</sup>

The powerful influence of the contrasting trends in income distribution in the two countries is consistent with what has been happening to their social class mortality differences. As Japanese income differences have narrowed so too have their social class differences in death rates.<sup>21</sup> In Britain, both income and mortality differences have widened.<sup>22</sup> It is these divergent trends in inequalities which lie behind the contrasting gains in average life expectancy.

### **Psychosocial interpretation**

The fact that health seems to be influenced more by the scale of income differences than by the average level of income suggests that cognitive processes of social comparison are involved at some stage in the causal chain. Instead of health being affected through purely material and physiological channels, as for instance air pollution or exposure to carcinogens affect health, the importance of relative income suggests that psychosocial factors related to relative deprivation and disadvantage are involved. That is to say, it is less a matter of the immediate physical effects of inferior material conditions as of what the social meanings attached to those conditions make people feel about their circumstances and about themselves.

The need to develop a psychosocial understanding of the impact of material differences on health is indicated by three other aspects of the relationship between health and socioeconomic status. The first is that the socioeconomic gradient in health does not distinguish merely between the poor and the rest of society, but goes all the way up the social ladder. Thus people who own their own houses and have two cars are healthier than those who have only one,<sup>23</sup> administrative civil servants are healthier than executives,<sup>24</sup> and people in the highest

income group are healthier than those only slightly less well off (figure 4). Second, despite rising real incomes which have undoubtedly reduced the incidence of absolute material deprivation - however defined - among blue collar workers and their families during the second half of this century, in several countries their absolute mortality disadvantage appears to have increased. Thirdly, even in a country such as Britain, which is now one of the poorer developed countries, absolute living standards among the poorest 20 percent of the population is remarkably high. In 1992 some 72 percent had central heating, almost all had televisions and fridges or fridge-freezers, 72 percent had telephones and almost 60 percent had videos.<sup>25</sup>

Talking to people who suffer relative poverty, what is striking is the sense of stress and hopelessness, the high rates of depression, desperation and anxiety. Dominating people's whole consciousness, feelings such as these have a much more devastating impact on the quality of people's lives than the presence or absence of various consumer durables does in itself. What is almost inescapable in modern society are the powerful psychosocial connotations of material differences.

The socioeconomic gradient in mortality could only be attributed to the direct effects of material factors if it is assumed that increasing standards of comfort higher up the income scale have as much affect as material improvements further down. However, if this were so, then there would be a close association between changes in GNPpc and mortality rather than the weak association shown in figure 3.

There are a number of different points at which psychosocial links might be involved in the association between socioeconomic status and health. First, psychosocial processes emanating from a perception of one's status, economic insecurity or relative deprivation may impact directly on health. Second, psychosocial stress may affect exposure to behavioural risk factors. Thirdly, attempts by people on low incomes to maintain socially acceptable standards in more visible areas of consumption are likely to involve saving on food and other necessities which could damage health.

Income distribution also appears to be a powerful determinant of mortality rates in the third world - as has been shown statistically and by examples like Kerala, Cuba, Costa Rica and China.<sup>26,27,28</sup> But because the absolute standard of living is also clearly important in developing countries, the mechanisms linking income distribution to mortality almost certainly involve access to food and the distribution of land over and above the psychosocial processes related to relative deprivation.

### **Implications for economic growth.**

The influence of income distribution on health in developed societies indicates the psychosocial importance of the sense of social status, inferiority and relative deprivation in human social life. Indeed, it prompts the question to what extent each individual's desire for more income may be a desire to improve their relative standing in society rather than a desire for a higher standard of material consumption regardless of standards elsewhere in society. If that were an important part of it, then it would not be well served by general increases in prosperity which left relativities unchanged. Indeed, it would mean that it was not legitimate to sum individual desires for more income into an aggregate societal demand for economic growth.

At a time when the environmental impact of economic development is a serious cause for concern, and in developed countries there is growing evidence that the social fabric is decaying, the importance of these relationships can scarcely be overestimated. They should caution us against two assumptions: first, that gains in the quality of life are more likely to come from further undifferentiated economic growth than from improvements in the quality of the social fabric; and second, that the benefits of economic growth are sufficiently substantial to justify its ecological risks.

Because health is sensitivity to quantitative as well as qualitative aspects of both material and social change, it could be claimed that it is a better guide to the quality of life than any mix of economic indicators. Economic indices can only handle change by reducing qualitative change to quantitative increase, and by ignoring the all important social dimension of life. In this context, the fact that in the developed world the predominant position of material factors as determinants of health has given way to social factors, has profound implications.

Though starting from a very different point, the various attempts to make adjustments to GNP per capita to produce better measures of economic welfare have also cast doubt on the benefits of undifferentiated economic growth. What they show - at least in Britain and the United States - is that economic welfare has been declining despite increases in GNP per capita.<sup>29,30</sup>

As Mustard and Frank pointed out, we do not understand the impact of economic growth on the quality of the social environment. There are however reasons to think that its quality may now be the primary determinant of the subjective quality of human life. If this is so, then perhaps our primary goal should shift from promoting economic growth to creating a high quality social environment. Social environments which are less divisive, less undermining of self-confidence and which put greater resources into developing people's skills and abilities, may well turn out to be more innovative and better able to respond to environmental problems. Growth may then take up forms in which there is less reason to doubt the relationship between economic growth and the real quality of life.

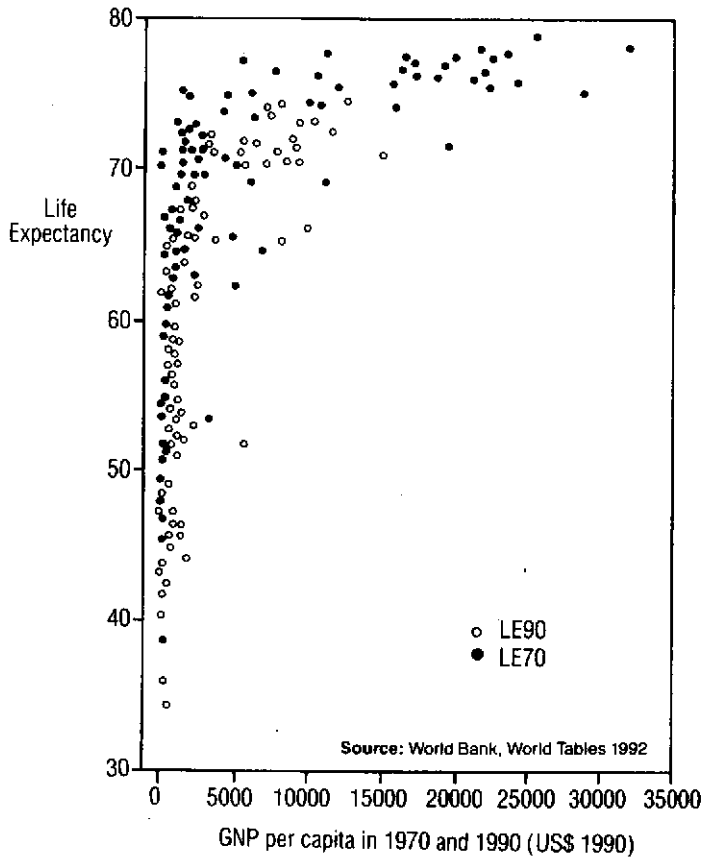
The task is to develop ways of being more selective about the nature and direction of growth. At the moment we need growth to avoid the discomfort of recession: to keep down levels of unemployment and to keep up profits. But more fundamentally, we need technical innovation which helps to safeguard the environment, and growth which serves to diminish relative deprivation. Fortunately, as well as the contribution which income redistribution makes to welfare, there is increasing evidence that it is also beneficial to productivity and growth.<sup>31,32</sup> But ultimately we must look beyond ways of making society fit for growth, and find ways of making growth fit for society.

## References

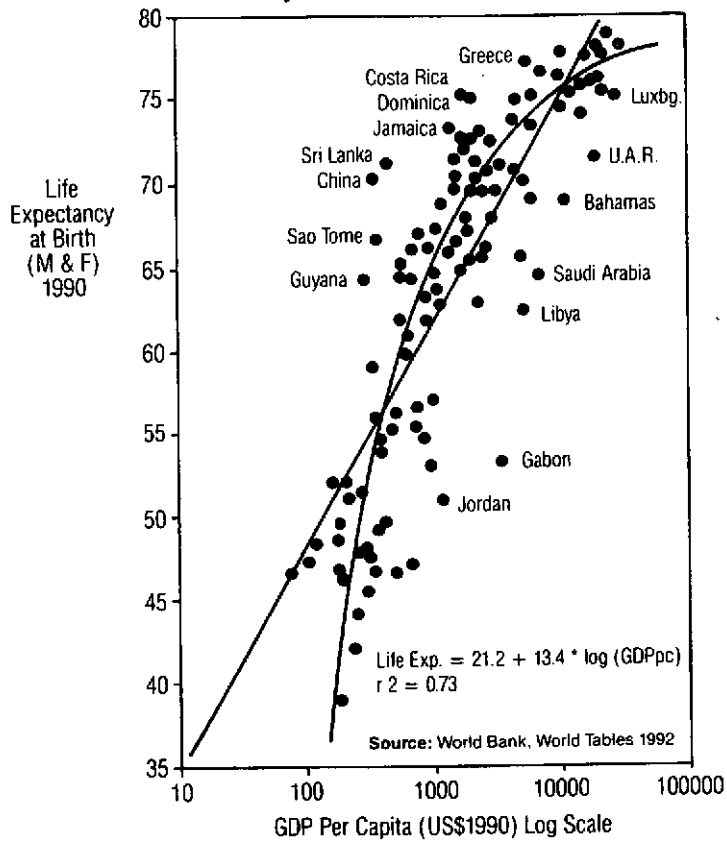
1. Burnet M, White DO. *Natural History of Infectious Disease*. Fourth edition, Cambridge, 1972.
2. Department of Health and Social Security, *Prevention and health: everybody's business*. HMSO, London 1976.
3. Wilkinson RG. The impact of income inequality on life expectancy. In: *Locating Health: sociological and historical explorations*. Edited by S. Platt, H. Thomas, S. Scott and G. Williams. Avebury, Aldershot. 1993.
4. Marmot MG, Adelstein A, Robinson N, Rose G. Changing social class distribution of heart disease. *British Medical Journal* 1978; ii: 1109-12..
5. Wing S. Social inequalities in the decline of coronary mortality. Editorial. *American Journal of Public Health* 1988; 78: 1415-6.
6. Antonovsky A. Social class and the major cardiovascular diseases. *Journal of Chronic Diseases* 1968; 21: 65-106.
7. British Medical Association. Report of the Committee on Nutrition. BMA, London 1950.
8. *The health of the school child, 1956-7 and 1960-1*, Report of the Chief Medical Officer, Ministry of Education, HMSO.
9. Stunkard AJ. Obesity and the social environment. In: *Recent advances in obesity research*, edited by A. Howard. Proceedings of the First International Congress on Obesity. London 1975.
10. Bourdieu P. *Distinction: a social critique of the judgement of taste*. Routledge, London 1984.
11. Macfarlane A, Mugford M. *Birth Counts: statistics of pregnancy and childbirth*. HMSO 1984.
12. Power C. National trends in birthweight and future adult disease. *British Medical Journal*, Forthcoming: accepted Nov. 1993.
13. *Health inequalities in European countries*, edited by AJ Fox. Gower, London 1989.
14. Wilkinson RG. Class mortality differentials, income distribution and trends in poverty 1921-1981. *Journal of Social Policy* 1989; 18 (3): 307-35.
15. Wilkinson RG. Income distribution and life expectancy. *British Medical Journal* 1992, 304, 165-8.
16. Wennemo I. Infant mortality, public policy and inequality - a comparison of 18 industrialised countries 1950-85. *Sociology of Health and Illness* 1993; 15: 429-46.
17. Wilkinson RG. Research note: German income distribution and infant mortality. *Sociology of Health and Illness* 1994; 16: 240-2.

18. Wilkinson RG. Income distribution and mortality: a 'natural' experiment. *Sociology of Health and Illness* 1990; 12: 391-411.
19. Kehrer BH, Wolin CM. Impact of income maintenance on low birthweight. *Journal of Human Resources* 1979; XIV: 434-62.
20. Wilkinson RG. Income and health. In: *Health, wealth and poverty*. Medical World / SHA, London 1993.
21. Marmot MG, Davey Smith G. Why are the Japanese living longer? *British Medical Journal*. 1989; 299: 1547-51.
22. Marmot MG, McDowall ME. Mortality decline and widening social inequalities. *Lancet* 1986; ii: 274-6.
23. Goldblatt P. (editor). 1971-81 *Longitudinal Study: mortality and social organisation*. OPCS Series LS 6. HMSO 1990.
24. Davey Smith G, Shipley MJ, Rose G. Magnitude and causes of socioeconomic differentials in mortality: further evidence from the Whitehall Study. *Journal of Epidemiology and Community Health* 1990; 44: 265-70.
25. Department of Social Security, *Households below average income 1979-1990/1*. HMSO, London 1993.
26. Sen A. Public action and the quality of life in developing countries. *Oxford Bulletin of Economics and Statistics* 1981; 43: 287-319.
27. Caldwell JC. Routes to low mortality in poor countries. *Population and Development Review*. 1986; 12: 171-220.
28. Rodgers GB. Income and inequality as determinants of mortality: an international cross-section analysis. *Population Studies* 1979;33:343-51.
29. Jackson T, Marks N. *U.K. Index of Sustainable Economic Welfare*, New Economic Foundation, 1993 (forthcoming).
30. Daley HE, Cobb JB. *For the Common Good*. Green Print, London. 1990.
31. Glyn A, Miliband D. (eds.) *Paying for Inequality*. Rivers Oram 1994 (in press).
32. Wade R. *Governing the market*. Princeton University Press, 1990.

**Figure 1**  
**Life Expectancy at Birth (M & F) 1970 and**  
**1990 In Relation to GNP Per Capita**

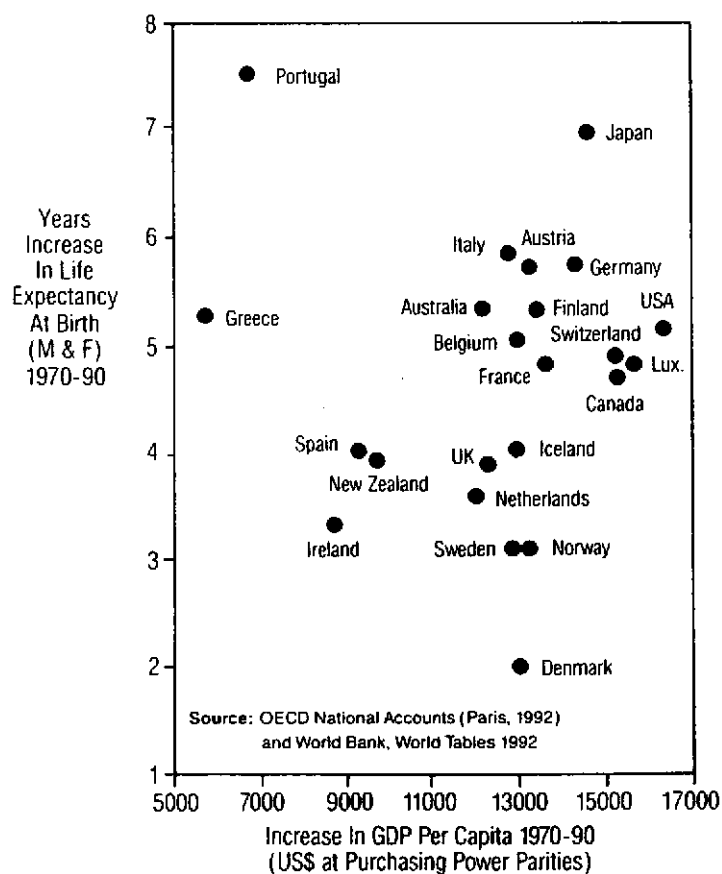


**Figure 2**  
**Life Expectancy and GDP**  
**Per Capita 1990**

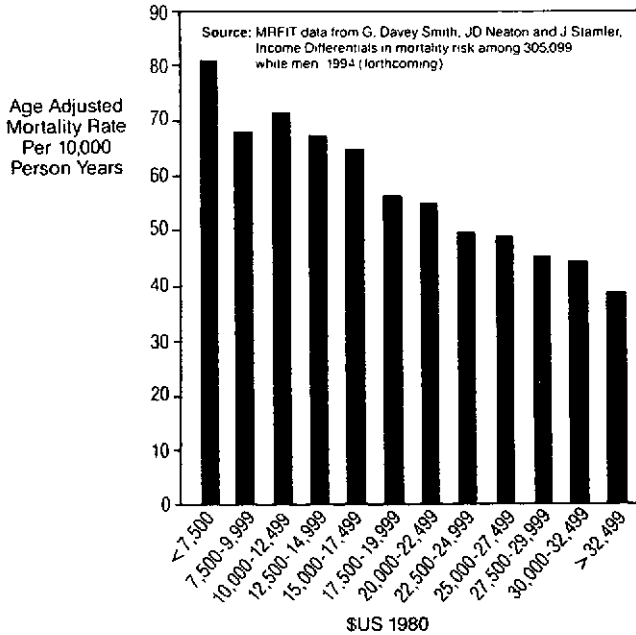


**Figure 3**

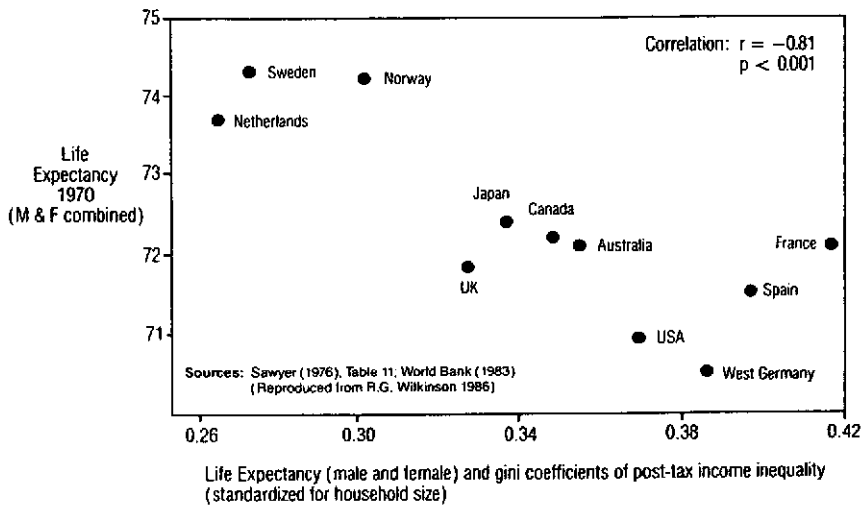
**Increases in Life Expectancy Related to Increases  
In GDP Per Capita in OECD Countries 1970-90**



**Figure 4**  
**Income and Mortality Among**  
**300,000 White U.S. Men**



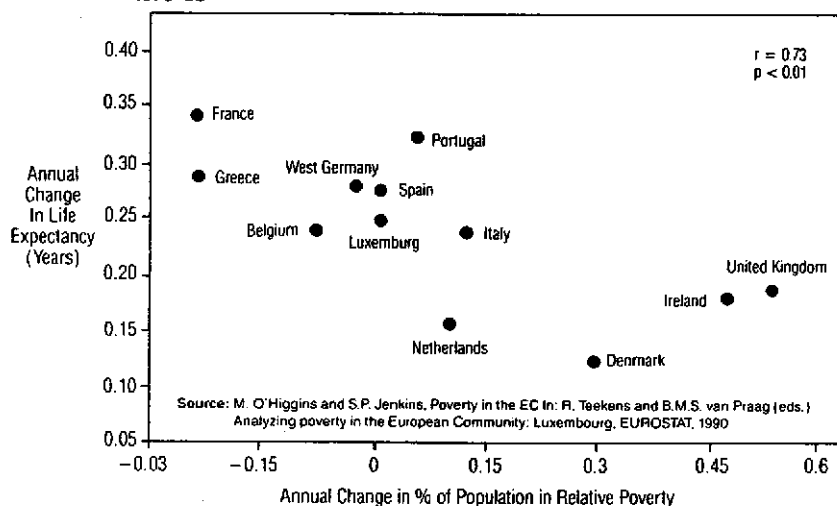
**Figure 5**  
**Life Expectancy and Income Distribution In**  
**OECD Countries**



**Figure 6**

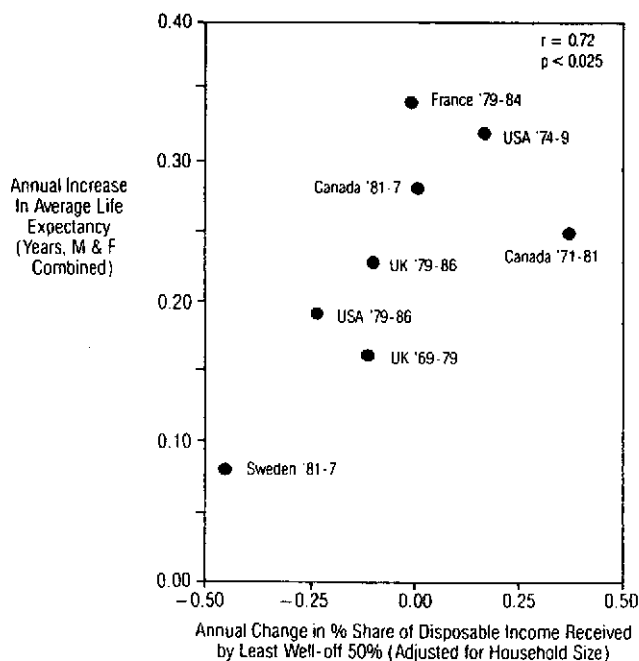
# **The Annual Rate of Change of Life Expectancy In 12 European Community Countries**

**With the Rate of Change in the Percentage of the Population in Relative Poverty 1975-85**

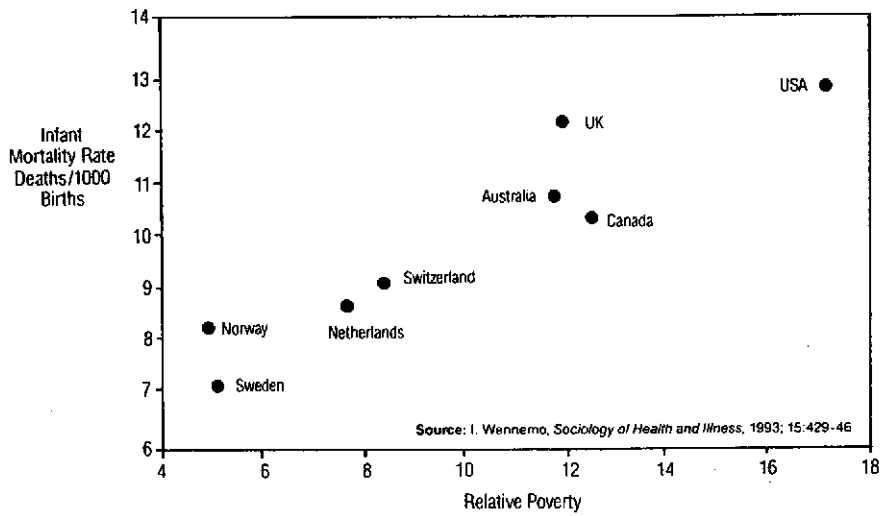


**Figure 7**

# **Changes In Income Distribution and Life Expectancy In Different Countries Over 5-10 Year Periods (LIS Data)**



**Figure 8**  
**Relative Poverty and Infant Mortality Rate**  
**for 9 OECD Countries 1980**



# Beyond Classical and Keynesian Macroeconomic Policy

---

*Paul M. Romer*

Ever since Adam Smith wrote his treatise on *The Wealth of Nations*, economists have offered advice about macroeconomic policy. In Smith's time, the goal of policy at the macro or economy-wide level was the one suggested by his title: Economic policy should increase the wealth of a nation. Today, the problem has been recast as one of sustaining year to year growth in a nation's Gross Domestic Product.

Smith observed that monetary tokens do not constitute the true wealth of a nation. Instead, he pointed to the inputs like tools, structures, and improved farmland that can be used in production. He concluded that a nation becomes more wealthy if its citizens and its government all refrain from excessive consumption and maintain a high level of saving. This will lead to the accumulation of more of these productive assets.

Smith's work contains many other perceptive observations about policy, but his emphasis of the importance of saving became a central part of what can be called the classical approach to macroeconomic policy.<sup>1</sup> In this approach, saving and capital accumulation are the central concerns. Excessive spending on current consumption, especially by the government, is perceived to be the most serious threat to sustained economic growth.

In the early part of this century, John Maynard Keynes argued that the classical fear of excessive consumption was misplaced, even dangerous. To Keynes and his followers, the biggest risk was that current spending might be too low, not too high. Classical fears of government deficits stood in the way of action by the government to increase spending in times when it spending was low.

By the middle of this century, Keynesians had carried the day. Since then, macroeconomic policy in many industrialized countries has been based on the Keynesian strategy of using monetary and fiscal policy to increase "aggregate demand." Keynesians admitted that in principle, there could be such a thing as too much spending and too much aggregate demand. Nevertheless, circumstances almost always seem to call for more stimulus and spending rather than less.

As dissatisfaction with the results from Keynesian demand management has grown over the course of the last decade, the classical approach to macroeconomic policy has made something of a comeback. Many economists now call once again for smaller government deficits, higher private savings, and more rapid capital accumulation. They argue that excessive government spending reduces capital accumulation and that capital accumulation is the key to growth.

For someone who is not an economist, this oscillation between calls for more saving and more spending is more than a little frustrating. When the newspaper reports that consumer spending has increased — and that private savings has decreased — is this good for the economy or bad? To people who are careful students of business, science, and history, there is also something suspect about the entire discussion. Doesn't the key to economic success lie in a nation's ability to introduce valuable new goods, to improve the quality of existing goods, and to find more efficient ways to manufacture and deliver goods? If so, why do economists seem to devote so much attention to monetary and fiscal stimulus on the one hand and pur-

chases of existing capital goods like blast furnaces and fork lifts on the other? Where is the discussion of innovation, invention, discovery, and technical progress?

Economists, for their part, have often tended to dismiss as poorly informed or misguided the kinds of concerns expressed by outsiders. During the "competitiveness" debate triggered by trade imbalances of the 1980s, economists were quick to point out any errors made by non-economists in their discussion of exchange rates and patterns of trade. They were not, however, as quick to confront the deeper issues that motivated the public expressions of concern. Nor were they sufficiently willing to explain their internal debates in terms that allowed outsiders to follow what was going on.

This essay tries to address the concerns of the baffled and skeptical outside observer of macroeconomic policy debates. It summarizes the logic behind both the Keynesian and classical policy positions in non-technical terms. It also describes how a body of work in economics, emerging under the label of "endogenous" or "neo-Schumpeterian" growth theory, validates some of the concerns about competitiveness, innovation, and discovery expressed by people who are not economists. This recent work suggests that both the save-more and the spend-more macroeconomic policy prescriptions miss the crux of the matter. Neither adjustments to monetary and fiscal policy, nor increases in the rate of savings and capital accumulation can by themselves generate persistent increases in standards of living. To achieve long-run success, economic policy must support the institutions that generate technological progress. This sounds simple enough, but policy-makers must also resist the temptation to impede change when it causes temporary disruption. In a modern industrial democracy, achieving a balance between support for economic progress and tolerance of economic change is not a simple task.

### **I. Three Overused Macroeconomic Metaphors**

Before showing why classical and Keynesian policies offer an incomplete description of the policies needed for long-run economic success, it is important to summarize the vision of the economy that motivates each approach. Three basic metaphors have guided most of the discussion in macroeconomics: The economy is like a farmers' market; the economy is like a family that should save rather than borrow; and the economy is like a pump that needs priming. The farmers' market and family finance metaphors can be traced all the way back to Smith and are central to the classical vision. Keynesian macroeconomics also makes use of the farm market metaphor, but augments it with the notion of pump-priming.

Taken all by itself, the notion of an economy as a farmers' market leads to the position of *laissez faire*, the assertion that governments should leave markets alone. In the mid-1800s, the French journalist and popularizer of economic ideas Fredric Bastiat gave one of the classic statements of the miracle of the market:

On entering Paris, which I had come to visit, I said to myself – Here are a million human beings who would all die in a short time if provisions of every sort ceased to go towards this great metropolis. Imagination is baffled when it tries to appreciate the vast multiplicity of commodities which must enter tomorrow through the barriers in order to preserve the inhabitants from falling prey to all the convulsions of famine, rebellion and pillage. And yet all sleep at this moment, and their peaceful slumbers are not disturbed for a single instant by the prospect of such a catastrophe. On the other hand, eighty [regions of France] have been laboring today, without concert, without any mutual understanding, for the provisioning of Paris.

Anyone who examines the wreckage of the centrally planned economies of the former Soviet Block cannot help but be convinced that there is a very large measure of truth in this testament to the power of markets.

Debate continues, nevertheless, over the kinds of institutional arrangements that provide the most effective support structure for a market system. With the possible exception of extreme anarchists, no one denies that formal institutions of governance are necessary to defend an economy from external aggression and to establish an internal system of law. Without these institutions, a market economy cannot exist. Most economists, even those who are most resistant to government intervention in other domains, are convinced that the government must also take the lead in establishing the basic elements of a monetary system of exchange. There is much less agreement about more subtle institutional questions such as how the government and the private sector can best structure their interaction with universities or how labor law, patterns of regulation, and the workings of financial markets can affect the rate at which new firms enter and compete in the market. In a modern economy, these kinds of institutional arrangements may be extremely important for the long-run economic success of an economy, but traditional macroeconomic policy analysis has tended to ignore them. It has focussed almost exclusively on monetary institutions and government expenditure.

Once one concedes the necessity of a government, one must address the question of how to arrange its finances. This is where the family finance metaphor comes in. A government that has outstanding debt is like a family that has a chronic outstanding balance on its credit card. If the income of the family goes up over time, the credit limit on the credit card will also go up, and so will its total outstanding debt. The consumer's total outstanding debt can continue to rise over time, provided that it does not rise faster than the consumer's income. So can the debt of a nation.

Popular discussion of government deficits conflates two distinct issues. The first is that a government cannot issue unlimited amounts of debt. Just like a family, it faces a limit on its total ability to borrow, determined by its ability to make the required payments on its debt. For a country as a whole, the ability to pay is limited by total income or Gross Domestic Product. Limits on borrowing mean that it is not possible for a government to consistently spend far in excess of its tax revenues. A government can do so temporarily, issuing large amounts of debt to finance the excess spending. Eventually, spending and revenue must be reconciled because the government will eventually reach its credit limit. In countries where debt is growing rapidly, economists urge immediate action because good policy, adopted in a timely fashion, can resolve a discrepancy between spending and income before access to credit is withdrawn. Bad policy and inaction will lead eventually to financial crisis with major economic costs when access to credit is abruptly cut off, as happened to many countries in Latin America during the debt crisis of the 1980s.

There is no dispute that the ratio of government debt to GDP cannot increase forever. In practice, the only debate in any particular country with large deficits is how high this ratio can go before the country starts having difficulty borrowing more. There is, however, a second point which sparks disagreement between classical and Keynesian economists. Suppose that a government behaves like a family that always carries the maximum possible balance on its credit card. Debt increases at about the same rate as total income in the economy, so access to credit is not in danger of being cut off. Is this policy bad for the economic growth of this nation?

For classical economists, the family finance metaphor applies once again. If a family that has a high outstanding level of debt reduced its level of spending and used the extra income to pay off its debts and accumulate money in the bank, its wealth would grow more rapidly. Extrapolating from the family to the nation as a whole, the implied policy position is that

income growth for the nation would be more rapid if the government did not issue debt. Income growth for the nation would also be more rapid if the government undertook measures that encouraged saving by private individuals, or at least refrained from imposing tax and transfer policies that discourage private saving. According to the classical economists, lower deficits and higher saving are valuable because they increase the rate of growth of national income. They are important policy goals even if there is no danger of a credit crisis.

The classical vision of economic growth is therefore one of a market system that functions fairly well but must be accompanied by a government that could dissipate the savings of its citizens with excessive deficit spending. Despite the simplicity and intuitive appeal of this policy model, support for it diminished steadily among economists in the middle of this century. Economists moved steadily away from the classical notion that maintaining high levels of public and private saving should be the primary goal of macroeconomic policy. In its place, macroeconomists accepted an entirely different model of the interaction between the government budget deficit and the market economy. In Adam Smith's time, a monarch who emptied the public treasury to build public works was likened to a spendthrift who dissipated the family wealth. Now a government that undertakes public works is said to be engaged in "short-run fiscal stimulus."

According to the Keynesian metaphor that justifies the concept of stimulus, the economy is like a pump that cannot raise any water from the ground until it has first been primed by adding some water to it. This metaphor of a pump that needs priming has a mathematical equivalent that has become a part of the public lexicon. Government spending is said to be associated with a "multiplier." One additional dollar of government spending is said to bring into existence many additional dollars of output from the private sector, just as one gallon of water poured into the pump will generate many more gallons of water from the pump.

This assertion sounds too good to be true. It promises something for nothing. The classical argument concerning savings says we can have more wealth only if we give up some current consumption. The Keynesian argument about pump-priming suggests that we can have both more government spending and more private income. Despite its implausibility, two developments helped build support for the Keynesian view and erode support for classical policies. The extreme circumstances of the Great Depression of the 1930s suggested that government pump-priming might sometimes be worth trying. In addition, theoretical analysis of diminishing returns to capital accumulation suggested that the costs of abandoning the classical emphasis on national savings and investment might not be so high as suggested by the classical economists.

During the 1930s, unemployment in North America was high and much productive capacity sat idle. If government spending could put these unemployed resources back to work, it was argued that this spending could dramatically increase current output. In the United States, there is some evidence that is partially consistent with this view. Between 1941 and 1944, the Federal Government increased its purchases of goods and services by \$130 billion (in 1958 dollars) because of the war effort. During this same 3 year period, total output increased by about \$100 billion, from \$260 billion to \$360 billion. An additional dollar of government spending translated into about 75 cents of additional output, so that private consumption hardly fell as government spending increased. Unemployed resources were also put back to work. The unemployment rate fell from about 10% to about 1%. After the macroeconomic difficulties of the 1930s, deficit spending by the government seemed to be a major policy success. The correct conclusion should probably have been that government spending can have unusual effects in unusual circumstances. The conclusion that was drawn, however, was that government spending had the power to dramatically increase economic well being.

The case for aggregate demand management and against too much concern about deficits and saving seemed to become even stronger during the 1950s and 1960s. As economists became more careful in their analysis of the economy, the classical family-finance metaphor seemed to crumble. Economists began to ask what would actually happen if everyone in an economy tried to save more and consume less. They recognized that the entire structure of production in the economy would have to adjust. Some factories and workers who used to produce consumption goods would have to stop producing them because the total demand for consumption goods would go down. These workers and factories would have to shift into the production of investment goods like machine tools, blast furnaces, fork lift trucks, etc. that could be used to produce more consumption goods in the future.

The point emphasized by Keynes and his followers was that this process of adjustment might get stuck halfway. If savings went up, some workers and factories would cease to produce consumption goods, but these workers and factories might not be put to use producing investment goods. There might not be enough demand for investment goods. In fact, as unemployment rose and income fell, total demand for investment goods might fall rather than increase because investment spending tends to fall when the economy enters a slump. If so, policy makers faced what Keynes called the paradox of thrift. More saving could lead to less investment by causing unemployment and underutilization of existing resources. It was in this context that government measures designed to increase aggregate demand were thought to have the potential for increasing total output in the economy.

During the 1950s and 1960s, economists pointed to a further difficulty with the family finance metaphor. Accept the optimistic assumption that when consumption goes down, all of the resources that are put out of work in the consumption good sector are shifted into the production of investment goods like fork lifts and blast furnaces. Would this really increase the output of consumption goods in the future by very much? Having more capital goods like fork lift trucks and blast furnaces might not be as much of an advantage as was once thought.

The problem an economy faces is what economists call "diminishing returns." In handling heavy objects, a fork lift is a very useful piece of equipment. The first fork lifts that were introduced in the economy were very productive. But as more and more of them accumulate, their value falls sharply. As we increase the total number of fork lifts that are installed and move from a position with one fork lift truck for every million workers in the economy, to one for every 10,000 workers, ending with one fork lift for every 10 workers or even one per worker, the value of each additional fork lift drops rapidly. Eventually, additional fork lifts would have no value and become a nuisance. The "return" on investment in an additional fork lift "diminishes" as more fork lifts are produced and eventually becomes negative.

Economists concluded during the 1950s and 1960s that even in the best case, an increase in the savings rate for a nation or for the world as a whole would increase the stock of factories and fork lifts, but the main effect of doing so would be to drive down the price of existing fork lifts and factories. More of these goods would not produce much additional output. There are, after all, only so many workers to go around, and only a limited number of them can usefully be put to work driving a fork lift. Economists therefore concluded that although saving more can be a successful strategy for a family that wants to increase its wealth, it will be far less successful if everyone in a nation tries to do it at the same time. There are not enough new opportunities to employ the additional savings, and the existing opportunities will not pay a high return on investment. More precisely, these economists argued that increases in saving could give a boost to output when an economy is short of capital and lacks things like fork lifts and factories, but will provide far less "bang for the buck" after the economy has become affluent and the return to capital accumulation has fallen.

The point of this argument is easier to understand if it is put in an historical context. Suppose that in 1800, the citizens of North America had decided to substantially increase the aggregate savings rate. This would have led to much more investment in the capital goods of the day – sailing ships, water wheels, ox carts, etc. Because of diminishing returns, the additional benefit from more water wheels and ox carts would have been small.

It was this emphasis on diminishing returns that first earned for economics the label “the dismal science.” The notion of diminishing returns suggests that the good times when the returns on investment were high are long gone and that the remaining room for income growth is limited. Yet as anyone knows, standards of living have continued to increase dramatically over time. During the nineteenth century, the economic performance of industrial nations was anything but dismal. Something caused economic growth to persist and even accelerate. If it was not the accumulation of extra ox carts and water wheels, what was it?

From the perspective of history, the answer is obvious. Improvements in standards of living during the nineteenth century came from new discoveries like the steam engine, the telegraph, better techniques for making steel, and machine tools that could shape metal. After the economy was relatively well provisioned with ox-carts and the return on investment in ox-carts had fallen, the invention of the railroad created an entirely new set of opportunities for productive investment. In this century, the returns to additional investment in railroads fell, but we have had the opportunity to invest in other new goods like trucks, skyscrapers, and computers. In the long run, it is invention and technological change that create new opportunities for investment and determine how rapidly an economy grows. Diminishing returns arise in any particular fixed investment activity, but discovery continually creates new activities where returns on investment are high.

In the 1960s, the classical prescription of increased savings therefore seemed doubly damned. In the worst case, a reduction in demand for consumption goods could lead to a recession or even a depression because of the paradox of thrift. In the best case, it could have only a small positive effect on output because of the fork lift problem — diminishing returns in the existing set of investment opportunities. As a result, Keynesian macroeconomists constructed a new image of how the government could support the operation of the market. Like the classical economists, the Keynesians in North America generally understood that markets had to be the central institution for coordinating economic activity. In their vision, however, markets suffered from an inherent instability. They might fail to generate enough demand for output. If they did, this could set off a cascade of events that could lead, in the worst case, to a repeat of the events of the 1930s. Keynesians therefore proposed that the market be supplemented by a government that stood ready to increase demand for goods – to prime the pump – whenever growth slowed down.

They concluded that there was no particular reason to worry about increasing savings and investment. Extra capital was not that important. Increases in standards of living would come from technological change, which, they assumed, the miracle of the market would provide. Without pretending to know in detail where improvements in techniques of production and new investment opportunities came from, macroeconomists assumed that we could rely on private markets to generate an unceasing flow of these improvements, just as we have always relied on the uncoordinated efforts of French farmers to provide a steady supply of food for Paris.

The government had two ways to stimulate demand. It could run a deficit by spending more than it received in tax payments or it could have the central bank increase the amount of money in circulation. Government deficits might reduce the total amount of investment in capital, but if capital accumulation was not that important, the government might as well try to replicate the experience of 1941 to 1944 and reduce the unemployment rate through deficit spending. Stimulus from the central bank raised the risk that inflation would occur. But just

as economists of the 1960s were convinced that the true costs of government deficits were low, they were unconcerned about the prospect of small increases in the rate of inflation. According to this policy of "fine tuning," the government would adjust fiscal and monetary policy on a month to month or year to year basis to avoid any recessions or slowdowns in the short run. Without any attention from policy makers, technological change would automatically take care of the long run.

In the 1970s, 1980s, and 1990s, things did not turn out as planned. After three decades of experience with fine tuning, we have little to show for our efforts. The average unemployment rate is no lower than it was when these policies were initiated. In fact, in many countries in Europe, it is higher than it has been since the Great Depression. Deficit spending has been used so frequently that there is growing concern that the government in several industrial economies may face chronic difficulties servicing the interest on its debt. Loose monetary policy may have increased economic growth during some short intervals of time, but these apparent benefits have largely been offset by the sharp short-run reductions in growth that followed attempts by the monetary authorities to control the inflation that earlier periods of stimulus had induced. Overall, there has been no obvious reduction in the variability of output due to recessions and booms.<sup>2</sup> Moreover, the average rate of growth has been lower than it was during the 1950s and 1960s. The only clear result from the decision to stand ready to prime the pump with fiscal and monetary stimulus is that industrial democracies have generally experienced larger government budget deficits and more inflation.

## **II. Stimulus, Conditioning, and Technique**

Because of the apparent failure of fine tuning, the economics profession is in the midst of a turn away from short-run economic stimulus. Unfortunately, this turn frequently takes the form of a pendulum swing back toward the classical policy stance and a family finance metaphor that equates saving with growth. Work done in the last 10 years on the theory of growth is sometimes invoked to support this return saving, as if this theoretical work had somehow diminished the force of the fork lift problem and removed diminishing returns to capital accumulation as a constraint on growth. This is not the right interpretation of recent growth theory. What this theory calls for is not a pendulum swing back to the classical emphasis on brute force savings and capital accumulation, nor even some middle position that balances the need for savings against the benefits of fine tuning aggregate demand. Instead, the thrust of this new work pushes policy analysis and economic theory in an entirely different direction, focussing attention directly on the determinants of discovery, invention, innovation, and technological progress.

Within the economics profession, this new direction in theory and policy analysis is being explored with the use of mathematical equations and statistical analysis.<sup>3</sup> It is possible, nevertheless, to give an interpretation of this work in terms of a metaphor that is as concrete as the image of the farm market, the family that needs to save, and the pump that needs priming. This interpretation makes use of the tired rhetorical device of a sports metaphor but relies on an unfamiliar sport. Imagine that an economy is like a swim team. Creating a higher standard of living is like helping swimmers achieve faster times in their races. The challenge for an economy is to achieve an increase in the standards of living each year. The challenge for a swim team is to achieve an increase in the speed of the swimmers each year.

To appreciate this metaphor, it helps to know a bit about the variety of methods used by coaches to increase the performance of their swimmers. Some methods increase race times in the very short run. Others improve physical conditioning or develop better technique and raise performance in the long run.<sup>4</sup>

Two approaches are used to give swimmers a short-run boost just before a race: tapering and blood doping. The principle behind blood doping is well understood. Hemoglobin is the

carrier of oxygen in the blood. The higher the concentration of hemoglobin, the more oxygen a swimmer can transport to her muscles during a race. Blood doping refers to the process of drawing blood from a swimmer for long before a race and then reinfusing the blood just before a race. When blood is withdrawn, the body replenishes its stocks of hemoglobin. When the withdrawn blood is reinfused, the swimmer will temporarily have an excess of hemoglobin and an increased oxygen carrying capacity.

There are two problems with blood doping. The first is that when blood is being withdrawn to build up a stock in storage, the swimmer's performance is reduced. The second is that if too much blood is reinfused, the swimmer can experience serious medical complications from sludging or thickening of the blood.

These "periodic injections of liquidity" are obviously intended to suggest the macroeconomic injections of liquidity (i.e. money) central to fine tuning. Liquidity injections lead to a short-run improvement in performance but are offset by the costs of liquidity withdrawals. Both kinds of injections can cause potentially serious complications if they are used too aggressively — sludging for injections of blood and inflation for injections of money. To appreciate part of the story behind post-war macroeconomic policy, imagine that the medical trainer for our economy has been using blood doping aggressively, achieving some short-run increases in performance, but pushing us right up to the limit where the complications from sludging (i.e. inflation) become unacceptable. Then the trainer withdraws blood, causing a sharp reduction in performance. This creates political pressure for more short-run stimulus, and the boom-bust cycle starts again.

In swimming, the other method for increasing performance in the short run is a taper. A taper refers to the process whereby an athlete follows a training schedule that tapers off in intensity in the weeks before an important race. Most areas of competitive sport are governed by the iron law of athletics: no pain, no gain. The taper, however, seems to exploit a "paradox of training" that parallels the paradox of thrift emphasized by Keynes. Training hard just before a race tends to reduce race times. Working less has the double benefit of reducing discomfort in practice and increasing speeds in a race.

Coaches and politicians find that their popularity goes up when they institute a taper or undertake fiscal stimulus. As a result, it is no surprise that in the politics of macroeconomics, both the left and the right have developed their own distinctive versions of fiscal stimulus. Politicians on the left promise an investment-led recovery of growth driven by deficit financed government spending. Politicians on the right promise a supply side recovery that is driven by deficit financed tax cuts.

The analogies between blood doping and monetary policy on the one hand, and between tapering and fiscal stimulus on the other, are consciously chosen to invoke other parallels. There are circumstances in which injections of blood and money are clearly appropriate. Despite the pejorative sound of the term, blood doping can be a valuable medical procedure. If someone knows in advance that he will have to undergo a surgical procedure that has a high risk of blood loss, it is a good idea to store up some blood in advance. If he suffers a major loss of blood, transfusions of his own blood can be a crucial life-saving measure. Symmetrically, economists now understand that a major contributing factor to the depression of the 1930s was the decision by the Federal Reserve Board to stand idly by and watch as the financial system in the United States hemorrhaged. Between 1930 and 1933, the money supply fell by about a third as the banking system collapsed.<sup>5</sup>

Although transfusions are valuable during real emergencies, this does not commend blood doping as a regular part of athletic training. Transfusions can be very effective in averting death, but they cannot permanently raise an athlete's performance by much. In the same way, the power of monetary policy to cause a sharp recession and to speed recovery from one

does not imply that we can rely on weekly or monthly adjustments in the supply of money or the level of interest rates as a means of achieving steady increases in standards of living.

The point of the analogy between tapering and government budget deficits is that both offer short-run benefits in exchange for long-run costs. The impression that tapering or deficits can give something for nothing is an illusion. A reduction in training does give a short term boost, but it impedes progress toward better aerobic conditioning and more strength. Conditioning, strength, and technique all improve with the total distance that a swimmer covers in practice. Tapers reduce this total. Symmetrically, if government deficits have a positive effect on short-run output, it comes at the cost of a reduction in national saving and the rate of accumulation of capital. In the long run, less capital will translate into less output.

This does not mean that tapers and budget deficits are useless. If you are a coach training a team for an important swim meet or a politician campaigning for re-election, you may be quite willing to give up some future performance to get some short-term gain. But it does mean that there is no free lunch.

The point that economists made about diminishing returns also has an exact analog in swimming. Economists observed in the 1960s that an economy cannot grow forever by accumulating more and more fork lifts. In swimming, it is equally true that a swimmer cannot continue to improve her times year after year merely by training harder and harder to improve aerobic conditioning and physical strength. The physics of drag in water and the physiology of the human body impose an upper bound on the speed at which anyone can move through the water using a particular swimming stroke. There is an absolute limit to how far a brute force approach to training can take a swimmer.

In swimming, as in macroeconomics, we nevertheless continue to see steady improvement. From the 1950's up through 1990, the number of world records that were broken in an average year has remained about the same.<sup>6</sup> This improvement derives fundamentally from the same source in both swimming and the economy. In swimming, there has been a dramatic improvement in the techniques that swimmers use to propel themselves through the water. This exactly parallels improvements in the techniques of production that have raised our standards of living.

One of the advantages of using a swimming metaphor to think about the sources of long-run improvement in performance is that it is relatively easy to see what the technical innovations are. From about 1500 to 1900, what we now know as the breast stroke was thought to be the most efficient swimming stroke in the Western world. It was modeled on the swimming motions of the frog. This represented an important advance beyond the previous model for human swimmers, the dog.

The breast stroke, in which the arms remain underwater throughout the stroke, was used throughout most of the nineteenth century in racing competition. The first person to swim the English Channel did so using the breast stroke in 1875.<sup>7</sup> For readers who do not appreciate how inefficient the breast stroke is, with its underwater recovery of the arms and legs and with the chest plowing through the water like a barge, swimming the Channel with the breast stroke today would be like running a marathon backwards.

Around the turn of the century, English and Australian swimmers copied techniques used by native peoples in Australia and Ceylon and developed what is now known as the free style or crawl stroke. As early as 1844, a version of this novel kind of stroke was demonstrated by two native Americans in England at a time when it was the center of worldwide swimming competition. But the inherently inefficient breast stroke had been refined, and early attempts to use an overarm recovery of the crawl stroke were crude and awkward. (One of the biggest initial stumbling blocks was the lack of a kick that would work with the new arm stroke.) As

a result, the efficiency advantages of having the arms recover out of the water were not generally recognized for more than 50 years.

One of the chronic problems that people have in understanding the potential for discovery is that we extrapolate our current circumstances both forward and backward in time. Most people expect that the basic swimming strokes have been known for centuries and are surprised to learn that something as obvious as the crawl stroke was discovered at about the same time as quantum mechanics. People also tend to underestimate the enormous gulf that separates the discovery of abstract principles from their effective application. The basic physical laws of fluid dynamics were established in the 1700s, but even with all the sophisticated laboratory experiments can be done today, there is still disagreement among coaches and swimmers about what constitutes the most efficient swimming stroke. New techniques continue to be tried.

To get some sense of how much scope there must be for improvement in stroke mechanics, consider the details of the arm motion used in the crawl stroke. The arm creates force through a combination of a pushing motion like that of a piston and side to side "sculling motions" that let the hand act like an air foil, generating force using the same principle as an airplane wing or propeller. According to the best current understanding of the crawl stroke, the hand makes a "Z" as it moves through the water, moving out away from the centerline of the body after entry, coming back in under the body, and then moving out again in the final motion. For each of these three motions, the swimmer has to choose the length of the movement in or out, the speed of the hand, and the angle "of attack" that the surface of the hand makes relative to its motion through the water. To quantify the scope for experimentation, suppose that for each of these three parts of the arm stroke, there are ten different lengths, ten different speeds, and ten different angles a swimmer could try. A simple mathematical calculation shows that there are one billion different variations on the basic stroke that can be tried.

Once one takes account of all the other choices that have to be made, it becomes clear that even this calculation only begins to suggest the scope for experimentation. The swimmer must adjust the length, speed, and frequency of the kick. In the 1950s, people found that a swimmer could reduce drag by rolling from side to side with each stroke instead of remaining flat on the stomach and plowing through the water like a barge. Finally, in addition to setting the parameters for the pull, the kick, and the roll, the swimmer must also decide whether to breathe on one side or on alternating sides, and how to time the breaths relative to the arm strokes. Lest anyone who is not a swimmer might think that some of these different variables are irrelevant, they should understand that the best swimmers adjust all of them according to the distance that they will swim in a particular race.

Given all of the possibilities, it is inconceivable that anyone has yet achieved the perfect swimming stroke in any event. In just the last 10 years, coaches have concluded that the length of the side to side motion of the hand should be much smaller than they had previously thought. In this, as in many other areas, there will always be room for disagreement, innovation, and improvement. In the same way, every productive activity in our economy is susceptible to an enormous amount of experimentation and improvement.

### **III. Policy Implications**

The swimming metaphor leads to a more nuanced view of how the government should set economic policy than either the family finance or pump-priming metaphors. It recognizes that the monetary and fiscal policies adopted by the government do matter. Macroeconomic mismanagement of the kind experienced in the United States during the 1930s should surely be avoided. A financial panic that risks the stability of the financial system also calls for

decisive intervention by the monetary authority just as a massive bleed calls for a blood transfusion during a surgical operation. Because tax revenue falls during a recession and rises during an expansion, it makes sense for the government to aim for budget balance on average but not in each year. It should tolerate modest deficits during recessions and run surpluses during expansions much as a swim coach might reduce the intensity of workouts when a swimmer has a cold and train more intensively when she is healthy. A coach may use a taper to get a special burst of speed just before an important race; policy makers might also decide to run a larger than usual deficit if there is a temporary need for government spending, during a war for example.

It is also clear that capital accumulation is as important to economic success as conditioning is to athletic success. The lesson from diminishing returns is that investments in capital goods alone cannot generate perpetual increases in standards of living, any more than increasing the length of swimming workouts can by itself lead to steady increases in speed. But this is no excuse for getting out of shape. In swimming, the trend over time has been towards more arduous training. In part, this reflects an increased appreciation for the benefits from training; diminishing returns to conditioning do not set in as soon as people once thought. It also is the result of accumulating knowledge about technique. It is not possible for a novice swimmer to internalize and make use of all of the knowledge of stroke mechanics that coaches and other swimmers have accumulated unless she is willing to spend many hours in the pool each day, practicing and perfecting what she has been taught.

The recent return to a classical policy stance has helped offset the constant pressure for increased spending and Keynesian stimulus. It reminds us that capital investment, like conditioning, does matter. Recent proponents of the classical view have also emphasized, correctly, that human capital is every bit as important as physical capital. Education and on-the-job training are investment activities that are just as important as building blast furnaces and fork lifts. Just as more investment in pool time is needed to take advantage of all the advances in swimming, more investment in education and on-the-job training will be required to take full advantage of technological change in economic affairs. New equipment will sometimes be required if an economy is to improve continuously. Curious, observant, well trained, and highly motivated workers will almost always be required. Worldwide competition has always forced the goal of continuous improvement on swim coaches, and it is increasingly forcing it on all aspects of business practice.

So governments do need to pay attention to the textbook fundamentals of macroeconomics and they do need to encourage, or at least avoid discouraging, the accumulation of human and physical capital. But these steps by themselves are not enough. A government must create an environment that fosters change – and progress – in the techniques we use. Especially in discussions of education and human capital accumulation, it is easy to be seduced by a suitably extended version of the family finance metaphor. If I am a high school educated worker and if all of my children become college educated professionals, then their standards of living should be higher than mine, or so this extended version of the family finance metaphor goes. I can invest, therefore, not by putting money in the bank, but by sending my children to college.

This makes sense at the level of the family, but in the economy as a whole, a strategy based exclusively on human capital accumulation eventually runs into the same limits from diminishing returns that arise with physical capital accumulation. Suppose that there had been no innovation and no technological change during the nineteenth century. We could have increased the fraction of the population that was high school educated or even college educated. We could also have accumulated lots more sailing ships, water wheels, and ox carts. But eventually we would be forced to admit that we had little use for one more college graduate who is employed driving one more ox cart. The increases in standards of living that

we achieved in the last century were possible only because of the discoveries and innovations that let new physical capital and new human capital be put to work in high return activities.

For a government, there are two parts to any strategy for creating an environment that fosters progress and change. The first is the one that most people think of in discussions of technological progress. The government has a role to play in supporting the research and training missions of science and engineering. An implicit assumption in both classical and Keynesian approaches to policy is that the farm market metaphor applies to the production of technological knowledge. The market, they assume, will somehow provide the right amount of technological advance. Yet economists have long understood that the production of knowledge is an area where it is impossible even in principle to "get the prices right." (Think of calculating the right price for the discovery that disease is spread by germs and that doctors can save lives by washing their hands between examinations of different patients. Before anyone discovered these facts, society should have been willing to offer a very high price for knowledge that can reduce sickness and death. But after these facts were discovered, the right price for them is zero. This knowledge should be given away freely because this leads to the most widespread and efficient use.)

Because it is not possible to get prices for knowledge right, markets by themselves cannot achieve the best possible solution to the problem of discovering and distributing knowledge. Other mechanisms like royal patronage, and its modern descendent, the government research grant, can help support the process of discovery. So can subsidies for education and training. But in many areas, market incentives also provide crucial guidance to the process of discovery. Imagine how much less efficient the market for computers would be if all computer software was paid for and distributed by a government bureaucracy like the postal service. As a result, the best institutional arrangement for supporting technological advance in any particular area requires a delicate mixture of private incentives and public institutions.<sup>8</sup>

Given its importance in the growth process, government support for technological progress deserves more careful study than it has received from the economics profession. Far more attention has been devoted to the study of the institutions and political economy of monetary policy than to the institutions and political economy of government support for research and education. To cite just one fact that should, but does not, attract attention from policy-oriented economists, the number of North American students attending graduate schools of engineering has plummeted in the last two or three decades. The general consensus seems to be that we can rely on imported foreign labor to teach engineering and to staff the engineering departments in our firms, much as we depend on it to pick our farm produce. In the next century, economists may look back on the current indifference to this trend with the same kind of horror that we now feel when we read how the Federal Reserve System calmly went about its business as while the money supply plummeted between 1930 and 1933.

In a discussion of macroeconomic policy, it is even more important to emphasize the other element in a policy that supports progress. Governments must not impede change. Change is disruptive and the citizens of industrial democracies increasingly demand that governments protect them from disruption. From a long-run perspective, the most serious side effect of the Keynesian revolution in macroeconomic policy may have been the intellectual legitimacy that it lent to the impulse to blame the government for all instances of economic misfortune. After all, if the government controls powerful levers that can be used to prevent job loss and unemployment, someone who loses a job would seem to have a strong claim of negligence against a government fails to act.

In economics, the real force of the athletic slogan of "no pain, no gain" lies not in an admonition to consume less and save more. Rather, it comes from the warning that in many cases, things have to get worse before they can get better. If you talk to a swimmer, she will

tell you that each time she tries a new variation in her technique, her performance falls. She typically has perfected the details of the old, less efficient stroke, much as the English swimmers had perfected the inherently inefficient breast stroke. When she makes a change, she initially gets many of the details wrong and performance falls. Then slowly the efficiency improves as she adjusts all the details to get the best performance out of the new the stroke. If swimmers were never willing to tolerate temporary reductions in performance, they would not be able to experiment and develop better technique.

The equivalent process of experimentation, deterioration, and improvement that takes place in an economy was called "creative destruction" by Joseph Schumpeter. A new or improved product typically replaces an existing one. The creation of economic returns for inputs used to make the new good is associated with the destruction of returns for inputs that made the old good. Resources used to produce the old good have to be shifted into some new activity. At best this results in a spell of unemployment. At worst, it can lead to a permanent loss in income.

If the old activity had been a particularly profitable one (the production of mainframe computers, for example), stockholders of the company engaged in this activity will lose when the new good (the personal computer) comes along. If the company that made the mainframes shared some of its profits with its workers, paying them well above the market wage, these workers may also suffer permanent reductions in income when the new personal computer industry destroys the profits of the mainframe manufacturers. The individual workers who lose their jobs may never find new jobs that pay them more than the market wage. At the same time, other workers at some of the new firms in the personal computer industry (for example, those who produce the microprocessors that power the personal computers) may earn above market wages as their firms expand and earn high profits. In addition, consumers and who use computers get the benefits of less expensive computing power.

Innovation and change creates winners and losers. Job losses and pay cuts are the most visible, and politically most powerful symptom of this disruption. But we know that for society as a whole, innovation, discovery and technological change offers large net gains because the new goods or processes are more efficient and more valuable than the old ones. A world with personal computers opens up opportunities that were inconceivable in the era of the mainframe. If we had not tolerated disruption in the past, we would still travel in ox carts.

If the firm that made the mainframes was also the firm that made the microprocessors, it might be possible to shift workers internally from mainframe assembly to microprocessor fabrication, without letting any workers go and without causing an intervening period of formal unemployment as they search for new jobs in the labor market. But in North America, many new innovative activities arise in new start-up firms. The job market is therefore characterized by constant turnover, with new hiring at some firms and job losses at others. Much of this turnover takes place during an economic expansion. The rest of it gets concentrated during recessions, when the vulnerability of older firms becomes plainly evident.<sup>9</sup>

Macroeconomic policy measures like fiscal or monetary stimulus cannot change the fact that many older firms need to shrink. Aggressive stimulus measures can perhaps change the timing of the job losses, but they cannot avoid them. If they delay the process whereby an existing firm comes to terms with new competitive realities, they might make things worse. More aggressive measures like bailouts, government loans, nationalization of failing firms, and prohibitions on firing workers are even more likely to have perverse effects in the long run. The ultimate contraction is likely to be associated with more job loss and loss of income than it would if firms had responded sooner.

This perspective suggests that much of the effort expended by governments to prevent recessions and avoid job losses may be misguided. The circumstances of the 1930s may have been exceptional (in part because of government actions that made the contraction so severe).

Stimulus measures that were appropriate then may not be useful in the ordinary course of events. Such efforts may only postpone processes of adjustment that are inevitable. There is a growing recognition, for example, that high levels of long-term unemployment in Europe may be the unintended consequence of policies designed to fight unemployment and recessions, not the result of the recessions themselves. A more productive approach to policy would then be to make the adjustment process as efficient and painless as possible, and to maintain the conditions that lead to rapid entry of new firms that compete for workers.

The greatest challenge for policy makers in may therefore be one of political leadership. Each year, new demands for security and protection from disruption are voiced. Meanwhile, evidence mounts that government efforts designed to provide security and prevent disruption have had a corrosive effect on the operation of markets, on the processes of entry and exit by firms, and ultimately on the kind of competitive spirit that any successful competitor — in the world of athletics or the world of work — must cultivate. This spirit is important in achieving a goal, but more importantly, may also lie at the center of a life that is well lived. Many workers suffer losses and setbacks during their careers, yet still have access to material advantages that would have been unattainable by even the very rich just two generations ago. Most swimmers loose races, but still benefit from participation in the competition.

To succeed in narrow economic terms, but also in broader human terms, leaders like coaches, must give people the confidence to compete. They must encourage people to believe that economic change brings real opportunity together with risks that are real, but manageable. Leaders must not cater to the demands of a generation that seems to believe that each person is entitled to a job, a house in the suburbs, and two cars just for serving time in school — as if everyone deserved a trophy just for putting in a specified number of hours in the pool. They must inspire people to dive in, strive for each small improvement in technique, compete to the best of their abilities.

Nations that can sustain this kind of policy stance can count on sustained economic growth that will carry them into the next century. Those that are most successful in creating institutions that foster discovery and innovation will be the worldwide technological leaders. Through mechanisms like free trade and transfers of technology by multinational firms, nations that are less successful in the cultivation and commercial exploitation of science and technology can still follow comfortably along in the wake of the leaders. But nations that try to resist change by protecting inefficient firms, impeding flows of goods and ideas, and making a high level of income an entitlement instead of a reward will slowly be left farther and farther behind.

## Endnotes

1 For example, Smith also emphasized the importance of free trade. He argued that larger markets would encourage technological progress through increases in specialization and the division of labor. Until recently, this important connection between trade and growth was neglected by economists. See Grossman and Helpman [1992] for the leading account of renewed work on this topic.

2 The conventional wisdom used to be that recessions in the post-war years were less severe and less frequent than during a historical comparison period that does not include the 1930s. Recent work by C. Romer [1986] has demonstrated that most of the apparent evidence for post-war stabilization came from a misinterpretation of the historical statistics on income growth.

- 3 Romer [1990] and Grossman and Helpman [1992] give technical presentations of this kind of theory. See Romer [1993a] for a non-technical discussion.
- 4 The description of training methods is drawn from Maglischo, 1993, and Colwin, 1992.
- 5 The extent of the fall in the money supply was first documented by Milton Friedman and Anna J. Schwartz [1963].
- 6 See Colwin [1992, p. 208.]
- 7 See Sprawson [1992] for an account.
- 8 For a discussion of the range of issues that an analysis of economic growth raises for conduct of science and technology policy, see Romer [1993b].
- 9 Recent work by Baldwin and Gorecki [1990] (for Canada) and Davis and Haltiwanger. [1990] (for the United States) has demonstrated how important entry and exit by firms is in causing job turnover in good times and bad.

## References

- Baldwin, John R. and Paul K. Gorecki, Structural Change and the Adjustment Process, Ottawa: Economic Council of Canada, 1990.
- Colwin, Cecil M., Swimming into the 21'st Century, Champaign, IL: Human Kinetics Publishers, 1992.
- Friedman, Milton and Anna J. Schwartz, A Monetary History of the United States, Princeton: Princeton University Press for the National Bureau of Economic Research, 1963.
- Grossman, Gene and Elhanan Helpman, Innovation and Growth in the Global Economy, Cambridge, MIT Press, 1992.
- Haltiwanger, John and Steve Davis, Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications, NBER Macroeconomic Annual, 1990, 123-186.
- Maglischo, Ernst W., Swimming Even Faster, Mountain View CA: Mayfield Publishing Company, 1993.
- Romer, Christina D., Is the Stabilization of the Postwar Economy a Figment of the Data?, American Economic Review 76 (3) (June), 1986, 314-334.
- Romer, Paul M., Endogenous Technological Change, Journal of Political Economy 98 (October), 1990, S71-102.
- \_\_\_\_\_, Economic Growth, in The Fortune Encyclopedia of Economics, David R. Henderson ed., New York: Time-Warner Books, 1993a.
- \_\_\_\_\_, Implementing a National Technology Strategy with Self-Organizing Industry Investment Boards, Brookings Papers on Economic Activity: Microeconomics 2, 1993b, 345-399.
- Sprawson, Charles, Haunts of the Black Masseur: The Swimmer as Hero, London: Jonathan Cape, 1992.

# Social Stability and Economic Growth

---

*Kei Takeuchi*

## 1. Natural and Social Sustainability

Recently, much has been made of "sustainable development." But there seems to be confusion and misunderstandings about the term. The purpose of this paper is to clarify some aspect of the conditions of sustainability.

First, we shall discuss some general framework. Sustainable development means a course of social or economic or whatever, development which can be continued indefinitely. And in order to enable a course of development sustained, outside or exogenous conditions, and inside or endogenous conditions must be satisfied. If we look at the humanity or human civilization as a whole, natural conditions, that is, availability of natural resources and environment, constitute the exogenous conditions, and social conditions, that is, capability of social systems determine the endogenous conditions for development of human society. And sustained development is possible when both types of conditions can be maintained. Therefore, we may classify conditions of sustainability into two types accordingly to the above as those for natural sustainability and those for social sustainability.

In the general discussion of global environment problems, attention seems to have been focussed on natural sustainability, which is understood to be the central issue of the global environment problems. But I would rather like to emphasize that two types of conditions are closely interrelated, and that social sustainability is equally, if not more important.

The first point which I would like to remark is that sustainability does not imply constancy. If sustainability has to imply fixed, unchanged conditions, "sustainable development" is an obvious contradiction, since "development" cannot be anything but change. I would also like to point out that natural conditions on the earth has always been undergoing major changes since its birth some billion years ago, and still is changing incessantly without any human intervention. Human civilizations, in the course of thousands of years of history, quite instantaneous in the history of the earth but long enough in human standards, have interacted with natural conditions and brought about substantial changes on the surface of the earth. Although recent progresses of technologies have made the power of human interventions in the natural process of the earth so much greater, it is still much below the level of natural potentiality of the earth.

It is sometimes said that we have to "be kind to the earth," which if straight forwardly understood, is not only meaningless, since the human civilization can service only within very limited conditions of the earth among the vast varieties of possible conditions of the life of the earth, but also absurdly arrogant since the earth may well do without humanity and would not care if the human civilization did vanish altogether from the surface of the earth. More proper expression should be "to let the earth keep conditions kind and congenial to the people", which is the real implication of natural sustainability.

Although the surface of the earth is always in dynamic process of changes either due to "physical" forces coming from within and from outside, and due to global ecological systems, and also due to human actions, there seem to be states of relatively stable equilibrium and of rapidly changing transiency. Natural sustainability means remaining in the state of relative stability, the adjective "relative" being essential since any equilibrium would be destroyed sooner or later by the "natural" forces. Also it is to be emphasized that there are

multitudes of relatively stable equilibria, some hostile, some congenial to humanity, and that the state of transiency also cannot continue indefinitely but a new state of equilibrium will be reached rather soon in any case. Therefore, the problem of natural sustainability means choosing among the possible states of relatively stable equilibrium, which is "best" to humanity, and attaining and maintaining it. It is a gross misunderstanding to assume that there is only one unique solution to this problem, and to assume that the present condition or those prevailing pre-industrial age are anywhere near to the "best". The CO<sub>2</sub> warming issue touches the point; although it is quite natural to be concerned with the possible damages which might be by global warming due to increased emission of the CO<sub>2</sub> gas, it is another to assume that the present level of the temperature, with its "natural" fluctuations, is the best possible case we can imagine, and that it is always harmful to affect the global temperature level even when it is possible to change it intentionally.

## 2. Concept Of Social Sustainability

The above are preliminary observations for our main theme of discussion, social sustainability.

The most important point is that the conditions of natural sustainability are made sense only in terms of human or social sustainability, and natural and social sustainability are inseparably interrelated.

Some "environmentalists" especially such people as the "deep" ecologists may raise the issue that there is a fundamental distinction between "natural" and "social" sustainability, or within "natural" sustainability, biological or "ecological" sustainability apart from the "physical" sustainability of the earth surface, and may argue that sustainability of ecological systems is independent of and somewhat above human or social sustainability.

I don't think that ecological systems as such have intrinsic values independent of and above all human values, (of course here I don't mean solely materialistic or commercial values but also include aesthetical or even ethical values which still cannot transcend human frameworks). Also I would emphasize that ecological systems on the earth have been always in a state of flux and substantially transformed by human activities at least since the advent of agricultural civilization. Ecological system is a part of the "nature" in which we ourselves are active participants.

In all discussions of sustainable growth, conditions of social sustainability are assumed at least implicitly, because natural sustainability really means maintaining natural conditions which enable social development sustainable. When we are not quite articulate about the conditions of social sustainability, whole discussion of natural sustainability becomes ambiguous and may lose the point.

Let us consider, as an example, rather "pessimistic" predictions made by the Club of Rome, (and also later versions by Dennis Meadows and others). They showed that if the present trend of economic growth were to be continued for another half century, the natural resources would be exhausted, natural environment irretrievably destroyed, and the size of population and the level of economic welfare must make sharp downward turn. And this means, not explicitly described in their quantitative analysis but well implicated, that there would be famine, epidemics, wars and social disturbances, and large scale retrogression, if not total disintegration, of the human civilization. Everyone would regard such a course as really a colossal disaster, still one might ask so what? Even though it seems to be an unimaginable tragedy, such cases actually have happened repeatedly to the past civilizations, and though the progress of human civilizations was often arrested and sometimes retarded, they somehow survived and revived after some years and new civilizations emerged and prospered.

So, even when the 21st century is going to see the global disintegration of the human civilization we could well assume that it would not be the end of the human race, nor even that of human civilization, and after some period of terrible confusion, new balance between natural conditions and human activities would be established at a much lower level than before and gradually a new form of civilization will emerge and then start to prosper.

Such prospects are surely distressing to us. But anyway it belongs to the history of the future, which none of us probably will ever live to see, so why do we need bother?

It is the basic assumption of such a conference as ours that, we do bother and worry – hence we have to consider the conditions of social sustainability into more detail and more precisely.

### **3. Levels Of Social Sustainability**

All discussions of conditions of social sustainability, or on the problems of sustainable growth, presume value premises which hold that sustaining human civilization is paramount.

But we have to distinguish different levels of conditions of social or human sustainability. We may distinguish three major dimensions of social sustainability, bio-physiological, socio-psychological and ethical.

Human beings must be fed, clothed and sheltered, and also given minimum level of health care in order to survive. To provide the whole population with necessities of subsistence is the minimum condition of social sustainability. Therefore as long as the world population continues to grow, provisions for physiological subsistence must grow proportionately. It is now mostly agreed that the world population would grow at least to the level of ten billion or twice as large as the present size, which almost automatically means that the world production of basic necessary goods must double within a proper time limit. And although it must be admitted that there are over-consumption and wasteful use of necessary goods in many parts of the world, it is also obvious that there are still many millions of people without even minimum necessities of subsistence, hence it would be safe to assume that economic growth must be proportionate to population growth even for the most basic level of social sustainability.

There is a subtle question about population growth. Plainly set, the question is whether population growth is the result or the cause of the level of economic welfare.

It is now generally agreed that population growth should be treated as an independent or exogenous variable when economic growth is considered.

But it does not mean that population growth is entirely autonomous and out of human control. Indeed population control is the most important issue in relation to the problem of sustainable growth, on which I would discuss later, and here would like only to mention the fact that lowering of the birth rate can be brought about only when the average per capita income reaches some level much higher than that of mere subsistence except for the case of such calamities as famine, epidemic, and war. This implies that some economic growth is absolutely necessary to attain socially sustainable conditions.

Such argument leads to the conclusion that for social sustainability, mere subsistence is not sufficient but some minimum level of cultural life must be guaranteed for everybody. Although it is difficult to draw a demarcation between cultural necessity and luxuries, but cultural aspect are indispensable to the life worthy of human beings and is also necessary in order to maintain valuable labour force for social functions. Also health belongs to this level of social sustainability conditions, and is its most important ingredient. Health means something more than simple physical adaptation to the environment, and has human or cultural aspects.

Education is a vital part of the cultural aspect of the social life and providing every one with some level of education is an essential part of social sustainability condition.

#### 4. Sustainability Of Socio-Economic Systems

The second level of conditions of social sustainability is related to the socio-economic system. Orderly social development is possible only when the socio-economic system can function.

The conditions under which a socio-economic system can survive and function depend on the nature of the system. Now most of the advanced countries in the world have liberal democratic political systems and capitalist market economic systems. Under democracy political systems can survive only when desire and aspiration of the people are mostly satisfied or at least not much frustrated. It means that for a democratic society to develop, it must not only provide the people with basic needs but also keep them socio-psychologically satisfied. This means that people must be assured of a standard of living which they think they deserve, and of possibility of betterment of life through their own effort, also of satisfaction of minimal level of social justice is important.

In abstract terms one can argue that living standards of the most advanced countries is far above minimum necessity and needlessly luxurious. Levels of consumption could be halved without rendering real economic welfare much lower, people could forgo the happily wasteful way of "modern" living. But in any country, the government could not remain in power if the peoples income did substantially fall, much less if the government deliberately manipulated it. Hence even the richest countries, keeping the level of per capita income more less at the present level is of the absolute political necessity.

As was impressively stated by Th. Veblen among others, people's aspiration for economic consumption is not determined solely by the absolute terms, but to a great extent by relative terms in comparison with that of their neighbours, other classes, other nations. And the more information technology progresses, the more people are influenced by what is going on in all parts of the world, and people tend to want the more strongly to obtain whatever they think other people are enjoying. Now every nation in the world thinks that she is entitled to the "fruits of the highest civilization". Unless all the people in the world can suppose that the way to the highest economic welfare is at least open to them in principle, and the goal is achievable if in the far future, the world wide political order and peace cannot be maintained.

On national levels, no government wants its country to remain permanently in the group of "backward" or underdeveloped countries, and wants to join the group of advanced countries through modernization by industrialization and rapid economic growth.

Also in most advanced countries, a minimum level of growth of economy is needed for smooth functioning of the economic system. Otherwise the economy will fall into "depression" and unemployment will increase with all social maladies associated with it. Such minimum growth level is not exactly specified numerically but it is certainly positive, and it is independent of the fact whether the society really needs more goods for consumption or construction. What matters is the "effective demand" but not the "social needs" to growth, whatever its content may be. In this context wastefulness to some, and often very great, extent is of often social necessity, a fact not necessarily limited to modern, affluent societies.

The important factor here is the international competition. Advanced nations want to get the "hegemony" over other nations, economically, politically, or militarily, all based basically on economic power. Thus it is most difficult to persuade any country to restrain from economic growth if there is a factor of international competition.

International economic competition of conditions is a reality which we cannot disregard in consideration of social sustainability.

## 5. Ideological Aspects

When we consider "limits of growth" imposed by the natural sustainability conditions, a paradoxical aspect of the social sustainability is that, a society with a democratic political system and free market economy has stronger dynamism and vitality than other societies, hence greater potential for sustained development, has also greater necessity of economic growth. In a free society people are allowed to express their desire and exert their power to achieve their own goals. Consequently their capabilities can be more aptly realized, but once they feel that they are denied the opportunity or frustrated in their effort, the society may easily fall into trouble.

Various types of pre-modern societies, the Japanese feudal system in Edo era for example, placed stabilization as its basic principle, hence their tendency to grow was much lower, but we cannot regard such societies having higher degree of social sustainability than the modern society. Their stability meant rigidity and inflexibility, which meant in turn lower capability to cope with changing circumstances and higher vulnerability to fluctuations, as was exemplified by repeated famines due to climatic changes. Also even in apparently most static societies, changes and growth are never completely ruled out, and in due course of time, they may suddenly collapse because of the accumulated frictions. Such societies were not always harmonious with natural conditions for sustainability. There are many cases when pre-modern civilizations irretrievably destroyed their natural resource-environment foundations, by over exploitation of soils, by over irrigation, by destruction of forests etc.

The last level of social sustainability condition is that of philosophy and ideology. Every civilized society has its own value systems and concepts of justice, and the society must prove itself worthy of its members' effort to sustain by achieving its professed objectives and realizing justice. Sustainable society must be the society worth sustaining.

In the contemporary world, most people agree, if only as principles, on the basic values of the modern society such as "liberty, equality and fraternity" and also "respect of individual human rights", and they are also our basic value premises in our discussion of "sustained growth". The world order worth being sustained must satisfy minimum standards of social justice and equity among all the peoples in the world.

From this viewpoint, it is obvious that there are much gross injustices in the present world, unjustifiable inequality, intolerable violation of human rights, and there are many aspects in the present social systems which must be transformed.

Proponents of the sustainable development must not advocate status quo of the present social order. Therefore, maintaining prevailing social or political order should not be included among the conditions of social sustainability.

Although maintaining social order is an essential part of social sustainability, it does not mean that whatever social order is worth sustaining. Thus the discussion of sustainable development inevitably involves ideological problems about in what direction the society should develop.

If we assume natural sustainability is absolute, the solution for it could be the worldwide totalitarian dictatorship, suppressing all human activities which may disrupt natural environment on the earth. But for us, such a "solution" seems to be as abominable as the disaster caused by the destruction of natural foundations by development of human civilization.

We need to note one special aspect of the value judgement on the sustainability questions. Sustained development is intergenerational. It means a process stretching over the periods of several generations, where conditions and behaviour of the present generation affects the future, yet unborn generations. To keep sustainability conditions means to provide the future generation with the foundation of the theatre of their activities. Here the interest of the present and the future generations must be balanced and harmonized. It is evident that such har-

monization cannot be achieved through negotiations and compromises among the parties of "enlightened individualistic egoists", since no one can represent the future generations in such negotiations. We need imagination and feeling for the unity of human civilization of past, present and future when we discuss sustainable development of the human civilization. Also we need some basic philosophy which transcends simple rational liberalism, which holds the absolute right of the individual.

## 6. Dynamism Of The Problem

It may be said to be the most basic feature of any civilization ever to develop and to expand, and the modern civilization is by far the most dynamic. Therefore it would be altogether impossible to contain it in a stable state prohibiting any further development. Any consideration of social sustainability rules out the immediate arrest of world economic growth as advocated by the Club of Rome.

But it also must be emphasized that as long as the human civilization is constrained on this earth, that is in the foreseeable future, some how natural conditions of the surface of the earth must be stabilized in the long run. It means that the energy flow and material circulation on the earth be stabilized. It is certain that if one directional change of energy flow or material circulation cannot continue indefinitely without causing serious threat to human civilization, and there are symptoms that disasters are not far away, although there are uncertainties exactly how far or how near.

It means that the present trend of economic growth must somehow change its direction and speed.

If we consider both the natural and social condition of sustainability together the most important is the dynamic nature of the problem. Not only various aspects of social and natural conditions change dynamically often in quite unpredictable ways, they interact in complex ways. Also we have to take into consideration the most important factor of technology.

Looking back into the past 100 years of technological progress, we can or must suppose that there shall be tremendous technological progresses in the forthcoming century. Without being over-optimistic we have every reason to expect that technological progress will be accelerated. Therefore it would be quite difficult to predict what will be the levels and forms of the technologies available to humanity at the end of the next century.

Still one cannot blindly rely on the future progress of technologies as a kind of deus ex machina, which can immediately resolve any difficulties. Progress of technologies is never uniform in different sectors and in different times and it does not necessarily respond to human needs or human wishes concerned with the contemporary people. Also, in some cases technological innovations are quite slow to be realized in actual social frameworks especially application is not limited to a specialized field of human activities such as military, industrial or even academic, but is involved in complicated social interactions (e.g. city planning).

Therefore we should be neither too optimistic nor too pessimistic about our technological capabilities for sustained development. We have to be careful to make any future predictions based on extrapolations of the trends of technological development.

Therefore, the most important factor in our discussion of sustainable development is the dimension of time scale. Sometimes it is emphasized that global environment problems require a very long time scope, say a few hundred years; hence we have to consolidate a plan of action for such a long period. But we have to admit frankly that there are too many and too great uncertainties to make up any practical plan for such a long period. With any confidence we can only foresee among 30 years or 50 years at most and make a program for action accordingly.

And it is quite impossible to stabilize the conditions of the world during the period of several decades if we seriously consider social sustainability conditions.

We are in a transient state, and what we have to try is not to attain socially sustainable situations once and for all, but to make the natural and social sustainable conditions more compatible to each other.

There is also an important factor of historical change of social systems. Modern world history is a continuous process of changes in social systems, some sudden, some gradual, but in all the changes have been tremendous and quite unexpected. The 20th century saw revolutions, wars, rise and fall of different regimes, transformation of social structures, social norms, life-styles, etc. Now social systems are often shorter lived than individuals. Who anticipated the sudden collapse of the former USSR socialist regime even five years ago?

And such changes also transform the framework of social sustainability. One of the worst features and the main causes of the fall of the former Soviet social system was its neglect of environment conservation. Its fall could improve the natural sustainability conditions in this part of the world. But the capitalistic system, as such, is not well qualified for compatibility of natural and social sustainability conditions with its strong drive for growth and expansion.

If the 21st century is also going to be a period of change, it is not necessary for us to assume that the present social systems will continue to prevail, nor is it desirable to maintain them. But then the issue becomes the basic problem of socio-political ideology about the choice of the best socio-political systems or the transformation of the present system for the better. Then the discussion will extend to well beyond the general framework of sustained growth, which is another strong reason why the scope of our discussion of sustainability is limited in time horizon and scope for social change.

## **7. Problem Of Human Reproduction**

One subtle aspect of social sustainability is concerned with human reproduction. Abstractly speaking, a social system is sustainable only if the size and structure of its membership is kept more or less constant, that is, its members are properly replaced by new generation.

Animal species are sustained both by their own system of reproduction and by the logic of the ecological system. All primitive societies and also pre-modern societies have maintained systems of human reproduction under duress of natural environment, and could sustain themselves for some time.

It is a basic contradiction of modern industrial society that it has no inherent logic of human reproduction to make it sustainable. On the one hand, it has pushed the ecological pressure aside by improving medicine and hygiene and also living conditions in general, but also on the other hand, it has abolished traditional rules of social behaviour concerned with human reproduction without replacing them by its own. Thus it destroyed both outside and inside systems to regulate the process of human reproduction, and made demographic changes quite unpredictable.

Malthus thought the only ecological pressure could check over growth of population, and Ricardo applied the rule of market economy to the population as the labour force and held that long range equilibrium between demand and supply of human labour could be achieved.

Both theories have been disproved by the actual course of the development of modern industrial societies. Technological progress outpaced the population growth, and level of economic welfare increased together with population, Malthusian ecological check did not work. But when economic living standard reached some level, the birth rate made a downward turn, the population growth, slowed down, then stopped, and now in almost all advanced countries, is going to be negative.

There are various *ex post* explanations given for this change in the trend of birth, but none are sufficiently powerful to be predictive.

Now in the contemporary world we are faced with a very complicated picture of demographic trends. In the least developed countries where social systems still keep mostly traditional rules of behaviour, the population grows very rapidly, producing over population problems, since the ecological check was diminished by the introduction of even minimum of scientific medicine and hygiene, while in rapidly developing medium-developed countries population increase and economic growth are accelerating each other as in the early stage of development of modern societies, and in well advanced highly developed countries birth rates are too low to replace the previous generation turning the long term population growth rate into negative.

Some nations including Japan moved through these three stages in a rather short period of less than 50 years. And it is now generally expected, sometimes wishfully, that the whole world will experience similar changes of demographic trends, and the world population will grow to some point, then stop and then gradually decline.

The fundamental weakness of such a prediction is that there is no coherent theory behind it. But that is not the demographers' or social scientists' fault. Modern industrial society pushed the process of human reproduction outside the regular social processes and made it unpredictable.

There is also a moral issue involved. The basic question is who has the right to determine whether a new baby is to be born or not to be born? If the liberal individualist moral principles are pushed to the extreme, it must be the basic right of the person him/herself to decide whether to be born or not to be born, an obviously ridiculous conclusion. But is it really obvious that the right to decide whether a new baby shall or shall not be born belongs solely and exclusively to its mother? Or the father has any right to be consulted? How about other members of the family? In any case it is usually assumed in the modern society that having a child belongs to a completely private sphere of the family life, hence there are strong objections against any social intervention into the matter.

The policy measures for population control, that is birth control, are quite limited by not technological but ethical considerations. Recent Chinese "Single Child Policy" will deserve moral criticism to the effect that it violates basic human rights. But it is almost obvious that without such strong and perhaps dubious measures, efforts to depress the birth rate could not have been successful.

There are also religious aspects. The Roman Catholic Church still maintains its stance against all methods of contraception and abortion and makes any birth control policy measures almost impossible. But then does the principle of religions freedom dictate that the Catholics are exempted from any birth control policy?

In reality, human reproduction cannot remain completely as a private matter. We have to find somehow to accommodate population control measures within the framework of modern society in order that its social sustainability is maintained. And here we again need some philosophical or ethical principles to guide the policies.

## 8. Conclusion

Economic growth should be regarded as a means to achieve social sustainability or to bring about a desirable social system all over the world and to perpetuate it. Economic growth as such cannot be considered to be an independent goal, but it is indispensable for social sustainability of the world system. It would not be easy to find a course of social development which makes social and natural sustainability conditions compatible, but it is an unavoidable task if we hope that human civilization continues to prosper.

# Discussion of Papers by Paul Romer and Kei Takeuchi: Economic Growth in Today's Global Economy - Implications

---

*Peter J. Nicholson*

Space permits only a brief and highly restricted commentary on the fascinating papers of Professors Romer and Takeuchi. My remarks will be framed in response to two rhetorical questions: (1) In response to Paul Romer – do deficits really matter? and (2) In response to Kei Takeuchi – is “sustainable development” an oxymoron?

## Do Deficits Matter?

**T**his question has been prominent in macroeconomic policy thinking since the time when Lord Keynes' ideas became the standard nostrum to keep capitalist economies performing at full throttle. Paul Romer puts the debate in a proper modern perspective concluding that “the medium term effect of reduction in the budget deficit would most likely be to increase the national savings rate and increase the level of investment.” Deficit reduction would thus encourage at least modest growth in the medium term. He concludes by saying that we should shift the focus of the debate away from questions of pump-priming through deficit spending and toward finding better *techniques* of production. This is a key message of the *new* economic growth theory and is well-taken.

But in the meantime, we cannot avoid facing the old, and now pressing question as to whether deficits do, or do not matter. Posed in the abstract, the answer to this question (like virtually all questions in economics) is: “that depends.” It depends on the magnitude of existing public debt relative to the size of the economy. It depends on whether or not the deficits were incurred primarily to finance productive investment which can be expected to generate sufficient extra growth in future to more than repay the investment. And, as a practical matter, it depends on whether the deficits are part of a *pattern* of chronic excess expenditure, both in bad times and good.

In most industrialized countries, governments have slipped into a pattern of chronic deficit. For example, the government sector in Canada last had a surplus in 1974. So much then for the Keynesian notion that governments should engineer budget surpluses when the economy is expanding and incur deficits only to help offset a contraction.

The effect in Canada of unremitting deficits for the past 20 years has been inexorably to increase the total stock of public debt relative to the size of the economy – for example, since 1980 the aggregate debt of the federal and provincial governments combined has increased from about 35% of GDP to somewhat more than 90% today, and rising.

But might this deficit spending nevertheless be justified as an investment in productive human and physical capital that will pay big dividends through much better future economic performance? The evidence, unfortunately, is all the other way. For example, government capital spending in Canada, expressed as a percentage of total expenditure, has been on a downtrend since the 1960s and is today about 4.5% as compared with more than 10% in the early 1970s.

Public spending growth has been due, almost entirely, to transfer payments of various kinds and, particularly during the past decade, to interest payments on surging public debt – reflecting the accumulation of perpetual annual deficits. Interest on the federal government's debt is running at about \$40 billion annually, or almost 25% of total expenditure, up from about 10% in 1974, just before the fiscal wheels in Canada began to wobble. Interest on existing debt is now the principal generator of continued deficits and thus of expanding public debt. The conditions for exponential growth of indebtedness have thus been established.

I believe, therefore, that we can say unequivocally that the string of large government deficits in Canada does matter a great deal. Apart from the fact that worthwhile public services are being inexorably crowded out by interest payments, the deficits matter *economically* for two fundamental reasons.

First, deficits have forced a rapid increase in taxation, and the *expectation* of further taxes, to the point where investment has been discouraged; business competitiveness impaired; consumer disposable income eroded; and the traditional voluntary compliance of tax filers undermined. To appreciate why Canadians have reached the limit of their tolerance for taxation, consider that between 1988 and 1992, total personal tax payments were boosted 31%, while the economy as a whole grew by less than 14% in nominal terms.

The second deleterious effect of chronic deficits is the growing diversion of the nation's savings to finance government activity – much of which is redistributive and not wealth producing – at the expense of private sector investment. In Canada, domestic savings have been inadequate to satisfy both the voracious appetite of the public sector and the investment needs of business. Canada has therefore had to borrow huge amounts abroad – now about \$25 billion additional every year – and has consequently accumulated an external debt of about \$325 billion or 45% of GDP. This is by far the largest relative foreign indebtedness among the G7 countries.

The effect of this extraordinary demand on capital markets has been to make Canadian interest rates higher than they otherwise would be – thus inhibiting growth-promoting investment. And ultimately, there is a limit to any nation's ability to rapidly run up its tab with foreign creditors. Unusually bullish conditions in global bond markets have made it relatively easy for Canadian governments and corporations to borrow abroad, at least for the time being, but the progressive downgrading of the credit ratings of most provincial governments is a signal that public sector borrowers in Canada may be nearing the end of their rope.

### Is Sustainable Development an Oxymoron?

Professor Takeuchi begins with the proposition that capitalist economies have the propensity, or even the *necessity*, to grow continuously in per capita terms. Thanks to television and other agencies in the business of "want creation," the material aspirations of consumers ratchet up relentlessly. A potential contradiction thus emerges between the *social-political* conditions of sustainability in capitalist democracies and the external *environmental* conditions of sustainability in the face of an apparently endless assault on the biosphere. This is Professor Takeuchi's dilemma.

It is obviously true that the consumption of terrestrial matter and energy cannot go on indefinitely. But this logical truism is almost beside the point since eventually the sun will burn out or some other astrophysical catastrophe may end the human experiment in the meantime. The literal finitude of the earth's potential material and energy resources is therefore not really at issue within the time horizons of existing human institutions. The relevant question is the *rate* at which net consumption (after recycling) takes place and the *rate* at which biologically harmful pollutants accumulate. If these rates are "sufficiently moderate" relative to the earth's stock of resources and bio-resilience, the question of material limits to growth

becomes moot. The politically crucial question, therefore, is whether through technological innovation, assisted by broadly-based attitude changes, it will be possible to reduce the rate of material/energy consumption sufficiently to satisfy the growth imperative of democratic societies at a cost the biosphere can sustain.

The dramatic rise in the prevalence of "green" concerns since the early 1970s indicates that increasing numbers of people fear that the answer is "no." More optimistically, one could interpret the emergence of a global environmental consciousness – still far from universal of course – as evidence that the required attitude changes are already underway. But obviously, there are no guarantees and the consequences of being wrong are apocalyptic.

Nevertheless, it is important to sketch an optimistic scenario, lest too many people should give up the ghost and stop using their ingenuity to resolve our common dilemma. Let us recall some salient facts and trends.

- The amount of energy and material resources consumed to produce a unit of global economic output has been declining for more than a century. Of course, the *total* consumption of energy and other resources has increased enormously as average material well-being has advanced. But at some point, the declining *material intensity* of the world economy might completely offset the impact of growing activity levels.
- Between 1967 and 1988, the physical weight of U.S. merchandise exports per dollar (after deducting inflation) declined 43%.
- Between 1965 and 1985, Japan increased its total volume of production two and one half times with virtually no increase in the amount of energy and materials consumed.
- The effect of relative price changes on the incentive to conserve resources has often been greatly underestimated. For example, dire predictions of the imminent physical exhaustion of many materials – as documented in the 1972 publication, *Limits to Growth* – were altered dramatically, if not falsified outright, by the oil price shocks in 1973 and 1979. The price system has provided a remarkably potent set of incentives to develop new products and processes that dramatically reduce consumption of what is scarce and/or expensive. Necessity is still the mother of a great deal of invention.
- Growth in the advanced economies has shifted very rapidly (by historical standards) from the production of tangible goods to the production of services. Now, about 70% or more of the GDP of countries like Canada and the United States is generated in the service sector. More profoundly, *information* is replacing *energy* as the key transforming resource in the economy and *intellectual* capital is replacing brute *physical* capital as the central agent of productivity improvement. In fact, it is increasingly the case that even tangible goods are really the embodiment of services such as R&D, design, engineering, advertising and financing. Today's automobile, created with the aid of robots, advanced materials, sophisticated engineering and lean production techniques, is mostly knowledge on four wheels.
- Consider finally the paradigm products of the new economy – (1) Only 2% of the value of a microchip attaches to the materials and energy of its production; (2) the material component of modern pharmaceuticals is even less. Today's drugs are embodied R&D, marketing and finance; (3) a fiberoptic cable can now carry 1.5 million transatlantic phone conversations as compared with the 140 conversations that were transmitted by the underwater copper cable in 1965; and (4) computer software, the quintessential representation of economic value in the information economy – and the means by which Microsoft brought IBM to its knees – has literally zero material content. Even the media in which software is stored have been shrinking inexorably to the scale of atoms.

These trends should at least create a glimmer of hope that the dilemma identified by Professor Takeuchi may be resolved through human ingenuity before it is too late. The answer depends principally on the nature of the development process underway in what used

to be called the Third World. There are at least 4 billion people in these nations – and their absolute numbers are still increasing rapidly – who have a very long way to go before achieving anything approaching the levels of resource consumption needed for a decent standard of living by industrial country standards. It is an open question – both politically and technologically – whether the reasonable aspirations of developing countries can be achieved before there is catastrophic insult to the biosphere. If able to surmount this hump – and it is a big if – I believe that the accelerating trend of “de-materialization” in the mature, advanced economies, affords a vision of sustainable growth that until recently has seemed unattainable.

# Discussion of Papers by Paul Romer and Kei Takeuchi: Economic Growth in Today's Global Economy

---

*Ken Delaney*

**T**he papers submitted by Prof. Takeuchi and Prof. Romer were quite thoughtful and certainly consistent with the high quality that is characteristic of work done for The Canadian Institute for Advanced Research. To their scholarly work, I can only add the perspective of a practitioner. Thus, as I am a representative of the United Steelworkers of America, it is a worker's perspective that I bring to the two papers.

Prof. Takeuchi warns that when we set sustainable growth as a policy objective, we must not forget social and political sustainability in our zeal to protect the environment. He also argues that social sustainability is not possible without some level of economic growth.

Prof. Romer argues that policy makers typically focus on either short term economic stimulus or the accumulation of physical capital as the primary economic policy objective and that something much more important, the role of innovation and technological change in sustaining economic growth, is too often ignored. More specifically, Prof. Romer argues that sustained economic growth is dependant on improving our "techniques" and that therefore we must develop policies that encourage innovation.

I certainly agree that focusing on social and political sustainability are critical economic policy objectives and that some level of economic growth is necessary to achieve these goals. I also agree that encouraging innovation and improving the "technique" of our economy is a critical objective as well.

However, missing from the analysis contained in either paper is the impact of globalization on the ability of individual nation states to shape economic activity and a recognition that not all technological innovations are consistent with the goal of social sustainability. The swimmer to which Prof. Romer refers might discover a technique that helps him or her win races in the short term, but which might injure the swimmer if used excessively.

The traditional view of increased globalization is that it will increase productivity through greater economies of scale and a more effective use of world resources. This view holds that while there may be some countries that suffer more than others during the period of structural adjustment, the world as a whole will be better off.

However, globalization is proving to be a double edged sword. While there may be greater economies of scale for some industries, there is also tremendous downward pressure on labour standards, levels of government regulation and social programs. Thus far, increased globalization has given rise to increased competition among nations and regions for available capital based on who can offer lower wages, lower environmental standards, lower taxes and greater government subsidy.

South American nations regularly use subsidies, low taxes and the promise of lower labour and environmental standards to lure Canadian mining companies to explore and develop operations in that part of the world. Since mining provides so many high paying and high skill jobs, and accounts for 4.6% of Canada's GDP and over 16% of its exports, this is indeed a unfortunate trend.

We also have to be aware of the limits of focusing exclusively on innovation and higher value added production as the means to solve our economic problems. Certainly there are diminishing returns to an increasing focus on innovation. More importantly, I don't see how all nations can compete internationally by focusing on innovative product design and high value added production. If they try to, the potential for sustaining a standard of living higher than that of other nations will certainly be diminished.

Even in the short run, exchange rate fluctuations can virtually wipe out any advantage gained by a national strategy focused on innovation and higher value added production.

I believe we must develop a better understanding of the role of institutions and regulation in constraining low wage competition and find new ways to constrain low wage competition on an international basis. Increased international economic activity can either increase economic efficiency or decrease it, depending on how firms compete. If a competitive edge is achieved primarily because of low wages, the absence of environmental regulation and some combination of low taxes and government subsidy, how can one argue that economic efficiency has been maximized?

I also believe we must develop a better understanding of the impact of technological change on social and political sustainability. The number of person-hours required to make a ton of steel has been the most critical productivity measurement in the steel industry for a long time. Similar measures are used elsewhere as well. By focusing, as we do so often, on how to lower unit costs by producing goods with fewer workers, we ignore the impact on aggregate employment.

The traditional view is that increased productivity will generate wealth and a demand for new products which will then employ the displaced workers. However, increases in productivity have outstripped increases in demand for most manufactured goods in Canada over the past twenty years and the number of new products and services developed have simply not taken up the employment slack. If the market is not generating new products or services that consumers want or need, what do workers who are displaced by technology do for a living? Once every retail outlet and every fast-food outlet is open 24 hours a day, I think our employment problem will get even worse.

As a trade unionist, I am disturbed by key labour market trends. In 1976 the unemployment rate in Canada rose above 7% for the first time since World War II. Since 1976 it has never been below 7%. We have experienced an increase in both long term unemployment and in part-time work. Real wages have been stagnant or declining despite increases in productivity.

Both globalization and technological change have contributed to these trends.

We are not yet in a truly global economy. Most economic integration has occurred regionally and largely between OECD nations. There are today, few truly multinational corporations. Nonetheless, the world is certainly moving toward greater integration and the slope of the trend line is increasing rapidly.

I believe we must respond to this by finding ways to regulate internationally. Various nations introduced collective bargaining, health and safety regulations, environmental regulations and tax systems to achieve certain social objectives that could never be attained through unregulated economic activity. If nations and regions are less able to shape economic activity individually because they must compete internationally for available capital, then they must do it collectively.

By negotiating labour and environmental standards as part of trade agreements or by establishing international bodies that establish and enforce standards, we must find a way to constrain international low wage competition.

We also have to better understand the social and political implications of technological change designed primarily to displace workers. Technology can eliminate dangerous or

repetitive work and increase the skill content of jobs. It can also eliminate large numbers of good jobs and create unsafe and repetitive work. We must learn to drive our technology agenda at least partially by broader social objectives.

Thus to achieve the social and political sustainability discussed by Prof. Takeuchi, we need not only the commitment to innovation discussed by Prof. Romer, but we need to ensure that innovation is directed at improving the quality of life and we need ways to ensure that international competition based on low standards is effectively constrained.

# Session Eight: Health Care and the Economy: Efficiency and Effectiveness in Health Care

---

*Naoki Ikegami*

## **I Strategic framework: Harnessing health care's unique characteristics**

### **1. Difficulty of evaluating effectiveness**

The basis of believing that health care is effective would seem to rest on the following. First, the treatment is justified because it counteracts the pathological mechanism as understood by the current biomedical research. Second, there is empirical evidence from the before and after treatment conditions to suggest that it is effective. Third, this evidence has been validated by a clinical control study. Only the third provides conclusive evidence that the results are better than placebo effects but very few procedures clear this criterion. Control studies require a great deal of time to undertake and they can never keep up with the rapid growth of new technology and the physicians' eagerness to try new methods (Wennberg, 1990). Even for the few which have been rigorously evaluated, they frequently come to be used in situations in which the effectiveness has not been established as witnessed by the expanded use of cimetidine (Hall, 1981). Claims are also made that improvements on the existing technology have made it effective despite the inconclusive results shown in the original study. Since clinicians witness an actual improvement in the patient's condition, there will always be strongly held beliefs about the technology's utility. Thus, it is likely that most of health care will continue to be justified only on the first and second criteria for which it is impossible to provide any conclusive evidence concerning their effectiveness.

It is useful to bear in mind that the classic pattern of health care is a closed, individual encounter between the physician and the patient. Since it is closed, the information that can be collected for the purposes of evaluation is limited and subject to bias because it can only be provided by either the physician or the patient. But physicians make their livelihood by providing care, while patients are seldom in a position to objectively evaluate the utility of the process. Moreover, since it is individualized, even if accurate information can be obtained, this will, strictly speaking, be only valid for the unique conditions in which the encounter took place (Ikegami, 1989). Given these circumstances, it is hardly surprising that it is so difficult to evaluate whether the health care service that is being provided is effective, let alone efficient. For to be effective is a precondition to be efficient.

### **2. Inbuilt mechanism which resists change**

Another aspect of health care which needs to be taken into account is that once the pattern of service provision is established, it becomes remarkably resistant to change for the following reasons.

First, as noted, health care is essentially a closed, individual encounter between two actors. Since it is extremely difficult to monitor the process, the solution has been to ensure

the quality by professional manpower. That is, to educate and train people so that they come to possess the standard skills that enable them to provide a standard level of care without supervision. Furthermore, since health care has become so complex and requires different skills, there has been a need to subdivide the profession into specialties. These two factors have combined to make the hospital a "professional bureaucracy" (Mintzberg, 1983), in which conflict is avoided by each agreeing not to encroach on the other's domain of expertise. This mutual accommodation is similar to that prevalent in craft unions except for the fact that the skills are required on an off-the-job academic setting rendering it even more resistant to change. It should also be noted that the resulting rigidity in the division of labor is not limited just to within hospitals and physicians, but is pervasive throughout the health care system and for all personnel with licensed qualifications. Thus, any changes will be uniformly resisted because they lead to a disruption of the existing order and pose a threat to the livelihood of those engaged in health care.

Second, in health care, it is a strongly held belief that, for the same clinical condition, the same quality of care should be provided for all, regardless of the ability to pay or geographical location. However, if equality is to be valued above all, then it becomes next to impossible to introduce any notion of competition. Because even if adverse selection can be eliminated, there would still be legitimate competition based on price and quality. However, should this information become available to the public, it will be extremely difficult to guarantee an equal provision of care. For if it became known that a certain provider had the highest quality but at a high price, then that difference would have to be paid by either the patient or the public payor. However, even the latter will not conform to egalitarian principles because the public would be paying extra for the patient who had the fortune to receive care from a high price, high quality provider. There would still be a problem even if the provider had both high quality and low prices because of the resulting concentration of patients which, in turn, can only be resolved by differential payment, queuing or lottery. Of these three, lottery may be the most egalitarian but, at the same time, operationally the most impracticable. Thus, it is necessary to let the public assume that every facility and physician is providing a satisfactory level of care. This is a claim that is often made by professional organizations to protect their members' interest which means that the proponents of equality will find powerful allies among the providers. Both would maintain that selection should be based only on the patient's convenience and affinity, and not on the quality of care. Since this would not lead to real competition, it will be very difficult to change provider behavior.

Third, consumer satisfaction is strongly related to consumer expectations and especially so in health care where real benefits are so difficult to evaluate. This means that patients tend to prefer what has been provided in the past or what they perceive as conforming to their notion of the current social norm. Furthermore, because the current pattern of services will be regarded as an entitlement, any revisions will be fiercely resisted. Thus, patients will willingly join forces with physicians so that the existing pattern of resource allocation may continue to be maintained even if the benefits are found to be, at best, marginal. Therefore, although the demand may have been initially induced by physicians, it will quickly be transformed into a legitimate "need".

### 3. Strategy for cost containment

It is possible to derive a strategy for cost containment from the above characteristics. Since the effectiveness of health care is established in only a relatively small area of health care, a global budget can be introduced without necessarily imperiling the quality at least in the developed countries where the vicious circle of poverty and illness does not affect the greater majority. However, the major problem that has been plaguing health policy makers is

the continuous introduction of new technology which has put a strain on the budgetary capabilities. It is here that the second aspect of health care may be utilized. That is, once a global budget is introduced, it is possible to rely on the medical profession's inbuilt propensity to maintaining the status quo and the existing allocation of resources. It follows that policy should be directed at bolstering this tendency so that changes, including cost escalation, can be kept to a minimum. The cooperation of the most powerful of the stakeholders among the providers should be sought by pointing out that any revisions in the allocation pattern is likely to be to their disadvantage. This is because should it be left to a natural course, their current favorable position would quickly become undermined by technological changes. The consumer's voice can be conveniently ignored without too much public protest because, as noted, their expectations are geared to what is currently provided and health care is very much a parochial industry.

Such a policy has, in fact, been pursued by Japan so that it has been possible to contain health care costs and maintain the same ratio to the GDP for the past ten years. This was not necessarily a conscious decision made by the government but was more the result of the country's unique socio-historical characteristics. Naturally, the maintenance of the status quo came at a significant social cost. Japan's health care system gives the impression of being under a time slip and contrasts sharply with the highly advanced condition found in the other industries (Iglehart, 1988). The fee-for-service solo-practitioner having an ill-defined clinical specialty is still the dominant figure. Most of the hospitals are small and physician owned, and have very basic amenities for the patient. Formal quality assurance programs are largely non-existent and providers are very reluctant to disclose any information. The providers are still able to successfully uphold, at least formally, the notion that their standards of care are equally good. Yet, despite (or perhaps regardless of) this situation, the health indices of infant mortality and life expectancy at birth are excellent. Moreover, there is no overt rationing and the per capita number of equipment such as CAT scans is the highest in the world (Ikegami, 1992). The following describes the underlying mechanism which has led to the present state, the problems it now faces, and the future prospects for reform.

## **II. The state in Japan**

### **1. Historical background**

Two aspects stand out as unique to Japan. One is that although medical licenses became restricted to those who had studied western medicine in 1883, provisions were made so that the pre-existing practitioners of Chinese medicine (and even their sons) were allowed to continue providing care and dispense medication. This meant that the country had an extensive, well-established network of medical practitioners which was readily available to the general population for looking after their common complaints. (At the time of the first national census in 1871, the number who gave medical practice as their occupation amounted to 86.6 per hundred thousand (Fuse, 1979)). These practitioners became mandatory enrolled and constituted the greatest majority of the pre-war (Japan Medical Association, JMA) when it was established in 1916.

The other is that there is only fragmented historical evidence of religious or secular orders which have provided institutionalized care (Sakai, 1982). Thus, the hospitals, which began to be established as part of the national policy of rapid westernization adopted in 1868, were the first to provide institutional care. Most developed from the clinics of the physicians who had the business acumen to expand their practices. Those established in the public sector were primarily for the purposes of teaching, the military and the control of communicable and ven-

eral diseases. Thus, prestigious hospitals remained largely confined to the few teaching hospitals and, furthermore, only their clinical departmental chairs were perceived as elite physicians.

These two factors meant that, when the government first legislated social insurance for manual workers in 1922, the two most powerful stakeholders within the medical profession were the private practitioners in clinics and a very small number of academic physicians. The hospital based physicians were entirely subordinate to the university clinical departmental chairs who had in fact assigned them to their positions. Under these circumstances, it was not surprising that the initial fee schedule, the point fee system, was largely for the payment of ambulatory services provided by the practitioners in clinics. In fact, its basic unit (the point) was the inclusive price of one standard drug per day reflecting the continued importance of dispensing for physicians. The original fee schedule was important because it was later adopted by other plans as the population that was covered by social insurance expanded. Private payment continued for the wealthy and was an important source of revenue for the prestigious physicians.

Thus, the JMA, representing the interests of the private practitioners, played a dominant role in the fee schedule negotiations from the very beginning and have continued to do so to this day. This was because although the number of hospital-based physicians has increased dramatically, their position has remained weak since they have remained under the patronage of the university professor. This hierarchical structure of the medical profession has hindered the development of collegial specialty associations despite the increase in specialized training. Furthermore, hospitals have been divided into the small but more prestigious public sector and the large private sector, each viewing the other with antagonism. These factors were skillfully exploited by Takemi, the charismatic chairman of the JMA from 1957-1982, to further consolidate the JMA's paramount position.

## 2. How the fee schedule is negotiated

The fee schedule has played a crucial role in both cost containment and in defining the scope and standard of services which are provided. The scope is defined because the procedure or the drug cannot be provided in conjunction with other services covered by social insurance before it is first listed in the fee schedule. This means that, for all practical purposes, it cannot be provided at all because everything would have to be provided out-of-pocket. The standard is also determined because the reimbursement for each item is made regardless of its cost to the provider and balance billing is strictly prohibited. The same fees are enforced on all providers regardless of their clinical setting, from the village clinic to the university medical center.

The actual negotiations for the revisions in the fee schedule take place in the Central Social Insurance Medical Care Council of the Ministry of Health and Welfare. The council comprises eight representatives from providers (five physicians, two dentists, and one pharmacist), eight from payors (four from insurers, which includes two from the government, and two each from management and labor), and four that represent public interests (three economists and one lawyer). All five physicians have come to be nominated by the JMA ever since the representative nominated by the Japan Hospital Association was excluded by JMA's pressure in 1963. Thus, the JMA has played a crucial role in the fee schedule negotiations because it is only through their channels that the physicians and hospitals can voice any of their demands.

The fee schedule negotiations are a two stage process. First, the estimated allowable increase in the total health care expenditure is decided by the negotiation between the Ministry of Finance and the Ministry of Health and Welfare. Data from the survey on the

financial state of hospitals and clinics (the latter being equivalent to a study on physicians' income) which is made every second year is presented to provide an objective basis for the change. However, in point of fact, the Ministry of Finance has had an decisive role, especially since the fiscally tight years of the 1980's. This is because the government provides indexed subsidies to the insurance plans having a weak premium basis from its general expenditure budget. These subsidies amount to a quarter of the total health care expenditures and are necessary for providing an egalitarian health care under a multiple payors system.

The second stage consists of negotiations between the Ministry of Health and Welfare and the JMA on how to reallocate this sum by individually readjusting the fees for each of the thousands of procedures and drugs (an across the board increase for all items cannot be made because the conversion factor from the "points" to yen has been fixed to 10 yen a point). The effect of revising the price of each item is calculated by multiplying it by its volume, the estimate of which is available from the national survey of claims data. Ultimately, the total amount must be kept within the already decided total health care expenditure target. Since virtually all the fees are changed periodically, the effect of any revision is never the same for any individual provider so that it is difficult to definitively say who has benefited most from the revision. The complexity of this process has worked towards containing conflict among the providers.

However, the main reason why conflict has been kept to a minimum lies in the cardinal principle of maintaining the status quo: continuous efforts have been made to keep the relative share of the pie between inpatient and outpatient care, hospitals and clinics, and among the various specialties as constant as possible. Should any unforeseen major imbalance or areas of grievance arise, it would be redressed in the next round of fee negotiations. This principle has worked very advantageously to the private practitioners because originally, health care consisted of mostly outpatient care provided from their clinics. Surgical fees were low to start with and have not, in general, increased as much as consultation fees (Takagi, in press). The JMA has demanded and obtained various measures such as additional consultation fees for internal medicine and pediatrics so that the private practitioners' share could be maintained. At the same time, the JMA and the government have collaborated to contain hospital costs by decreasing hospitalization fees as the length of stay increased, and lowering their outpatient consultation fees. By such measures, the government has achieved the goal of cost containment while the JMA has been able to protect the interests of the private practitioners, its major constituent.

The power of the private practitioners is also apparent at the regional level. The local medical associations nominate members to the peer review board of the clearing houses for claims which are established in each prefecture. This board reviews the detailed item-by-item claims sent by every health care facility at the beginning of each month and denies payment if they judge that the numbers of procedures or drugs has been excessive. The actual ratio of the claims denied is less than 1%, but because these are concentrated on the high cost claims, the review does have some effect. Also, the denials tend to have a sentinel effect on the provider because should they gain a reputation for excessive services, the ad hoc nature of the claims review process means that their claims would in the future be subject to a more intense examination. These measures have restrained aggressive revenue maximization and market expansion by hospitals. The other area in which the local medical associations wield power is through the prefectural health planning committee (in most cases, the committee's chair is held by the chair of the prefectural medical association). The establishment of new hospitals or the expansion of bed capacity has to be approved by these committees. These regulations were imposed by the 1985 regional health planning legislation, the objectives of which lay in cost containment for the government, and the prevention of hospital expansion for the private practitioners.

### 3. Pricing of new technology

The above description of the fee schedule would seem to be incompatible with the highest per capita number of CAT scans and renal dialysis in the world. The answer lies in the general policy which is adopted towards new technology. The initial fee is usually set at a level which is only slightly above that of the nearest existing technology. This does not normally cover the full costs but many providers will try to make the technology available in order to attract patients and physicians. At the same time, the manufacturers will compete to market an inexpensive model that would have all the essential features of the more expensive type. This will, in turn, trigger a downward spiral in prices, a classic example of which is CAT scans. It should also be noted that diagnostic equipment is usually not labor intensive, nor would poor quality equipment or manpower likely to lead to a direct adverse effect on the patient. Both would become more problematic for invasive technology which may explain the fact that the per capita number of surgery is about a third that of the United States (Hasegawa, in press). In any event, through the low pricing of new technology, it has been possible to maintain the appearance of professional freedom and avoid accusations of rationing, at the same time contain costs.

One possible exception to the above process is renal dialysis. In this case, a policy decision was made to speed its usage by deliberately setting its fee above the estimated costs. When this was more than successfully achieved, fees were lowered. However, since the manufacturers have continued to lower their prices for the equipment and for the disposable material, it is still possible to generate a substantial profit if the facility is large enough to have the merits of economies of scale. Patients with end stage renal disease are sought after by the providers and it is possible that dialysis may have been started before it was truly warranted.

This illustrates the difficulty of using the fee schedule as a device to control the use of technology. It becomes even more problematic in the case of pharmaceuticals. When a new drug is approved, the price is set at a maximum level of 10 percent higher than the nearest pre-existing product. However, the merit to the manufacturer is much more than this figure would suggest. This is because the price paid by social insurance has been periodically reduced following the national survey of the actual price paid by the provider which is conducted by the government every other year. Since the provider purchases the pharmaceuticals in the open market, competition among the manufacturers regularly drives their prices down, thus, setting the stage for the next reduction in the price paid by insurance. Through this process, the insurance price of some drugs have been reduced by over half in the past ten years. Therefore, new drugs are lucrative for the manufacturers even if their prices are set at only 7% more because the price of the pre-existing drug would have been reduced at the time of the introduction of the new drug.

If the marginal benefit of using the new drug is justified by the marginal increase in price, then the present system that of using competitive market forces to drive prices down would be ideal. However, as has been stated at the beginning, it is very difficult to evaluate the effectiveness of technology, including drugs. The present system gives incentives to prescribe quantitatively more drugs, and to qualitatively shift to new drugs with high prices from old drugs and generics. This is the probably one of the reasons why Japan's pharmaceutical expenditures are the highest in the world despite the fact the individual prices for brand name drugs are lower in Japan when compared to the United States (Ikegami, in press).

### III. Structural problems facing Japan

Japan has been able to achieve the seemingly impossible combination of cost containment, universal coverage and no overt rationing. This has not been the result of pursuing effectiveness and efficiency, but by maintaining the status quo. Given this situation, the best option may appear to be the continuation of the present policy. However, structural changes have occurred which has made the maintenance of the status quo increasingly difficult. First is the rapid expansion in long-term care due to the demographical and social transitions. The ratio of the elderly 65 and over has increased from 5% in 1960 to 12% in 1990 and will reach 25% in 2020. Furthermore, over half of married women are working outside of their homes so that the traditional family support has become increasingly difficult to provide. These factors have led to a massive increase in the ratio of the elderly institutionalized, the increase being from 1% of the elderly 65 and over in 1960, to 6% in 1990. Three quarters of the elderly institutionalized are in hospitals due to the lack of development in alternate forms of care (Ikegami, in press). However, the share of long-term care to total health care expenditure has been compressed because of the need to maintain the balance among the providers. This has led not only to a poor quality of care, but also significant out-of pocket payment for the patient (67,000 yen according to a survey made by Niki (1992)). Although these extra charges are legally prohibited, they are in reality unofficially condoned.

Second is the changing expectations of the consumer, especially of the younger urban population. The globalization of the economy has led to an enhanced awareness of the world outside of Japan which is beginning to effect even the parochialism in health care. Concepts such as informed consent and patient's rights have become familiar to the general population through the influence of mass media. More concrete examples such as the poor standard of amenities in hospitals has also drawn attention. At the industrial level, the process of harmonization in the pharmaceutical industry has highlighted the need to have more rigorous clinical trials and a more open distribution system for drugs. In all these aspects, it is nearly always with the United States that comparisons are made. While the United States' health care system leaves much to be desired, it does provide an "other" for the growing dissatisfaction with the state in Japan. One manifestation of the disenchantment with private practitioners is the growing concentration of patients to the large prestigious hospitals which are mostly in the public sector. Since these hospitals receive subsidies which supplement the low revenue coming from the regulated fees, they have been able to maintain standards which are closer to those found in Europe and North America.

Third is the structural changes which are occurring within the medical profession. At the time when Takemi became chairman of the JMA, private practitioners were the majority and most were in their forties. Their ranks have swelled due to doubling of medical school enrollment during the war time years. Now due to the doubling of the medical school graduates and the decline in the number of young physicians going into open practice, private practitioners constitute only a third of the total. Their average age is 62. While the JMA continues to be politically powerful, it is hardly surprising that their leadership lacks dynamism. The prospects for young physicians are none too bright because not only are they less well adapted to go into open practice due to the specialized training they have received, but also the increase in initial capital investment that is required to set up the practice has become a major hurdle especially in the urban areas.

Both the government and the leaders of the JMA are to a certain extent aware of these monumental changes. But so far, their hands have been tied by the general policy of maintaining the status quo. Any radical changes would, at least in the short term, firstly, raise costs, and secondly, place the private practitioner at a disadvantage. The former would be precluded for fiscal reasons should it be financed by social insurance, or for egalitarian princi-

ples should it come from private payment. It is worthwhile noting here that unlike many west European countries, Japan does not have any officially recognized private payment system. Extra-payment is made only to elite physicians in the premier medical centers. It would be difficult to open this black market and expand it. Concerning the latter, it would be suicidal for the JMA establishment to disregard the interests of the private practitioners since they are their major constituents.

#### IV. Strategies for change

To use a medical terminology, it would be next to impossible to find the "complete cure" for Japan's health care system. There is not going to be a magic bullet solution to any of the problems because of the resistance to change which has been outlined in the beginning which has been accentuated by the mutual accommodation between the government and the JMA. To recapitulate, first, changing the pattern of health care delivery would entail changing the way professionals earn their living. Since this has come to be regarded as their entitlement, and since professionals have autonomy in their decisions, this would be difficult to realize. Second, the value placed on egalitarian principles precludes consumer choice and tends to foster a rigid, monolithic provision of services. Third, while it is true that the consumer has become more dissatisfied with the system in Japan, they would, at the same time, like to retain the easy access to outpatient care which they feel is their entitlement, and the low health care costs.

Given these circumstances, change would seem to be only possible on an incremental basis and in areas where it is both budget neutral and does not put the private practitioners at a disadvantage. It is for this reason that the author has proposed reforms in the two areas which meet these conditions (Ikegami, 1993). The first is aimed at the acute care provided by large public hospitals. The present fee-for-service system would be replaced by inclusive per diem rates for inpatient care and inclusive per visit fees for outpatient care which would be negotiated at the prefectural level for each hospital that opts for this new payment mechanism. These rates would be initially set at each hospital's present average earnings per patient. However, from the next year, it would be tied to the hospital's performance indicators. These would be selected from the following: productivity indicators such as bed occupancy rate and average length of stay; indicators for the provision of tertiary care such as the number of complicated surgical operations and tertiary emergency care; indicators evaluating the establishment of networks with the private practitioners such as the ratio of referrals among new outpatients; and indicators evaluating patient satisfaction. The number of indicators to be selected, their weighting, and finally their performance targets would be negotiated at the local prefectural level by an organization having a similar membership composition as the Central Social Insurance Medical Care Council.

This new form of payment for acute care would be budget neutral and, indeed, could lead to a decrease in total costs because public hospitals would have less need for subsidies. Efficiency savings could be realized because there would no longer be a need to dispense medication and order laboratory tests in order to generate revenue. The JMA has already proposed its implementation in its report from the Health Policy Committee. From their standpoint, it has the merit of freezing the share of public hospitals and also making these hospitals responsible for referring back patients to private practitioners.

The second area is in long-term care. The government has already established an inclusive per diem rate for which geriatric hospitals may opt instead of the regular fee-for-service payment if they meet the required staffing level. The problem with this method is that although it solves the problem of over-medication and the ordering of too many laboratory

tests, it gives the incentive for hospitals to admit only light care cases. To deal with this new situation, the author has been conducting validation studies on RUG-III (Resource Utilization Group, Version III), a case-mix grouping system for long-term care developed in the United States. In addition, in order to take the first steps towards ensuring quality, validation is also concurrently being made on the MDS (Minimum Data Set) and RAPS (Resident Assessment Protocols). The MDS will provide the comprehensive assessment while the RAPs will be useful in drawing care plans. Both are virtually unexplored areas in Japan. While the introduction of these measures would not necessarily add more resources to long-term care, it would contribute to its more efficient use and, by highlighting the inadequacies in the present system, may lead to the way for more radical reforms.

High-tech acute care and long-term care are the two areas in which the greatest cost escalation is likely to occur if the present system were to continue. Coincidentally, they are also the areas which are marginal to the interests of the private practitioners, and indeed, can be regarded as their serious competitors to scarce health care resources. Therein lies the greatest hope for reform. Once these changes can be effected, it may be possible to set the stage for the next round of reforms.

These strategies for change are unique to the peculiar circumstances of Japan. Certainly, maintaining the status quo in countries which allocate relatively more resources to high-tech care than Japan would not lead to cost containment. However, a similar examination of the health care policy arena in each country may bring to light the weaknesses in the forces which maintain the status quo. Based on this analysis, a beginning can be made for realizing effectiveness and efficiency in health care.

## References

Fuse, S, History of Doctors. Its Japanese Characteristics (Ishi no rekishi. Sono Nihonteki Tokusei) (Tokyo: Chuou Kouronsha, 1979).

Hall, W.J, "Use of Cimetidine in hospital patients," Canadian Medical Journal 124 (1981): 1579-1584.

Hasegawa, T, "A comparative study on the hospital admission rate between the US and Japan", in Ikegami, N. and Campbell, J.C. eds., Controlling Costs in Japanese Health Care (Ann Arbor: Michigan University Press, forthcoming).

Iglehart, J.K, "Health policy report: Japan's medical care system", New England Journal of Medicine 319(1988):807-812 and 1166-1172.

Ikegami, N, "Best medical practice: The case of Japan", International Journal of Health Planning and Management 4 (1989): 181-195.

Ikegami, N, "Japanese health care: Low cost through regulated fees", Health Affairs 10 (3) (1991):87-109.

Ikegami, N, "The economics of health care in Japan", Science 258(1992):614-618.

Ikegami, N, "A new fee schedule to realize functional differentiation - Reconstructing the fee schedule and subsidies (Kinou Bunko notameno Shinshihari Houshiki-Shinryou Houshu to Hojyokin no Saikouchiku o Mezashite)", Shakai Hoken Jyunpou 1800(1993):6-9, and 1801 (1993): 10-13.

Ikegami, N, "Comparison of long term care for the elderly between Japan and the US", in Ikegami, N. and Campbell, J.C. eds., Controlling Costs in Japanese Health Care (Ann Arbor: Michigan University Press, forthcoming).

Ikegami, N, "Comparison of pharmaceutical expenditure between Japan and the US", in Ikegami, N. and Campbell, J.C. eds., Controlling Costs in Japanese Health Care (Ann Arbor: Michigan University Press, forthcoming).

Mintzberg, H, Structuring in Five. Designing Effective Organizations (New Jersey: Prentice Hall, 1983), 189-214.

Niki, R, Health Care and the Fee Schedule of the 90's (90 Nendai no Iryou to Shinryou Houshu)(Tokyo: Keisou Shobou, 1992), 198-230.

Sakai, S, History of health care in Japan (Nihon no Iryoushi) (Tokyo:Tokyo Shoseki, 1982)

Takagi, Y, "The effectiveness of the Japan Medical Association in maintaining the financial position of private practitioners", in Ikegami, N. and Campbell, J.C. eds., Controlling Costs in Japanese Health Care (Ann Arbor: Michigan University Press, forthcoming).

Wennberg, J.E, "Outcomes research, cost containment, and the fear of health rationing", New England Journal of Medicine 322 (1990):1202-1204.

# Can We Have Too Much Health Care?

---

*Greg L. Stoddart*

*John N. Lavis*

**T**he health care industry uses resources - and large amounts of them in most industrialized nations. In an era of slow economic growth this creates increasing economic and political pressures for governments concerned with sustaining their health care systems, regardless of the public-private mix in the financing and provision of health care.<sup>1</sup>

The tensions are exacerbated by the strong internal momentum for expansion which is a universal characteristic of the health care industry. This in turn is fed by several factors, including the attitude of providers of health care who, as Evans has commented, often "take for granted that their work is more important than any other uses a society might have for its resources".<sup>2</sup>

Another important factor is a view of the determinants of health which accords health care the primary role, despite accumulated - and accumulating - evidence to the contrary.<sup>3</sup> An oversimplified version of this view, yet one which, judging by their actions, is implicitly held by the general public and policy-makers as well as health care providers, is depicted in Figure 1. In this "heating system" model, people become sick or injured through a variety of external and unspecified reasons. They present themselves to the health care system, where their needs for care are professionally interpreted, and the professional providers then respond by delivering the 'necessary' services, within whatever institutional arrangements for access and payment exist. The health care system is both thermostat and furnace; the response to colder temperatures (more "disease") is to turn up the heat (provide more health care). Health policy effectively becomes health care policy.

A more appropriate conceptual framework for consideration of the determinants of health is both more complex and more subtle. An again oversimplified view of one such framework is presented in Figure 2. There are several notable features of this framework. First, health care has been removed from centre-stage; it is now but one of several determinants of health. Second, the framework allows for the importance of population- as well as individual-based choices and interventions. Third, health as an objective is placed in a broader policy context. In addition to health care and health policies, which are targeted at health, the framework illustrates the potential significance of public policies in other sectors which may be targeted at health or may affect health indirectly.

The 'big picture' in Figure 2 was developed by Evans and Stoddart in order "to provide meaningful categories in which to insert the various sorts of evidence which are now emerging as to the diverse determinants of health."<sup>4</sup> How a problem is framed is important. It determines which types of evidence are deemed relevant, and therefore given weight, and which are ignored or discarded. It also determines the set of admissible - or probable - questions for discussion. In this respect, the progression from the "heating system" view of the determinants of health in Figure 1 to the broader view in Figure 2, with its complex interactions and trade-offs, allows a question of increasing importance at least to be posed for discussion: Can we have too much health care?

## Clarifying the Question

Our intent in this paper is to begin to explore this question by suggesting some analytically distinct ways in which it may be possible to have too much health care, heretical though that may sound to some. Our purpose is more to raise basic issues and questions rather than to resolve or answer them, and in the process to consider what types of evidence exist or might be generated to address the complex trade-offs involved in the use of resources for health care instead of other activities. Some of the ways which we identify will seem obvious. Others will be less straightforward, and less subject to agreement, by experts and non-experts alike. Nevertheless, the explicit placement of the question on the agenda of admissible items for analysis and discussion (in our view) represents the achievement of a modest goal. We anticipate that the exchange of comments at this conference will assist in clarifying, revising and completing the survey of this difficult terrain that we have begun.

We readily acknowledge that the manner in which the question has been posed creates a potential minefield of further questions, definitions and value judgements. Who is "we"? What is included in "health care"? What is meant by "health"? And against what yardstick (and by whom) is "too much" to be judged? The last of these questions is the substance of the body of the paper below; however, the others can be clarified at this stage.

By "we" is meant both a society (or population) committing scarce resources to the production and consumption of health care and, at times and depending upon the specific context below, an individual receiving a particular health care intervention. "Health care" is to be interpreted in a rather traditional (some may say 'narrow') way to be the set of services, procedures or programs (and the personnel and facilities allowing the capacity for their delivery) that physicians and other professionals (including the so-called 'allied' health professions and administrators) provide for the primary purpose of restoring, maintaining or improving health. This includes services which may be thought of as part of 'public health' programs, although again with a relatively narrow focus on such things as immunizations, maternal and child health, and control of infectious diseases, for example, rather than activities such as safety campaigns or environmental cleanups.

Health itself may also be defined in a narrow or broad manner. The World Health Organization definition that "health is a state of complete physical, mental, and social well-being" is the broadest possible definition, but is so all-encompassing as to make all human activity health-related and all policy health policy. While in some ultimate sense - and in the context of a broad view of the determinants of health - this may be so, we follow Evans and Stoddart in labelling the WHO concept as "well-being", to which health is an important but not the only contributor. Again following Evans and Stoddart, and as illustrated in Figure 2, we restrict the interpretation of health for the purposes of this discussion to the absence of negative biological circumstances associated with disease or disability, but do make the important distinction between this phenomenon as defined and treated by those in the health care system (labelled "disease" in Figure 2) and as perceived and experienced by the affected individuals themselves (labelled "health and function"). Thus there is a "separately identifiable realm of specifically health-oriented activities"<sup>5</sup> but they may be seen and valued quite differently by patients than providers because the two may differ in their implicit notion of health. Patients' notions of the absence of illness as defined by capacity to function need not match providers' notions of the absence of disease, even for the same negative biological circumstances.

The question of whether a population or an individual can have too much health care is of course quite a different one than the question of whether a particular population or individual does have too much health care. We limit the scope of this paper to the former, although the examples that we use for illustrations are often directly relevant to the latter question as well.

As a final introductory comment, we also wish to emphasize that the intent of this paper is not to deny the significant contribution that health care does make to health, nor to denigrate the dedicated efforts of its practitioners. No thinking person would suggest that we dismantle the health care system, nor that acceptable levels of individual and population health can be maintained without health care. Rather, the discussion here seeks to provide balance and perspective - a small 'voice on the other side' - on the seemingly insatiable appetite of health care systems everywhere for more resources. It is a reminder that health care, like all resource-using activities, has opportunity costs which warrant consideration in the 'big picture' of the determinants of health and well-being.

The following sections identify and briefly discuss some analytically distinct ways in which it may be possible to have too much health care. A tentative taxonomy is introduced after the listing, along with several caveats. We then revisit one theme of particular importance to this conference - the relationship between health care spending and economic performance - and offer some concluding observations.

### Health Care That Is Not Effective

If health care is not doing what it is supposed to do, i.e., prevent or cure disease, alleviate symptoms, or restore, maintain or improve function, for individuals (B)<sup>8</sup> or populations (C) then the resources devoted to care are simply wasted.<sup>7</sup> In economic terms, there is no final output (health), even though an intermediate output (health care) has been produced.

The failure of health care procedures, services or programs<sup>8</sup> to perform their task may occur in several different ways and for different reasons. Some interventions may actually do harm to patients, or do more harm than good.<sup>9</sup> Other interventions may be ineffective in practice. The provision of ineffective care may be due to any of the following: a lack of information (or use of information) on the effectiveness of a given intervention; the way in which available information is presented to health care providers; or the way in which such information is interpreted by providers.<sup>10</sup> It may also be the case that an otherwise effective service is rendered ineffective because it is applied in the 'wrong' clinical circumstances, i.e. when it is not clinically warranted.<sup>11</sup>

### Effective Health Care That Is More Costly Than It Need Be

This may occur in two ways. First the resources (personnel, equipment, facilities and know-how) used to produce health services may be employed in ways or combinations that yield lower output (i.e. fewer services) than is technically or physically possible (A). An alternative way of stating this is to say that by deploying or combining the resources committed to health care differently, a higher level of output (i.e. more services) could be achieved with the same resource commitment.<sup>12</sup>

Second, "costly" may be defined in a pecuniary sense, thereby taking into account the prices of the inputs used in the production processes for care. In this case, care is more costly than it need be if a larger quantity of services could be obtained with the same level of expenditures or, alternatively, if the same quantity of services could be obtained with a lower level of expenditures (D). In other words, although the services may be effective, they are not cost-effective.<sup>13</sup>

### Health That Is More Costly Than It Need Be

Health care is only an intermediate output. Resources are committed to it for its ability to contribute to another production process, the output of which is health. Just as health care

can be produced in more costly or less costly ways (in either a technical or pecuniary sense) so too it is possible to produce health in more costly than necessary ways. Once it is recognized and accepted that health care is but one of the determinants of (inputs to) health, it becomes apparent that another possible interpretation of the phrase "too much health care" refers to the circumstance in which a given level of health (again for either an individual (E) or a population (F)) could be achieved with a lower level of overall resource commitment involving either less of some types of health care relative to others<sup>14</sup> or less of health care and more of 'other things' which contribute to health.<sup>15</sup> (Alternatively, the same is true if a higher level of health could be achieved with the same overall resource commitment but one involving less health care in either of the ways mentioned immediately above.) In other words, even cost-effective health care can be seen as competing for resources with other activities that also contribute to health.<sup>16</sup>

It seems useful to divide these other, competing uses for resources into two groups, distinguished by their orientation or purpose with respect to health. In the first group are activities outside the health care sector whose primary purpose (or at least one of their purposes) is nevertheless health-related (eg. industrial safety programmes or environmental regulations). In the second group are activities with a more general social purpose (eg. education, day care, social supports, pensions) but ones which are increasingly being shown to have important health consequences. Economic growth can be considered another such activity. We return to the implications of this in a later section.

#### Health Care That is Valued At Less Than Its Cost

Thus far, interpretations of "too much" have been based on production, or supply-side, considerations. The production of either health care itself or the health that health care in turn produces in conjunction with other determinants has been more costly than necessary for one reason or another. But cost and supply constitute only one side of an economic ledger. Just because something is or can be produced cost-effectively does not mean that it should be produced, either at all or at a particular level or rate. That judgement depends on an assessment of benefits as well as costs, and brings in consumption, or demand-side, considerations. In other words, the benefits of health care must be valued in relation to its costs.

This valuation process turns out to be much more difficult for health care than for other commodities, traded in private markets, because of the peculiar characteristics of health care and the institutional responses to them which typically remove the mechanism - commodity prices - that normally maintains the linkage between the resource cost of a commodity and its value to users.<sup>17</sup> Another difficulty is that the criteria for judging whether different allocations (distributions) of production and consumption - and costs and benefits - are 'better' or 'worse' are complex, value- and ethics-laden, and subject to dispute.<sup>18</sup>

These difficulties aside, however, it does not seem unreasonable conceptually to suggest that if the benefits of health care are perceived<sup>19</sup> by those receiving or paying for care to be smaller than the costs (i.e. other things foregone) of care, then there is "too much" care being produced and consumed (G).<sup>20</sup> This may be translated as too much health care in some aggregate sense or, more likely, too much of certain types of care (and perhaps not enough of others), i.e. the wrong mixture of health care outputs.

#### Health That is Valued At Less Than Its Cost

The primary benefit of health care and the reason that it is of value is the contribution it makes to health. But health as an output is also subject to the same criterion of "worth". Therefore another circumstance in which it may be possible for both individuals (H) and pop-

ulations (I) to have too much health care is that in which the health produced by health care and other resource-consuming determinants is not valued as highly as other outputs that could have been produced with the same resources.<sup>21</sup> In other words the mix of health versus other things (eg. education, security, justice, consumer goods) or the mix of different types of health, could be improved.

### Well-Being That is More Costly Than It Need Be

The suggestion that it is possible to have the 'wrong' mix of health versus other outputs that individuals or societies value implies that health is not the final output in the production process - or the final goal of public policy - after all. There must be some higher-order concept to which health and its competitors for resources make their ultimate contribution, for both individuals (J,L) and populations (K,M), and upon which the worth of the contributions is judged. This higher order concept may be termed "well-being" and is indeed the 'end of the line', for it is presumably not possible to have too much well-being!

Just as health care is but one determinant (input to the production) of health, so too is health only one of the determinants (inputs to the production) of well-being. It is therefore in principle possible to produce well-being too in more costly or less costly ways, depending upon the amounts of health and other inputs that are used and the respective costs of the inputs. It may therefore be possible to have too much health care if the level of health resulting from the care is too high, ie. if a lower level of health (facilitating a higher level of other contributors) would result in a higher level of well-being. In this case, well-being is not being produced cost-effectively.

### A Tentative Taxonomy

In Figure 3 we collect the ways in which there can be too much health care from the sections above and position them within a tentative taxonomy, defined (horizontally) by a progression from lower to higher level outputs and (vertically) by a progression from supply and cost to demand and value considerations.

The letters in the cells of Figure 3 correspond to those used in the sections above. For example, cell C (technical efficiency in the production of population health) corresponds to health care that is not effective (at the societal level); cell E (cost-effectiveness in the production of individual health) corresponds to health that is more costly than it need be; and cell G (allocative efficiency in the production of health care) corresponds to health care that is valued at less than its cost.

The three rows of the grid in Figure 3 correspond to the three levels of efficiency in economic theory - technical efficiency, cost-effectiveness and allocative efficiency. Although the grid may be very useful as a conceptual "sorting device"<sup>22</sup>, one should be careful not to draw overly fine distinctions among some of the cells. As the nature of the output changes the label of the cell may change; however, the fundamental economic problem - how to allocate scarce resources to competing uses in order to give society its most highly valued mix of outputs (outcomes) - does not change. Witness that allocative efficiency in the production and consumption of intermediate outputs (health care, health) is directly related to cost-effectiveness in the production of final outputs (health, well-being).

Much more important however, is that the grid in Figure 3 masks complex issues of equity in the distribution of both the benefits and costs of health care and other activities. Aggregation across individuals in a population is not simple in either theory or practice, because individuals are not identical (although much of economic theory assumes that they are) in their needs, resources or preferences. Therefore policies which may 'improve' perfor-

mance in the sense that they remove inefficiencies in some or all of the cells of Figure 3 are not unequivocally 'better'. That still depends upon a social judgement about the distribution of the gains and losses created by the reallocations.

### Health Care Spending and Economic Performance

Economic performance and health care spending are linked in several ways that reduce to two main causal paths. In general, the ability of a society to consume depends on its ability to produce; countries with productive economies can, and generally do, sustain higher levels of health care spending. Our interest in this section, however, is in the reverse causal path, i.e. the effect of health care spending on economic performance and prosperity.

Here we revisit and expand upon the final theme in the section above concerned with Health That Is More Costly Than It Need Be. This theme highlighted the opportunity cost of health care spending which might have otherwise found its way into economic growth, which in turn could both augment the capacity of the economy to sustain future health care spending, and improve health directly and/or indirectly. This is illustrated in Figure 4.

Exactly how health care spending affects economic performance and prosperity is a subject with more questions than answers, especially empirical ones. Nevertheless, it is possible to outline briefly the main arguments linking health care spending to prosperity, both positively and negatively.

On the positive side, it can be argued that health care spending represents an 'investment' in health which increases the capacity of the population to work and produce wealth, although the increasing concentration of health care on persons outside the labour force, such as the chronically ill or the very old weakens this argument.<sup>23</sup> Another positive effect of health care spending, it is sometimes argued, is that it creates beneficial economic activity. This usually means jobs and incomes which, although it may redistribute wealth, still begs the question of whether wealth has in fact been created. Perhaps if the jobs are associated with new products or services which are highly valued, then health care spending generates wealth; however, if the jobs are those traditionally associated with the health care sector - health care professionals and hospital employees - many would argue that such spending, although it might be termed "secondary" wealth creation<sup>24</sup>, represents a net drain on society's wealth.

On the negative side, the main argument is that among the "other activities" reduced by health care spending are savings and investment, which might have otherwise found their way into the research and development necessary for technologic innovation, which in turn is increasingly necessary for so-called "primary wealth creation" deemed critical in modern, global economies. In this view, health care spending is not only a wealth drain instead of gain, but also has a very high opportunity cost: it threatens the engine of growth which will provide the future capacity to sustain the social infrastructure (including, but not limited to, the health care system) essential to population health and well-being.

We look forward to discussion of these views, especially the latter one, at this conference, because our own limited examination of the research literature indicates that very little, if any, research exists which bears directly on the "health care spending - primary wealth creation" link, and its dynamic economic effects.<sup>25</sup> For example, what is known about the opportunity costs associated with the steady expansion of health care systems internationally until recently, or with their subsequent contractions in some European countries? What public or private activities were displaced or facilitated? And which research designs and methodologies might be used to investigate these questions? What proportion of the current level of health care spending contributes to wealth, either through enhanced worker productivity or through the development of new products in the tradeable goods and services sector, both on

average and at the margin? What is the relationship, if any, of health care spending to the competitiveness of a nation's firms, and through what micro-channels? It seems, at least to us, that there is ample scope for useful research on these and other related questions.

### Reflections on the Exercise

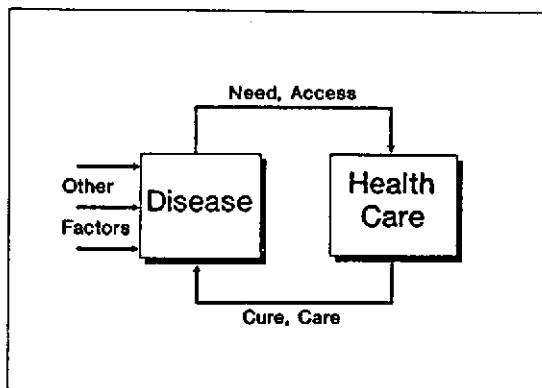
In this paper we have explored one of the admissible questions which a broad view of the determinants of health allows. We have sought to identify conceptually distinct ways in which individuals and societies might have "too much" health care and to illustrate these with examples where possible. It would perhaps be too strong to say that this exercise has yielded any new conclusions, but it has generated (at least for us) two principal observations.

First, it is possible to have too much health care, in one or more of several meaningful ways. Whether a society does have too much health care is a different, normative question; the answer to it depends both on the strength of the relevant evidence which exists for any particular cell depicted in Figure 3 and on social judgements about the equity of re-allocations involving less health care.

Second, the availability and strength of the evidence varies widely across cells. Some cells, such as A, B, C and D, are supported by a reasonably extensive empirical literature. Only a selection of examples contained in such cells have been referenced due to space constraints. Other cells suffer from a limited or non-existent supporting literature. For example, in cells E and F, there is little empirical research on the trade-offs between different mixes of inputs involving less of health care and more of "other things" which contribute to health. The inter-related cells of H, I, L and M, which bear on the trade-offs between health and "other things" which contribute to well-being, are also poorly researched. Yet these cells clearly represent the types of choices which societies make.

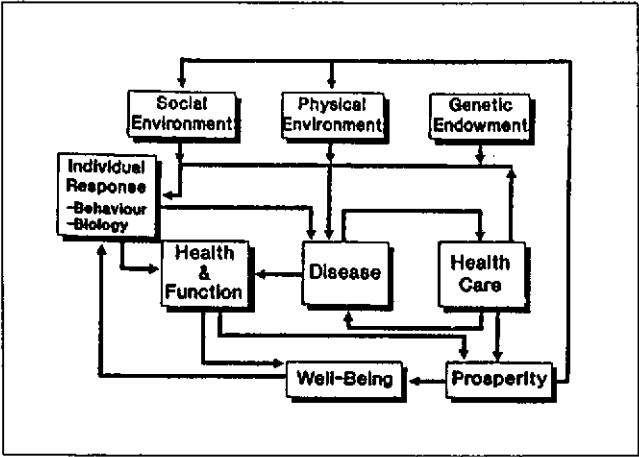
One of the most important cells (and one with perhaps the most relevance to this conference) has perhaps the least research of any cell. There is a major research agenda waiting to be set and executed on the effects of health care spending on economic performance.

Figure 1:<sup>1</sup>



Source: Robert G. Evans and Gregory L. Stoddart, "Producing Health, Consuming Health Care", *Social Science and Medicine* 31 (12) (1990): figure 1, page 1350.

Figure 2:<sup>2</sup>



Source: Evans and Stoddart, “Producing Health, Consuming Health Care”, figure 5, page 1359.

Figure 3: A Taxonomy

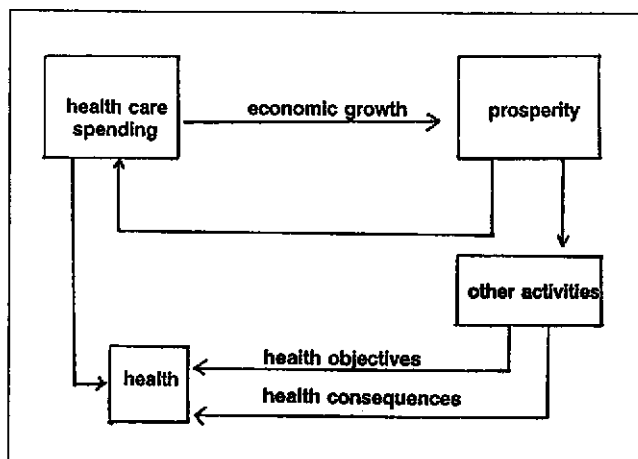
		Output				
		Health Care	Health		Well-Being	
			Individual	Population	Individual	Population
Efficiency	Technical <sup>3</sup>	A	B	C	J	K
	Cost-effectiveness <sup>4</sup>	D	E	F	L	M
	Allocative <sup>5</sup>	G	H	I		

<sup>3</sup> Technical inefficiency: more output could be produced with inputs

<sup>4</sup> Cost-ineffectiveness: more output could be produced with the

<sup>5</sup> Allocative inefficiency: a more highly valued level or mix of o the same resources

Figure 4:



#### Endnotes

- 1 Nowhere in the industrialized world is there a purely private system, although models of publicly financed and (largely) privately provided health care exist, like the Canadian system. The percentage of public financing varies considerably, however, from 36% and 42% in Turkey and the United States, respectively, to 90% or more in Sweden, Luxembourg and Norway in 1990. (Canada and Japan are somewhere in the middle of the OECD pack, at 73% and 72%, respectively). See George J. Schieber et al., "U.S. Health Expenditure Performance: An International Comparison and Data Update", *Health Care Financing Review* 13 (4) (1992): 1.
- 2 Robert G. Evans, "Health Care Reform: 'The Issue from Hell'", *Policy Options* (July-August, 1993): 38.
- 3 Robert G. Evans and Gregory L. Stoddart, "Producing Health, Consuming Health Care", *Social Science and Medicine* 31 (12) (1990): 1347-1363, and references therein.
- 4 Evans and Stoddart, "Producing Health, Consuming Health Care", 1349.
- 5 Evans and Stoddart, "Producing Health, Consuming Health Care", 1347.
- 6 Readers can ignore letters in parentheses for the moment. They are used in the taxonomy introduced below.
- 7 See Jan Blustein and Theodore R. Marmor, "Cutting Waste By Making Rules: Promises, Pitfalls, and Realistic Prospects", *University of Pennsylvania Law Review* 140(5) (May 1992): 1543-1572, for a taxonomy of waste that includes some of the elements in the broader taxonomy we develop below.
- 8 Hereafter referred to collectively as "services" or "interventions".

- 9 The harmful effects of health care interventions are termed iatrogenesis. The extent of clinical iatrogenesis, or harm at the level of the individual, has been estimated in the Harvard Medical Practice Study. Adverse events, defined as injuries caused by medical management, occurred in 3.7 percent of hospitalizations, with 2.6 percent of such events causing permanently disabling injuries and 13.6 percent leading to death (Troyen A. Brennan et al., "Incidence of Adverse Events and Negligence in Hospitalized Patients", New England Journal of Medicine 324 (6) (7 February 1991): 370-376. Ivan Illich has extended the discussion of iatrogenesis to include the harmful effects of health care at the level of society in Limits to Medicine (London: Penguin Books, 1976). Therein he describes social iatrogenesis as the directly health-denying effects of medicine's social organization. He posits the medicalization of prevention as a symptom of such iatrogenesis (p. 97). The personal responsibility of an individual for his or her future is transformed into the management of that individual by an agency. Illich also describes cultural iatrogenesis which denotes the transformation of pain, impairment and death from a personal challenge to a technical problem (requiring technical solutions from the health care sector).
- 10 The provision of ineffective health care is rarely an intentional act. Information on effectiveness is often lacking or not readily available to providers (or not acted upon when available). Iain Chalmers et al., in Effective Care in Pregnancy and Childbirth (Oxford: Oxford University Press, 1989), have identified that many obstetrical interventions require further evaluation or should be abandoned in light of the available evidence. Similar reviews of the available evidence on the effectiveness of clinical services have not been undertaken for other clinical specialties. Even with information in hand, however, the manner in which it is presented can affect treatment decisions. For example, physicians have been shown to indicate a stronger inclination to treat patients after reading of the relative (rather than absolute) change in the outcome rate. See Lachlan Forrow et al., "Absolutely Relative: How Research Results Are Summarized Can Affect Treatment Decisions", The American Journal of Medicine 92 (February 1992): 121-124. However, factors other than information on effectiveness (such as fear of litigation, or method of reimbursement) may influence treatment decisions. For example, A. Russell Localio et al. have demonstrated a positive association between malpractice claims risk and the rate of cesarean delivery in "Relationship Between Malpractice Claims and Cesarean Delivery", JAMA 269 (3) (1993): 366-373. The effectiveness of health care can also be considered at the societal level. Investigation of the association between health care services and (admittedly crude) measures of health outcome cast doubt on the effectiveness of health care services at the aggregate level (although there is some debate about this). See, for example, John B. McKinlay et al., "A Review of the Evidence Concerning the Impact of Medical Measures on Recent Mortality and Morbidity in the United States", International Journal of Health Services 19 (2) (1989): 181-208; and John M. Neutze and Harvey D. White, "What Contribution has Cardiac Surgery Made to the Decline in Mortality from Coronary Heart Disease", British Medical Journal 294 (14 February 1987): 405-409.
- 11 Inappropriate care is care provided to a patient who does not meet the indications for which a specific intervention has been shown to be effective. John N. Lavis and Geoffrey Anderson ("Hospitals in Canada: The Appropriateness of Setting and Services Provided", Work in progress for submission to the Queen's -University of Ottawa Economic Projects) have recently reviewed the international literature on the appropriateness of service provision. Restricting the discussion to studies wherein the panel of

experts involved in the rating of indications was drawn from the same country as patients receiving the procedure, ranges for inappropriate use were 4-27 percent for coronary angiography, 2-16 percent for coronary artery bypass surgery, 13-32 percent for carotid endarterectomy and 11-24 percent for gastrointestinal endoscopy.

- 12 Human resources may be employed in ways that yield lower output from the health care system than is technically possible. For example, Woolhandler et al. have shown that, because of duplication of administrative departments and personnel, U.S. hospital administrative costs are more than twice those in Canada ("Administrative Costs in U.S. Hospitals", New England Journal of Medicine 329 (6) (5 August 1993): 400-403). Facilities may also be deployed in ways that yield lower output than is technically possible. Hospitals may have low occupancy rates and laboratories may have long periods of down-time. Health policy reforms which reduce the costs of bureaucracy or reduce excess capacity could yield substantial savings or permit the production of a higher level of output.
- 13 Health care provision may be more costly than it need be if the same services could be provided by a less costly but equally effective health care provider. For example, if nurse practitioners were utilized in the Canadian health care system as fully as possible (given existing evidence of their safety and effectiveness), it has been estimated that savings would accrue in the range of 10-15 percent for medical services as a whole, and 16-24 percent for ambulatory services (Frank T. Denton et al., "Potential Savings from the Adoption of Nurse Practitioner Technology in the Canadian Health Care System", Socio-Economic Planning Sciences 17 (4) (1983): 199-209). The provision of health care may also be more costly than necessary if it is provided in a more capital-intensive setting than needed. Estimates of the extent of non-acute hospital days of care in the United States range from 20-48 percent for adults and 5-21 percent for children. These days of care could have been provided in nursing homes, ambulatory care facilities or other less capital-intensive settings (if available) (John N. Lavis and Geoffrey Anderson, "Hospitals in Canada: The Appropriateness of Setting and Services Provided", Work in progress for submission to the Queen's - University of Ottawa Economic Projects).
- 14 Consider, for example, a fixed pool of resources (\$1 million) and the years of life which could be bought with these resources. Spending the money on monitoring low-risk patients in coronary care units would buy 3 life years; providing propranolol for mild to moderate high blood pressure would buy 73 life years; and providing bypass surgery for left-main coronary artery disease in middle-aged men would buy 134 life years (Louise B. Russell, "Opportunity Costs in Modern Medicine", Health Affairs (Summer 1992): 164). A similar exercise could be performed using measures of health such as improvements in functional status or improvements in the quality of life. The point is the same. "More" health could be achieved with the same level of overall resource commitment but a different mix of inputs. "League" tables of the cost-effectiveness of alternative interventions are not without their difficulties, however. See, for example, Michael Drummond et al., "Cost-Effectiveness League Tables: More Harm Than Good?", Social Science and Medicine 37 (1) (1993): 33-40.
- 15 There are few examples of empirical research which extend the analysis of health-producing strategies beyond those within the formal health care system. Theodore Joyce et al. have compared the cost effectiveness of various health inputs (such as neonatal intensive care) and government programs (such as a supplemental food program for women,

infants and children) in reducing race-specific neonatal mortality or death in the first 27 days of life ("A Cost-Effectiveness Analysis of Strategies to Reduce Infant Mortality", Medical Care 26 (4) (April 1988): 348-360). The early initiation of prenatal care is the most cost-effective means of reducing the neonatal mortality rate with blacks deriving more benefit per dollar of input use than whites. Neonatal care, although one of the most effective means of reducing these rates, is one of the least cost-effective strategies. The need for further work on health-producing strategies beyond the formal health care system is suggested by international comparisons of total health expenditures (as a percentage of gross domestic product) and population health indicators (such as infant mortality rate or life expectancy). See, for example, George J. Schieber et al., "Health Spending, Delivery, and Outcomes In OECD Countries", Health Affairs (Summer 1993): 120-129. In 1991, Japan, Canada and the United States devoted 6.8, 10.0 and 13.2 percent of GDP to the health sector (respectively). Yet the pattern of infant mortality rates in 1990 is the reverse of that expected based on the expenditure data: 4.6, 6.8 and 9.1 per 1,000 live births in Japan, Canada and the United States (respectively). Similar gradients can be found for life expectancy at birth. Furthermore, life expectancy has increased rapidly in Japan over the past several decades compared to countries such as England (Michael G. Marmot and George Davey Smith, "Why Are the Japanese Living Longer", British Medical Journal 299 (23-30 December 1989): 1547-1551). This pattern is not particular to these four countries. J.P. Mackenbach ("Health Care Expenditure and Mortality from Amenable Conditions in the European Community", Health Policy 19 (1991): 245-255), in a study of European Community countries, was unable to show an association between a higher level of health care expenditure and a larger degree of success in eliminating mortality from preventable and curable conditions. Although this is far from definitive evidence, some countries (such as Japan) appear able to produce "more" health with a lower level of health care sector resource commitment.

- 16 Of course, if health care is not effective, as discussed above (B,C), it cannot be cost-effective. But cost-effectiveness in the production of health care is but one step along the road to cost-effectiveness in the production of health.
- 17 See Robert G. Evans, Strained Mercy: The Economics of Canadian Health Care, (Toronto: Butterworths, 1984), chapters 1-5.
- 18 See Anthony J. Culyer, "The Normative Economics of Health Care Finance and Provision", Oxford Review of Economic Policy 5 (1): 34-58 and Uwe E. Reinhardt, "Reflections on the Meaning of Efficiency: Can Efficiency Be Separated from Equity", Yale Law and Policy Review 10 (302) (1992): 302-315.
- 19 Or would be so perceived if the recipients/payers were fully informed. Another way of thinking of such "allocative efficiency" might be to imagine what fully informed individuals or societies might wish to 'spend' on health care, and compare it to the current pattern.
- 20 Examples of situations in which the benefits of health care are valued at less than its cost can be found both for individuals and populations. At the individual level, patients' and providers' valuations of benefits in relation to costs may differ and hence the decisions of fully-informed patients may differ from those arising out of a traditional patient-provider interaction. For example, John E. Wennberg has studied the effects of informed patient decision-making on the treatment of benign prostate gland enlargement ("Better

Policy to Promote the Evaluative Clinical Sciences", Quality Assurance in Health Care 2 (1): 21-29). Patients differed in how they were bothered by their symptoms and their willingness to undertake the risk of operation. Preliminary results have shown that, among patients who used interactive videodiscs which explained treatment options and associated risks, 25 percent of those who initially preferred surgical removal of the prostate gland (after discussion with their physician) changed their mind and opted for "watchful waiting" ("Educational Videodiscs Give Patients Information to Make the Right Choice", Medical Post, 16 March 1993). Although several studies available suggest that fully-informed patients choose less health care, it is not clear that enhancing the role of patients in decision-making will always lead to less health care (see, for example, Mark N. Levine et al., "A Bedside Decision Instrument to Elicit a Patient's Preference Concerning Adjuvant Chemotherapy for Breast Cancer", Annals of Internal Medicine 117 (1) (1 July 1992): 53-58). Research on the elicitation of patient preferences for lung cancer therapies has shown that the manner in which information is presented can affect the valuation of the net benefits of treatment. This is analogous to the framing effects demonstrated with physicians and discussed in endnote 10. For example, the attractiveness of surgery, relative to radiation therapy, was substantially greater when the problem was framed in terms of living rather than dying (Barbara J. McNeil et al., "On the Elicitation of Preferences for Alternative Therapies", New England Journal of Medicine 306 (21) (27 May 1982): 1259-1262). (In addition to ensuring that the benefits of health care are perceived by those receiving care as greater than the costs, the involvement of patients in their care has been shown to improve patient outcomes. See, for example, Sheldon Greenfield et al., "Expanding Patient Involvement in Care: Effects on Patient Outcomes", Annals of Internal Medicine 102 (1985): 520-528.) Those paying for care also have reason to ensure that the benefits of health care are greater than the costs. Oregon's effort to explicitly ration the medical care resources it is prepared to allocate to the poor provides an example of an attempt to systematically solicit and apply citizens' values in a prioritization process. See Daniel M. Fox and Howard M. Leichter, "Rationing Care in Oregon: The New Accountability", Health Affairs (Summer 1991): 21 for a description of the process used.

- 21 As with the valuation of health care (in terms of its ability to produce a desired health state), the health that is produced by health care and other resource-consuming determinants must also be valued relative to its cost. At the individual level, such valuations may, for example, take the form of advance directives which specify the medical care which an individual wishes to receive when they are no longer competent. Linda L. Emanuel et al. ("Advance Directives for Medical Care — A Case for Greater Use", New England Journal of Medicine 324 (13) (28 March 1991): 889-895) surveyed outpatients of primary care physicians and found that 93 percent desired advance directives. These individuals refused life-sustaining treatments in the event of a persistent vegetative state (85 percent) or dementia with a terminal illness (87 percent). Hence these health states were valued less than the (non-monetary) cost associated with life-sustaining treatment. Despite such valuations, advance directives are not always complied with. Marion Danis et al. ("A Prospective Study of Advance Directives for Life-Sustaining Care", New England Journal of Medicine 324 (13) (28 March 1991): 882-888) demonstrated that actual care was inconsistent with advance directives 25 percent of the time and more aggressive than requested 6 percent of the time.

- 22 Cutler, "The Normative Economics of Health Care Finance and Provision".

- 23 Evans and Stoddart, "Producing Health, Consuming Health Care".
- 24 J. Fraser Mustard, "Background on Key Concepts", Ontario 2002: A Report of the Task Force to Review the Ontario Technology Fund in the Context of an Innovation-Based Society to the Premier's Council on Economic Renewal, (Toronto: Premier's Council on Economic Renewal, 1993): 90-96.
- 25 Or, for that matter, on what social arrangements might best protect against the disruptive or destructive effects of growth as it creates gainers and losers in a society.

# Health Care As a Threat to Health: Defence, Opulence, and the Social Environment

---

Robert G. Evans

## What "Good" is Health Care?

Very few people own tanks (battle) for personal enjoyment. Indeed very few, other than psychopaths and small boys, have any desire to do so. Such vehicles are not offered as prizes in lotteries or advertising programs, or featured in the lifestyles of the rich and famous. New homes, BMWs, tropical vacations, yes, but not tanks. And yet, under some circumstances, one might literally die, if not for one, then at least for lack of one.

The parallel with heart-lung transplants should be apparent. The services of the health care system, like those of the military, are not wanted for their own sake. When they are called on, they are needed, not wanted, and the need may be a matter of life or death. But in both cases, while having the services at need is much better than not having them, it is far better still not to need them at all. The most sophisticated and effective health care in the world cannot produce results as good as simply remaining healthy in the first place. This much is obvious to anyone who has ever been a patient – or been in battle – and our common understanding is pretty clear on the value of both health and peace.

Yet this understanding becomes clouded when we consider health care policy, because most health care services (unlike those of the military) are personal services provided to specific, identifiable individuals. By a progressive elision, or modification of the metaphor, the observation of a discrete commodity or service provided to/utilized by an individual can be interpreted as a "good" being "demanded" by a "consumer" rather than an intervention being undergone by a patient. The overtones of uncertainty, anxiety, and suffering which accompany the experience of health care are replaced by the confidence and satisfaction of the knowledgeable (and affluent) buyer making choices in a well-stocked market.

In introducing this alternative set of terms, with their heavy load of (untested, rather implausible) behavioural assumptions and (personal, rather idiosyncratic) values, the naive economic analyst can dissolve the difference between "needing" a surgical procedure or a diagnostic intervention (or long-term care for the cognitively impaired), and "demanding" an ice-cream cone or a new pair of shoes. The same analytic apparatus is then applied to each type of transaction; one should not be surprised when similar policy recommendations emerge.

But as the *Economist* newspaper said about theories of economic growth: "It takes many years of training to ignore the obvious." Most people do not have that training, and may not have the intellectual capacity to absorb it in any case. Thus they have not lost their grasp on the obvious, and understand that health care is not a "good" in the same sense as a hamburger. When we hear of two different people, one having just "consumed" an expensive vacation in the south seas, and the other having "consumed" an equally expensive episode of hospital care, we all know that the first event is to be celebrated (or envied) and the second to be commiserated (or pitied), even if one paid the full cost of her vacation, and the other received

her large allocation of "goods" for free. A successful outcome for the second is to be celebrated, given the prior circumstances necessitating the care, but not the episode in total.

### Collective Financing and Collective Priorities

This broad public understanding is reflected in the fact that all countries in the developed world fund health care collectively, and almost all fund it through public or quasi-public agencies. Table 1 shows, for the countries of the Organization for Economic Cooperation and Development (OECD) the proportion of total health care funding that comes through the public sector, defined broadly to include "social insurance".

TABLE 1

International Comparison of Public Share of Health Care Spending		
	<u>1980</u>	<u>1990</u>
Australia	63%	68%
Austria	69%	67%
Belgium	83%	83%
Canada	75%	73%
Denmark	85%	83%
Finland	79%	81%
France	79%	74%
Germany	75%	73%
Greece	82%	76%
Iceland	88%	87%
Ireland	82%	75%
Italy	81%	76%
Japan	71%	72%
Luxembourg	93%	91%
Netherlands	75%	71%
New Zealand	84%	82%
Norway	98%	85%
Portugal	72%	62%
Spain	80%	78%
Sweden	93%	90%
Switzerland	68%	68%
Turkey	27%	36%
United Kingdom	90%	84%
United States	42% (?)	42% (48%)*
*adjusted to include tax subsidy		

SOURCE: G.J. Schieber, J.-P. Poullier, and L.M. Greenwald "U.S. health expenditure performance: An international comparison and data update" Health Care Financing Review 13:4 (Summer, 1992)

Even the exception, the United States, the home of private finance, is less exceptional than appears at first look. Although nearly two-thirds of the population rely on private insurance for their primary coverage, this only covers about one-third of total expenditures, compared with forty percent from public sources. And even that private coverage depends on heavy public subsidies through the income tax system. If these are netted out, "private" coverage raises only about one dollar in four, compared with roughly half from governments.

The particular form of collective financing makes a great deal of difference to the distribution of the total burden among the population – to who pays for the system, and how much. When health care systems are funded from general tax revenue, people at higher incomes pay a significantly greater share of the overall costs than when they are funded through specially earmarked "social insurance premiums". Private insurance, along with user-pay or co-insurance charges, takes a particularly large proportion of the incomes of lower income people. Moreover, systems which rely more heavily on user payments and on multiple forms of private insurance also offer higher income people better access to services. Compared with tax financed systems, the more wealthy pay less and get more. But regardless of how the load is distributed across the population, the predominance of collective funding is universal in the developed world.

The issue of who pays, and how, rarely arises in the financing of military services; it is the personal linkage and the misleading labelling of health care as a "good" that supports continuing debates over health care finance. The fact that different ways of organizing the collective funding of health care have such very different distributional consequences, and that changes can have the effect of transferring substantial amounts of money from one set of pockets to another, assures that conflict over the forms of funding will never be far from the top of the public agenda in most developed societies.

Moreover it is not difficult to see that a very small proportion of the population – those unfortunate enough to be quite ill – account for most of health care spending. It is for this reason that all systems rely on collective funding. In the modern world there is no possibility that those who use the services could ever pay a significant share of the cost from their own resources, even if we thought that that was fair.

Private insurance, operating on true insurance principles and without heavy public subsidy, is equally incapable of covering more than a small share of the overall expenditures, though private insurers will happily enrol large numbers of people with minimal needs. The fatal flaw in the private insurance approach is precisely that they are able to do so, because use is not only very unevenly distributed over the population, but also to a large extent predictable. At the same time, however, the fact that most people use very few services, and a few use most, further highlights the large interpersonal transfers of resources that are an inevitable part of any modern health care system.

By contrast, it is difficult to say exactly who in a society most benefits from military services, so awkward questions as to who is paying for whom do not arise in so transparent a form. Correspondingly, there is little controversy either over funding from general tax revenue, or over the necessity of setting collective, "national" priorities to determine how much to spend in the pursuit of national military objectives, and how to spend it. That is not to say that this collective decision-making process is not highly contentious, or subject to extreme uncertainty. But it is generally accepted to be properly a collective, political debate, without the intrusion of the individualistic rhetoric that so confuses public debates over health care policy.

### **Who Are the Beneficiaries? Outputs and Inputs**

In both the military and in medicine, benefits (and costs) come in several different forms.

There is a particular confusion between those who benefit from the services per se – being protected from invasion or cured of illness, for example – and those who benefit from the opportunity to provide the services themselves – the defense contractor or drug company shareholder. Indeed it is very often argued that everyone gains, collectively, from the expansion of such opportunities, through the creation of jobs, whether they be in the military, or the medical, or any other sector. Particularly in times of high unemployment, as now, one finds health care celebrated as a growth sector even while it is being criticized for absorbing an ever larger share of national resources. The financial press is particularly prone to this confusion.

Economists tend to have a clearer grasp of the distinction. No one has ever improved on Keynes' example, that if the Treasury were to bury bottles full of banknotes (secretly, at night) in various parts of the realm, a highly profitable private employment could be created as people went out to look for them and dig them up. Jobs would be created. Yet no one would imagine that the average real output, or income, or well-being, of the United Kingdom would be increased as a result. The gains of successful finders, their increased claim on national output, would be offset by the reduced share of those who did not find bottles. Indeed if the bottle-finding industry absorbed, not only the unemployed, but some who would otherwise have been producing commodities of real value, then average incomes in real terms would in fact fall. It is the output that matters, that creates value or wealth. If there is no (useful) output, the job or activity is not of value in itself.

(The follow-on impact of an increase in the money supply is another and much more complex matter. But whatever the effects of monetary policy, they do not depend upon whether or not people are also digging holes in the ground.)

The same idea is expressed in the more recent phrase "real jobs". The collateral suggestion that "real jobs" are found only in the private sector, and "unreal jobs" in the public, is ideological nonsense; the American health care financing system, for example, currently wasting about \$100 billion a year in activity without (net) output, is probably the largest single concentration of "unreal jobs" in the world. But the distinction is real enough. To maintain living standards, a society must concentrate its scarce resources in activities that create wealth, not waste motion.

### Wealth Through Production, Not Destruction

Again the military analogy may help to clarify the point. Imagine two countries engaged in a major war that, to separate the human factors from the economic, is being fought only with and against "smart" weapons, not human beings. National income accountants on both sides dutifully record the production of various weapons platforms and ordnance in the Gross Domestic Product; these are then destroyed in battle the next week. Employment and output in each country are at an all-time high, as are recorded incomes. But is this prosperity? Keynes' banknotes in bottles have been updated to the ultimate video game.

The reality is that, on average, citizens in both countries are receiving a large part of their real "incomes" in the ephemeral form of military equipment. To reflect the rather limited value to them of this type of income, particularly in its post-combat form, their monetary incomes must be reduced by a corresponding amount. This can be done through contemporary taxation, but typically wars are financed in part at least through borrowing. Inflation then or subsequently reduces the value of the loan, resulting in de facto taxation of the lenders.

One way or another, the diversion of resources into "products" that are not valued in themselves (and in any case do not last) must be matched by a reduction in the availability of commodities that people do value, either for consumption or as productive investments. The

elementary economics texts often refer to a choice between “guns or butter” – more of one, less of the other. And the costs of military spending may not just be reductions in current consumption. If resources are diverted away from maintenance, repair, and up-grading of the nation’s capital stock (except for that needed for war production), long-term economic growth will suffer.

This is not to argue that the military activity itself is wrong-headed; it may be eminently sensible policy for one or even both countries. But the “income” and “output” generated in the process do not create wealth or prosperity. Furthermore, given the economic costs of the war, rational people on both sides might well ask whether the ultimate objectives of peace, security, or national fulfilment might possibly be pursued in a less costly manner by other means. Recall again that the production and operation of weapons have (for most of us) little or no value in themselves, but are means to these real, though poorly defined, ends. (The rational contemplation of alternative means, however, may be more possible in the bloodless context of a super video game. When people are dying, further killing quickly becomes an end in itself.)

### III-Health Is Not Wealth

To bridge to the health care system, consider a nation faced with a really serious epidemic. [An HIV mutation is transmissible through airborne droplets; Bovine Spongiform Encephalopathy (“mad cow disease”) jumps to humans on a large scale; multi-drug resistant tuberculosis becomes widespread.....] If the problem is severe enough – on the scale of the Black Death, unlikely but not impossible – we would presumably see a massive transfer of resources into the health care sector for treatment, prevention, and research. One might predict that health care cost containment would rapidly lose its position on most national agendas.

This major expansion of activity in the health care sector would all be recorded as national income and output. And, assuming that the resources were being deployed in a rational manner, given the available information — as in a war, information is always incomplete and you try to make the best guesses you can — presumably most of us would accept that the national effort was “needed”. But as in a war, no one should be confused by the economic statistics or by the increased health care activity into believing that we had become wealthier or healthier. Such an epidemic would make us unambiguously worse off, on both counts.

Indeed the numbers can be even more deceptive. Suppose the epidemic tended to carry off the very old and the very young – not an uncommon situation – and that treatment, despite heroic efforts, was not very effective for these. The net effect of the epidemic would be an increase in per capita output and income. The less effective the treatment, and the heavier the death toll, the greater the increase.

Such an epidemic is, of course, purely hypothetical, although the scenario is a good deal less far-fetched than it would have seemed twenty-five years ago. In developed societies the health care sector devotes a very small share of its activity and resources to dealing with infectious disease; even the extraordinary phenomenon of HIV/AIDS is highly localized in its impact. Attempts to recruit AIDS treatment as an explanatory factor in the escalation of health care costs in the United States have no empirical support, and are merely part of a broader effort to divert attention from the real forces at work.

Nevertheless, the hypothetical epidemic highlights characteristics of the broader health sector. It illustrates very clearly the status of health care as not a good in itself, but as a means, sometimes powerful but often imperfect, to the real end of health. There are certainly circumstances in which health care, like military preparedness, is desperately needed. Health care is usually needed only by individuals, but as in the case of the (we hope) hypothetical

epidemic it may be needed by the whole society as well. But those circumstances are defined by the existence of threats to health, against which care is effective; while we may need care, we do not want (to need) it.

### Collective Coping with Threats – Quis Custodiet...?

Health care systems, like military establishments, embody our collective responses to perceptions of need, whether in the form of threats to health, or threats to national security. These threats are interpreted, and responses formulated, by complex and highly professionalized establishments in each case. And the status, incomes, and professional self-fulfilment of the people in those establishments depend upon the level of collective support they receive from the rest of us, which in turn depends upon our perceptions of the seriousness of the threats, and of the effectiveness of the responses which they can offer.

It is no accident that the self-perpetuating "military-industrial complex" that President Eisenhower identified and warned against just before leaving office, was so quickly transposed as the "medical-industrial complex" – the large, interdependent collection of professional and commercial interests, people and institutions, who offer the rest of us protection against threats to health. In return, they require that we tithe – taking from us between 8% and 10% of national incomes in most developed societies (and cutting much deeper in the United States, now over 14%).

Health, like peace and security, is well worth "defending". But the same questions arise in each case:

- i) How serious is the threat?
- ii) How effective are the responses offered?
- iii) How efficiently are those responses being mounted?
- iv) Are there other ways we could try to respond?
- v) What are we giving up, in the process?

These questions do not bear on nonsense issues such as: "Should we or should we not, as a society, support a health care system/defense establishment?" Of course we will. The real policy issues are: "How large should it be?" and "What do we want it to do?" When such issues are left to the professionals themselves, the answers are: "Somewhat bigger," and "Whatever we think appropriate". Implicitly the answers that they offer to the five questions above, are: i)-iii) "Very", iv) "No", and v) "Nothing of comparable value."

Well, the threats are real enough, and some of the responses are outstandingly effective. Few would be so foolish as to deny the powerful contribution that health care can make, in particular circumstances, to the health and well-being of individuals and whole communities. But there are also egregious examples of ineffective and inefficient responses, in every modern health care system, detailed for example among the various ways in which one can have "too much" medical care, in the companion paper for this session. These illustrate yet again the force of the ancient question, "Quis custodiet custodiet ipsos?" Who keeps the professionals up to scratch? They cannot do it themselves.

### Grasping Shadows, Losing Substance?

But question i) cuts deeper. What is the threat? Death? disability? pain and suffering? In some absolute sense death per se cannot be a meaningful threat, since it is simply an ineluctable consequence of life. "Mortality is the price of identity." Premature, agonizing, terrifying death – these are threats. But death as such just is. As Eliot's Becket says: "...I am not in danger; only near to death." Similarly functional disability goes with age – I cannot do now some of the things I did at twenty. The process of deterioration can be restrained,

and some deficits repaired or compensated – for a time. Some capacities improve with age – again for a time. But perfect health, indefinitely maintained, like perfect security, is not on offer in this life at least. “Here is no continuing city.”

If we misread the possible, we misperceive the threat. The result can be, often is, wholly inappropriate responses – like pouring water into a sieve. Let no one depart this life with unbroken ribs, for example, or with an orifice un-intubated. During life, any discrepancy from the ideal must be met with some intervention, regardless of the chances of success. Raising and contributing the necessary resources is then presented as an ethical obligation on the rest of society – but whose ethics demand this? What is the ground of obligation?

At some level, we all understand the foolishness of this, just as we understand that increased defense spending, if it triggers off an arms race, can easily reduce national security. Indeed we now have a further ironic twist. One interpretation offered for the highly-publicized American arms build-up of the early 1980s is that, whatever the underlying reality, it forced the Soviet Union to try to keep pace, thereby over-straining an already weak economy to the point that the regime collapsed, eliminating the Great Adversary. Clearly building ever more tanks did not make the U.S.S.R. more secure; the world's largest tank armies, never committed to battle, are now a major source of high-grade steel scrap.

But are the United States and its allies now more secure? That seems far from clear. Apart from the nation, are individual Americans more secure now than ten or twenty years ago, either economically or physically? And what was the cost, in terms of lost opportunities to address the profound problems facing American society? Which brings us back to questions iv) and v), and health care.

### **The Determinants of Health: More Ways to Skin a Cat**

The understanding that health, at both the individual and the population levels, is powerfully influenced by factors other than health care per se is very ancient. It has been somewhat eclipsed, over the past century, by the extraordinary effect of the application of science and technology to medicine, and the wide distribution of the results among the populations of developed societies. But it is now enjoying a considerable resurgence, not only because of the development of a more measured and critical view of the contributions of health care itself – more careful evaluation, less cheerleading – but also because of the convergence of new research findings bearing on the determinants of health. Some of these are coming from sociology and psychology, always a bit sceptical of the claims of “mainstream” medicine. But surprisingly supportive evidence is also emerging from epidemiology, molecular biology, neurophysiology, and even primatology.

Possible biological foundations are beginning to emerge for the widely observed correlation between prosperity and health, a correlation which includes but goes far beyond the obvious connection between poverty and deprivation, and corresponding physiological damage. There is extensive evidence for a correlation, within populations, between health and hierarchical position. This may reflect the effects on health of differences in material circumstances, education, status, or sense of control over one's environment. If the relationships between health and “other things” – prosperity, self-esteem, sense of control – within populations also underlie the differences between populations (either different societies at the same time, or “the same” society at different times), then quite conceivably economic growth and development in itself is a powerful contributor to health.

If so, the connection is not simple. At a very low level of development, economic growth may mean more and better food, housing, clothing – (relatively) well-understood pathways to better health. But in more highly developed societies, the connections may be more subtle. Is it prosperity per se that contributes to health, or is that a marker for other forms of

progress, in the power of individuals and groups to control and shape their environments? Economic output is an indicator of what a population can do. But it is not the only one. The ability to do is a powerful source of self-esteem; but again, not the only one. So the pathways are not clear. But it is clear that such pathways exist. The connections between health and social environment, broadly defined, are very real and may be very powerful.

### **Overshooting the Mark**

This raises the possibility that over-investment in activities which are not themselves directly valued – such as medicine and the military – could have negative effects on health through the “opportunity costs” which they impose through the rest of the social environment. But again, the linkage may be complex.

The withdrawal of resources from more productive activities, to build tanks and aircraft which then destroy, and are destroyed by, someone else’s tanks and aircraft, is not difficult to perceive as a loss of national wealth. But suppose we (think we are) winning? The sense of self-esteem and control may be enhanced, even while the standard of living is being reduced. If it is the psychological state, rather than the material conditions, which are most important for health (at least above some minimum standard) then health may be improved, even though prosperity is reduced. The perception of victory probably cannot be indefinitely sustained on both sides, however, and even the victors must count the cost on the morning after.

The transfer of resources to the health care system is less clear-cut in its effects, because health care may generate significant benefits both as consumption – improving the quality as well as the length of life – and as investment – restoring people to productive roles in society. But the efforts to contain the expansion of health care systems, which are virtually universal throughout the developed world, are prima facie evidence that the members of those societies, or at least their political representatives, do not regard the benefits, at the margin, as worth the costs. Particularly in the United States, one hears increasingly the concern that excessive health care costs are weakening the nation’s ability to compete in world markets, thus threatening the prosperity (and perhaps the health) of its population.

### **Easy Answers Do Not Lighten the Load**

At a superficial level, economic analysis has a ready response. A nation cannot “price itself out of world markets” in general. If domestic costs of production go up, for whatever reason – higher wages or higher health care costs – a downward adjustment in its exchange rate can always restore the balance. And if American workers choose to take more of their wages in the form of health care benefits, and less as take-home pay, why should that be a source of public concern? If health care is a “good”, then “consumers” are simply substituting one “good” for another, presumably because that gives them greater satisfaction. No problem.

But this superficial response is weak at two points. If, as seems self-evident, health care is not a “good” in the usual sense, then the transfer of an increasing share of national income to its purchase, particularly in the absence of any observable off-setting increases in health, will be quite correctly perceived by workers, and payers of insurance premiums and taxes, as a fall in their real incomes. “Too much” health care, in one or other of the senses described in the companion paper to this, at “too high” a price, is being produced, through processes over which the individual, the employer, or the insurer has in practice little or no control. Hence the distress.

The threat to international competitiveness then arises as a consequence of people’s efforts to maintain their real standard of living, net of increasing health care costs, whether

those costs are channelled through taxes, payroll deductions, or premiums. If these efforts result in lower rates of saving – expressed in lower personal savings, reduced corporate retained earnings, and larger government debt – this in turn reduces the resources available for investment in physical plant and in research and development.

In wartime, patriotic propaganda (or fear) may encourage people to accept higher taxes, and voluntarily to increase savings (buying Victory Bonds), while rationing can be used to reduce the availability of consumer goods. But no such mechanisms can be used to win acceptance of an easy transfer of resources from personal consumption to health care.

The “threat to competitiveness” then arises not because a country prices itself out of world markets, but because its products are no longer competitive. No one wants products that embody obsolete technology. The second flaw in the superficial argument is that it implicitly assumes a fixed set of commodities and an unchanging technology – a static world, not a dynamic one. Only prices change, so that any increase in domestic price levels can be cancelled out by an adjustment in exchange rates. But the world we live in is dynamic. It is true – it has to be – that eventually exchange rate adjustment (or default) will bring a country's international accounts into balance. But in the process, its capacity to create wealth may have been severely damaged.

### Old Ideas, New Evidence and Applications

The argument that too much health care may be a threat to health, just as too much military effort may be a threat to national security, links together very old and very new ideas in both biology and economics. The ancient understanding that how we live has profound effects on our health is being substantiated and made much more specific – although also much more complex – by recent advances in both the biological and the social sciences. If an over-developed health care system drains away resources that might otherwise have improved the social environment, then it could be a threat to health.

At the same time, the most recent attempts to understand economic growth – the advancement of the wealth of nations – is reviving the concerns of Adam Smith and the Classical economists who wrestled with the same problem. Their distinction between “productive” and “unproductive” labour (dismissed as wrong-headed by the marginalists over a century ago) points towards the modern concern with wealth-creating versus wealth-dissipating activities.

Quite obviously, activities which yield no output, or output that is of no value, are wealth-dissipating. Not all commodities are “goods” – it depends upon the circumstances. But even valued outputs may be “unproductive” if they tend to reduce the capital base of a society. Modern theories of economic growth focus more on the importance of intangible capital – “know-how”, or “science-based innovation” – than on physical capital like canals, railways, or factories. But know-how is commonly embodied in, and indeed arises from, investments in new physical capital as well.

Adam Smith would not have been surprised to find that a society devoting a relatively high proportion of its economic resources to “unproductive” activities had a lower rate of growth as a result. He was not a pacifist; indeed he pointed out that “Defence is more important than opulence.” and he would presumably have included defence against non-human threats to life and limb. But the associated activities, however necessary, do not contribute to economic growth.

The linkage of national with biological defence suggests a more modern metaphor for the argument sketched out in this paper. Each of us is defended against external threats by an immune system whose extraordinary complexity and sophistication is only beginning to be revealed. Without it, we could not survive. But that system can in a variety of ways be stim-

ulated to over-react to threats, or to react to non-existent threats, in such a way as to become itself a threat to health. Both the military and the health care system represent important parts of the social, or population level, "immune system".

But the social immune systems, like those of the individual organism, can go out of balance and react inappropriately to real or imaginary threats. Autoimmune disease in the individual has its parallel in the military coup. Allergies arising from an over-zealous immune system can range from minor irritations to life-threatening emergencies. The steady pressure for expansion which one sees within health care systems throughout the developed world, and which each society struggles with greater or lesser success to contain, may likewise be thought of as an over-zealous social immune system. That system is still essential to health. But in its efforts to respond ever more forcefully to a wider and wider range of perceived "threats", it has itself become a threat. This social "allergic response" is at its most extreme and most painful in the United States, where it is also rapidly growing more severe. But the discomfort is widely shared – we are all scratching and sniffing over health care.

# Discussion of Papers by Naoki Ikegami, Gregory Stoddart/John Lavis and Robert Evans: Health Care and the Economy

---

*Jonathan Lomas*

**T**he common theme that flows through each of these papers is "market share" and the mechanics underlying its maintenance. Individuals and organizations compete within sectors and collaborate against outside sectors to maintain and increase their market share. Because of the truism that resources are finite, an increase in one's market share leads to a decrease in another's.

Dr. Ikegami addresses the reactions of physicians to competition, and therefore threatened loss of market share, within the sector of Japanese health care. Drs Stoddart and Lavis draw attention to some of the targets used in the battles for market share both within health care – for instance, effective health care that is more costly than needs be – and between health care and other sectors of the economy – for instance, does health care spending adversely affect economic growth? Finally, Dr. Evans, using the analogy of the military, addresses the techniques used by an entire sector to combat assaults from other sectors on its rightful share of the larger economic pie. He points out that need for weapons (or health care) is an infinitely elastic concept that can be manipulated by encouraging the population to confuse needs and wants. The confusion can serve to fuel expansion even in the face of no real increase in needs.

Dr. Ikegami points out that the powerful private practitioners in Japan will collaborate on reform of long-term care (not a significant part of their current market share) but obstruct major revision of payments for ambulatory care (the core of their current market share). By tracing the historical roots of the private practitioners' power, Dr. Ikegami also reminds us about the importance of history. The Japanese Medical Association, under the entrenched influence of the private practitioners, can only accommodate to incremental change with budget neutrality. While the nature of the constraint from history varies across countries, the fact of history as a constraint does not.

The paper by Drs Stoddart and Lavis not only addresses some of the targets used in the battle for market share within and between sectors, but also points up the importance of framing the debate accurately. It was Henry Kissinger who said that "about 90 percent of any policy issue was determining the intellectual framework within which the issue was discussed. Only 10 percent was working at the details of the policy and implementation." Stoddart and Lavis highlight that the "appropriate" market share (for medicine in health care or for health care in the larger economy) will depend upon the criteria one adopts for valuation. They have designed a comprehensive map of the potential valuation criteria that may prove to be a powerful "intellectual framework within which the issues can be discussed".

Much of Dr. Evans' paper is a disposition on how various myths can be created and sustained with little effort when maintenance or expansion of a sector's market share is at stake. The American Medical Association can set aside their differences with the Health Insurance Association of America on micro-management of medical care to oppose health care reform that might actually slow the inexorable growth in American health care spending. As I

sought to summarize his paper I was reminded of Robert Heilbroner's observation that "advertising is the propaganda of the private sector and propaganda is the advertising of the public sector". Evans argues persuasively that health care, and particularly medical care, uses propaganda (read advertising) very effectively to maintain market share, independent of the need for health care.

Finally, I was struck in reading these papers by their avoidance of the relationship between well-being and either health care or economic growth. My own roots in psychology lead me to interpret data on health inequalities that are not attributable to deprivation (e.g. the "Whitehall" studies) as suggesting that poorer health may be due to perceptions of relative status. Perception of disadvantage may be the crucial variable in determining lack of well-being and, subsequently, poor health. Such perceptions are driven by one's status relative to one's immediate (work, living, family) environment.

If, in fact, well-being is seen as our ultimate objective (isn't that why we want economic growth, we don't want it just for itself, do we?) and well-being is based on perceptions, what is the relationship between well-being and health care?

At the level of the individual, health care can undoubtedly improve well-being while leaving health unaffected. For instance, the placebo effect is based upon generating perceptions of benefit in the absence of demonstrable gains in health. Ruling out a feared diagnosis leaves health unaffected but probably does much to improve well-being. (To be fair, I should also note that there are examples where health is unaffected but well being is reduced by health care – labelling someone with a disease for which we can do nothing, for instance.)

At the level of the community one can trace a similar impact of health care on well-being while it leaves health unaffected. The classic example is the community hospital which, whether or not it actually affects the community's health, confers a sense of community well-being. Attempts to close such a hospital, under the judgement that it is not "needed" to maintain health, come up hard against the perception of the community that it is needed as a major contributor to social cohesion, local identity and, ironically, the community's capacity to attract new industry for economic growth. The latter are likely significant contributors to well-being.

If, therefore, we accept that maintaining and improving our well-being is our ultimate aim (rather than economic growth for its own sake), the nature of the question may change: "At what point does the level of funding and the provision of health care balance a) its net positive impact on improved (perceived) well-being and b) its net adverse impact on the capacity of society to deliver other things that improve well-being?" In answering this question it will not be enough to focus only on the clinical epidemiologists' judgements of "effective care". To these judgements we must add the individual's and the community's perceptions of benefit. If indeed well being is a matter of perception then we must either alter what we deem as misguided perceptions about the value of health care, or we must live with the impressions created by unanswered and effective "propaganda" on the merits of health care. We cannot on the one hand claim that well-being is our ultimate objective, but on the other hand ignore that the current well-being of many individuals and communities is inextricably tied to perceived benefits from much health care that is, by clinical epidemiologists' standards, ineffective.

# Discussion of Papers by Naoki Ikegami, Gregory Stoddart/John Lavis and Robert Evans: Health Care and the Economy: A Paradoxical Relationship

---

André-Pierre Contandriopoulos

All three lecturers consider the efficiency of the health care system to be at the heart of the debate on the relationship between health care spending and economic prosperity.

Comparing the health care system with the military system, Evans argue that: *"To maintain living standards, a society must concentrate its scarce resources in activities that create wealth not waste motion."* (p.7)... *"Quite obviously, activities which yield no output, or output that is of no value, are wealth-dissipating... even valued outputs may be 'unproductive' if they tend to reduce the capital base of the society"* (p. 21)... *"This raises the possibility that over investment in activities which are not themselves directly valued such as medicine and the military could have negative effects on health through the 'opportunity cost' which they impose through the rest of the social environment"* (p. 17).

Stoddart and Lavis inquire: *"Can we have too much health care?"* (p. 3) ... *"The benefits of health care must be valued in relation to its costs"* (p. 9)... *"This theme highlighted the opportunity cost of health care spending which might have otherwise found its way into economics growth, which in turn could both augment the capacity of the economy to sustain future health care spending and improve health and well-being directly and/or indirectly"* (p. 14).

Ikegami describes how Japan managed to control the costs of its health care system. His discussion is based on the assumption that : ... *"It is difficult to evaluate whether the health care services that is being provided is effective, let alone efficient"* (p. 3) ... *"Since the effectiveness of health care is established in only a relative small area of health care, a global budget can be introduced without necessary imperiling the quality at least in the developed countries"* (p. 5).

While the speakers agree that the efficiency of the health care system must be maximized in order to reduce the opportunity cost of health expenditures, they also recognize that assessing the efficiency level of the health care sector is not an easy task.

Stoddart and Lavis try to shed light on the matter by putting forth a classification of the various types of efficiency. However, they are led to conclude that it is almost impossible, given our present state of knowledge, to determine the *"trade-offs between health and 'other things' which contribute to well-being"* (p. 17).

This uncertainty is used by Ikegami as a justification for maintaining the status quo in the organization of the health care system so as to control costs and thus, paradoxically, increase the effectiveness and efficiency of health care expenditure. As with any decision pertaining to life, pain and death, the problem is seen by Evans as being dependent *"upon the level of collective support [the health care system] receive from the rest of us, which in turn depends*

upon our perceptions of the seriousness of the threats and the effectiveness of the responses which they can offer" (p. 11).

Rather than it being a technical matter, the efficiency of the health care system is a function of the belief systems of individuals and populations concerning health, well-being, and their determinants.

As pointed out by Ikegami, the efficiency of the health care system is largely dependent upon what happens between patients and health professionals and thus, upon the behaviour of individuals acting in a given society, at a given time and within a specific technical, economic, political and social environment.

Individuals behave according to their belief systems which derive from their interests, goals, and outlooks on life, death and pain. Inevitably, these belief systems constitute a very complex set of perspectives that are based, although not explicitly, on different outcome criteria aiming at different units of analysis (treatment of individual illness, improvement of population health, personal growth, etc.), within different timeframes, and often largely influenced by available information (life expectancy, mortality rate, utilization rate) that barely reflect the various goals of the individuals involved.

Under these circumstances, measuring the efficiency of the health care system and determining its underlying factors is pointless unless one recognizes, as Cameron (1986) proposes, that the concept of organizational efficiency, and all the more that of interorganizational system efficiency, is in essence paradoxical: *"Organizational effectiveness is inherently paradoxical. To be effective, an organization must possess attributes that are simultaneously contradictory even mutually exclusive"* (p. 545). Cameron uses Slaatte's definition in order to explain his thought further.

*"A paradox is an idea involving two opposing thoughts or propositions which, however, contradictory, are equally necessary to convey a more imposing, illuminating, life-related or provocative insight into truth than either factor can muster in its own right. What the mind seemingly cannot think it must think; what reason is reluctant to express it must express"* (Slaatte 1968, p. 4).

*"Paradox, then, involves contradictory, mutually exclusive elements that are present and operate equally at the same time ... no choice need be made between two or more contradictions. Both contradictions in a paradox are accepted and present. Both operate simultaneously."* Cameron, (1986, p. 545)

A good example of the paradoxical nature of the health care system is provided by Ikegami who believes the status quo in resource allocation to be the main instrument of change within the health care system. *"Japan has been able to achieve the seemingly impossible combination of cost containment, universal coverage and no over-rationing. This has not been the result of pursuing effectiveness and efficiency, but by maintaining the status-quo"* (p.2).

Once the "inherently paradoxical" nature of health care system efficiency has been recognized, it is easy to understand why there are so few empirical, quantitative, solid, reproducible, widely accepted examples of evidence for what Stoddart and Lavis call *"the trade-offs between different mixes of inputs involving less of health care and more of 'other things' which contribute to health... (and) to well-being"* (p. 17). The very nature and complexity of these problems prevent them from ever being settled definitely through accepted scientific approach and research. They belong to what Schumacher (1977) calls divergent problems as opposed to convergent problems which *"deal with distinct, precise, quantifiable, logical ideas that are amenable to empirical investigation. Convergent problems are solvable problems, and as they are studied more rigorously and more precisely, answers tend to converge into a single accepted solution. Divergent problems, on the other hand, are problems that are not easily quantifiable or verifiable and that seem not to have a single solution. The*

more rigorously and precisely they are studied, the more the solutions tend to diverge, or to become contradictory and opposite." (Cameron, 1986, p. 548).

Postman (1988, p. 5) makes a somewhat similar distinction between the study of "processes" which have nothing to do with human intelligence and can be scientifically accounted for and "practices" resulting from human decision and action, taking place in a specific environment and never fully predictable.

Problems may be classified as either convergent or divergent. Borrowing Fig. 3 from Stoddart and Lavis (p. 19). We can assume that questions A and B on the effectiveness of resources in producing services, and questions D and E on the effectiveness of services in modifying individual biological dysfunctions, are convergent problems. They can be solved. Research results should lead to specific recommendations. It is not surprising that most of the resources devoted to research bear on these matters.

The remaining questions, however, the most important ones, those that could enlighten the choices facing society, are divergent problems. *"The moment we deal with problems involving the higher levels of Being (life, consciousness and self awareness) we must expect divergence, for there enters, to however modest a degree, the element of freedom and inner experience. In them we can see the most universal pair of opposites, the very hallmark of life: growth and decay. Growth thrives on freedom... while the forces of decay and dissolution can be contained only through some kind of order. These basic pairs of opposites, Growth Versus Decay and Freedom Versus Order, are encountered wherever there is life, consciousness, self awareness."* (Schumacher, 1977, p. 125).

In the field of health, we are reminded by Stoddart and Lavis at the beginning of their presentation their (Fig. 2) of the complexity at play between the various factors affecting individual and population health. In their discussion of the various meanings of efficiency, however, one gets the impression that this complexity is lost, that health care is there to improve the health of individuals and therefore, that of the population, that the well-being of individuals depends upon their health and that the well-being of a population is simply the sum of well-being of its individuals (full arrows in Fig. 1, attached).

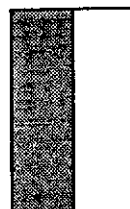
But, inasmuch as we know that the reverse relationships also exist (dotted arrows in attached Fig. 1), and that, for example, the health of individuals depends on their well-being or that there is a direct relationship between well-being and health care independently from the latter's effect on health, it becomes obvious that there can be no definite answers to questions about the health and well-being of populations.

These problems may not be solved once and for all. Research results constantly challenge one another. Uncertainty arises from the tension created by this apparent incompatibility. This, in turn, leads to the negotiation of compromises, calling for higher values (such as compassion, justice, equity) which "transcend" contradictions.

There is no definite, technical answer to the question of how to improve the efficiency of the health care system. At any given time, each society establishes temporary compromises based on the information provided by inevitably partial and contextual research, and the values of the various social groups.

FIGURE 1 - TAXONOMY

		Output			
		Health		Well-Being	
		Care	Individual	Population	Individual
Efficiency	Technical	A	B	C	K
	Cost-Effectiveness	D	E	F	M
	Allocative	G	H	I	L



Source: Stoddart G and Lavis JN. "Can we have too much Health Care?"

Honda Foundation Discoveries Symposium, Toronto, October 1993.

# Discussion of Papers by Naoki Ikegami, Gregory Stoddart/John Lavis and Robert Evans: Health Care and the Economy

---

*John Creighton Campbell*

I would like first to express my thanks for being invited. This conference has been extremely interesting to me. As a Japan-specialist political scientist who started working on health care only rather recently, this has been the first time I have encountered the ideas of population health research in so coherent and concentrated a form. For someone from south of the border, it has been refreshing to spend two days discussing health care and not hear the phrase "managed competition" or the name "Hillary" mentioned even once.

My comments begin with some remarks about earlier mentions of Japan. It is striking how often the Japanese have been able to identify the key points in a problem and get them right. It is as if they had heard all the lectures this weekend thirty or forty years ago. Specifically, Japan has emphasized to an unusual degree the following policies that seem to be positive for economic growth, good health outcomes, or both: equality of incomes; heavy human capital investment; concentration of resources on the early childhood years; attention to incremental improvements in techniques, or Japan as an "industrial lab;" and such basics as a high rate of investment, hard work on bringing down deficits and the like.

We could also say that Japan got it right on health care itself. That is, the Stoddard and Lavis paper points out that one can invest too much in health care (implying that many countries do), while Evans demonstrates that such overinvestment can hurt competitive advantage. Japan of course spends little on its health care system per se. In fact, its gap in health care spending with the U.S. (around 7 percent to about 14 percent) may well be a bigger source of competitive advantage than the often publicized gap in defense spending (about 1.5 percent to 5-7 percent).

So Japan seems also to have heard some lectures by Population Health people, perhaps even before the field was invented. It seems to me that people who think that regular medical care tends to use up too much money should be very interested in strategies for cost-containment. Here there are some aspects of the Japanese experience well worth examining. I will mention four that draw upon the Ikegami paper as well as his and my work together.

1. Japan's entire medical care system is oriented toward providing strong incentives for low-cost care. Under its fee schedule, visits to the doctors and ordinary tests are profitable, but open-heart surgery and other sophisticated procedures are often unprofitable. When you make money on cheap stuff and lose it on expensive stuff, overusage is not going to become a big financial problem.
2. The necessary conditions for effective cost-containment are universal coverage and a mandatory fee schedule. However, these are not sufficient conditions. In the 1970s, Japan had very much the same system as today, but its medical costs (even relative to GNP) were going up at about the same pace as in the United States. Containing costs

required a tough political decision in the early 1980s. Incidentally, what probably made this decision possible was widespread perception of impending fiscal crisis.

3. It is very difficult to serve multiple goals at once. In Japan as in most other countries, universal coverage and overall equality of benefits were achieved well before the main emphasis shifted to cutting costs. And as Ikegami pointed out, when cutting costs, the delivery system and the allocation of roles, power and money among providers was essentially left unchanged. There are lessons for the United States here.
4. Similarly, on the revenue side, Japan met its financial difficulties by cutting down on government subsidies to medical care from tax revenues. However, the increases in co-pays were minor; it relied mainly on a larger share for health insurance contributions, but mainly those from the better off systems through new measures for cross-subsidization. Japanese tend to see the most egalitarian "share-the-pain" strategies as most feasible politically.

The Japanese approach to health care delivery and finance is quite different from either Canada or the United States, but it does resemble various European systems in many respects. Many countries do a better job than North America in providing good health outcomes while controlling costs, and the political strategies they follow are worth close examination.

# Reconsideration on the Concept of Investment and Consumption in Health and Well-being

---

*Shuzo Nishimura*

## 1. Introduction

When we look at the history of economics, it has been crucial to distinguish the concept of investment from that of consumption. By inquiring the nature of investment, many economists have established theories to show how economic development is achieved. It would not be an exaggeration to say that economics, as an academic discipline, has its *raison d'être* by distinguishing these two concepts.

In early days of economics, economic units were simply divided into two: the household and the firm. There, the household was an economic unit which was solely engaged in consuming economic goods except for supplying labor. However, this distinction was obviously for the sake of simplicity.

G. Becker[1966] was one of the the pioneers who realized that even the household is engaged in several productive activities in addition to labor. It is called the human capital approach nowadays.

His work has been applied in many fields where investing in human capital is relevant. Economic analyses of such activities as education and job-training were remarkably developed owing to his work. Corresponding to these studies, people realized the importance of human capital for the economic development.

In the field of the economics of health and medical care, Grossman[1972a, b], by applying Becker's work, first constructed the model which considered the investment nature of medical care. Grossman tried to show how individuals demand health care under the income constraint. Following his model, many economists such as Cropper[1981], Newhouse and Phelps[1974], Dardanoni and Wagstaff[1987] paid attention to the investment nature of health care.

Since, however, these studies seem to be confined to individual behavior of the demand for health care, they are not clear concerning the macroeconomic implications, such as the role of national health expenditures in the national economy as a whole.

Moreover, because of the confusion of the concept of investment and consumption in health and health care, the role of the public sector in health either is underestimated or is exaggerated. For example, it seems to me that most of the public and the government officials, at least in developed countries, tend to think of health care only as a consumption good.

The purpose of this paper is to suggest a research agenda which will clarify the investment goods nature of the health expenditures at the aggregate level. I believe that this kind of research will give insight into problems such as how much public health expenditures should be appropriate.

Contents of the paper are the following: firstly, in section II, I will briefly review the Grossman's model. And in section III and IV, I will inquire into his model from the viewpoint of filling the gap between the behavior of individual and the behavior of the public policy.

## II. Review of Grossman's Model

Grossman's model is a thoughtful model. Here I briefly review his model by referring to Folland et al.[1993]'s exposition. One of the most important ways in which he took health care into account is that the demand for health has both of the following characteristics:

- (1) pure consumption aspects: health is desired because it makes people feel better; and,
- (2) pure investment aspects; health is also desired because it increases the number of healthy days available to work and thus to earn income.

Investment for health is attained by two kinds of behaviors: one is derived from medical expenditure, and the other is derived from conscious health-enhancing activity in the non-market sector.

This is shown as

$$I_t = I_t(M_t, TH_t; E_t) \quad (1),$$

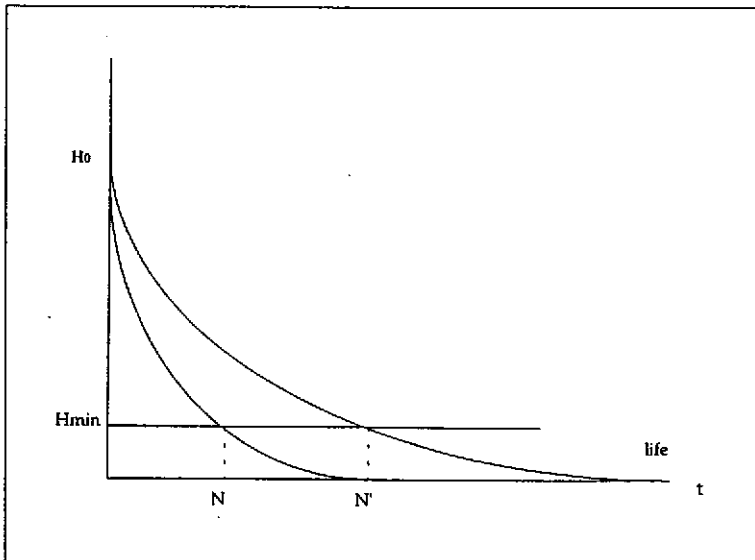
where  $I_t$  is investment,  $M_t$  is medical care,  $TH_t$  is the time spent for health-enhancing activity, and  $E_t$  is the stock of human capital. The variable  $E_t$  in this function is included to suggest that productivity in producing it may vary from person to person. Though  $M_t$  can contribute positively to increase the  $I_t$ , it will be diminishing.

The investment is related to health status as a following equation:

$$H_t = \int_0^t I_t dt + H_0 e^{-\delta t} \quad (2),$$

where  $H_t$  is the health status (which is measured by the stock level) at time  $t$ ,  $H_0$  is the inherited health status at the time of birth,  $I_t$  is investment for health at time  $t$ , and  $\delta$  is the rate of depreciation. The rate of depreciation is assumed to be exogenous. As an example, this relation is depicted in Figure 1.

Figure 1 Grossman's Model



Health Status depreciates at the rate of  $\delta$

Death takes place when  $H_t = H_{\min}$ . This means that the larger the investment for health is, the longer the length of life is. Life expectancy is an endogenous variable.

Consumption aspects is expressed as follows

$$U = U(\phi_0 H_0, \dots, \phi_t H_t Z_0, \dots Z_t) \tag{3}$$

where  $U$  is an utility function of a typical consumer,  $\phi_t$  is the service flow per unit of stock, and  $h_t = \phi_t H_t$  is total consumption of health services, and  $Z_t$  is total consumption of another commodity in the  $t$ th period.

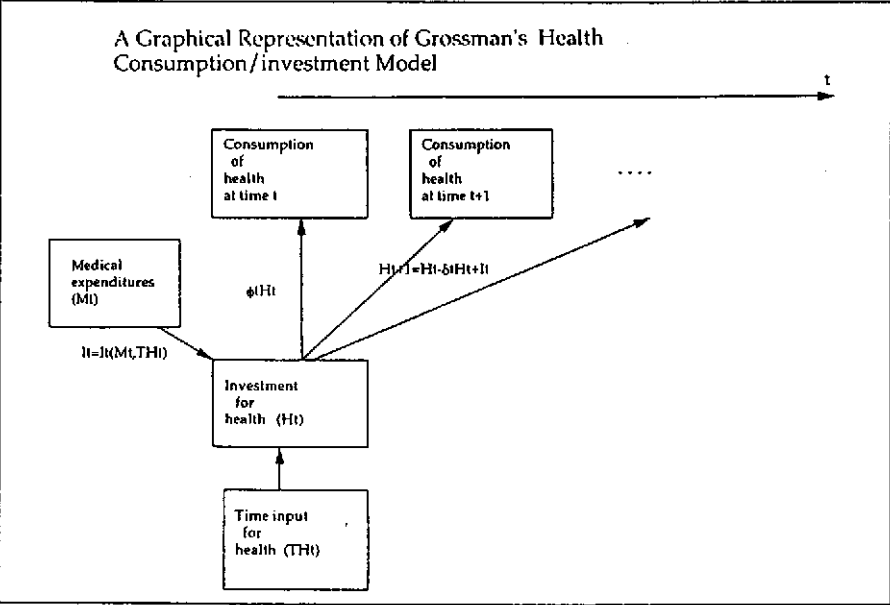
According to Grossman's model, one's utility of consumption of health depends on one's health status, that is, it increases depending on how healthy people are.

Another important aspect of his model is that health status is assumed to be enhanced by one's allocation of time. This is shown in equation (1). Based on Becker's model, people are assumed to have the time constraint as

$$W_t = TW_t + TH_t + T_t + TL_t \tag{4}$$

where  $W_t$  is the total amount of time available in period  $t$ ,  $TW_t$  is hours of work,  $TH_t$  is time input for health investment,  $T_t$  is the time necessary to consume  $Z_t$ , and  $TL_t$  is the time lost from market and non-market activities due to illness and injury. Grossman's model can be graphically shown as Figure 2.

Figure 2



Grossman not only constructed a theoretical model, but also tested his model on the U.S. panel data. Though his results are interesting, hereafter I will investigate the problems of his model at the theoretical level.

### III. Contributions of Grossman's Model

Though Grossman's model has several drawbacks, it helps us clarify the relationships among several concepts such as health, medical expenditures, and people's efforts to enhance their own health status.

Before discussing the policy matters in general, I would like to raise one example of the recent policy issues. One of the most important policy issues which most countries now face is the increase of medical expenditures for the elderly.

For example, share of medical expenditure for the aged 65 and over, among total medical expenditures in Japan, is more 40 percent. This share is still dramatically increasing. In most developed countries, these medical expenditures on the elderly are increasing as the society becomes older.

Thus it would be natural to ask the question "To which should we give priority in allocating medical resources, for the elderly or for the non-elderly?"

Relying on Grossman's model, first of all, I will try to answer this question as an individual choice. I will compare characteristics of medical care for the elderly with those for the non-elderly. As for investment aspects of medical care, health( $H_t$ ) is related to medical expenditure( $M_t$ ) in the equation (1). Though marginal effect of the elderly may be larger than that for the elderly in general, and therefore though cost-effectiveness of health care might be lower for the elderly, it would be natural to think medical expenditures are demanded more by the elderly. Moreover, since, health status depreciates as people grow older, the amount of investment can be larger when they are old. As an investment, people invest more when they grow older.

On the consumption side, the situation is different. Though the marginal effect of medical care for the elderly is also lower than that for the non-elderly in general, whether people consume more or less will depend on their own preferences.

This conclusion which was deduced from an individual choice perspective, seems to be different from societal judgment, that is of setting priority on resource allocation of health care between the elderly and the non-elderly. Society's general sentiment seems to be as follows: since the elderly do not live longer, as an investment, it is less effective for the elderly to invest for health.

This gap between individual choice and social judgment seems to be explained at least by two factors: one is the problem of time preference and the other is the interpretation of the concept of investment.

First, it is important to examine the discrepancy between individual time preference and the social time preference. Analytical methods such as cost-benefit analysis and cost-effectiveness analysis, which aim to show how health resources are used by utilizing social discount rate, will mislead to unfavorable conclusion for the elderly, if this discrepancy is significant.

As a matter of fact, time preference of the elderly is not always high, that is, they do not necessarily give higher weight to the present than to the future. For the elderly, health expenditures as an investment rather than as a consumption might be more important.

Secondly, it seems to me that there is a social prejudice that investment is more important than consumption. As economics tells, the ultimate determination of people's happiness depends on consumption and does not depend on investment. However, under the pressure of international economic competition, there is a tendency for the society to invest more for the survival of each nation.

As a whole, common-sense logic about the allocation of health resources is as follows: since health expenditures mainly consists of consumption goods and consumption is less valuable than investment, it is better for the society to consume as few resources as possible.

I would like to indicate another approach in which investment aspect of health expenditures are neglected. Thanks to OECD's laborious efforts, we now have ample internationally common health-care-related data.

Many studies have been done to inquire into the relationship of health, medical expenditures, and national income. Since Newhouse and Phelps[1974]'s pioneering work, many economists have made clear that there is a strong positive correlation between national medical expenditures and the GDP or the national income. However, the effect of medical expenditures as an investment good is not clear.

Thus, many economists consider medical expenditures only as an consumption good.

Since income elasticities of medical expenditures were shown to be higher than one, some people have concluded that it is a luxury good. It was said not to be a necessity good. However, as Parkin et al.[1987] pointed out, this is the conclusion deduced from a misunderstanding of the concept.

This kind of misunderstanding also seems to have strengthened the common view that medical expenditures for the elderly are less valuable.

#### IV. Limitations of Grossman's Model

What I wanted to emphasize in the preceding section is that ultimate well-being of human being depends on consumption. The only significance of the investment is that it will contribute to consumption in the future. This indicates that we should pay much attention to the time preference. However, recent economic analyses for health care do not sufficiently consider the possible discrepancy of time preference between the elderly and the non-elderly.

Though Grossman's model provides the analytical tools for this kind of analysis, there is an important limitation. That is, his model does not consider the uncertain nature of health care. In his model, people can plan how long they live. In reality, death sometimes comes suddenly. In some cases, it irrespective of one's past health. This is an example of the limitation of his model.

Another important limitation of Grossman's model, as well as other economic models, is that investment function is considered to be known for each individual. In spite of many laborious clinical epidemiological studies, there are many kinds of health care, the effect of which is quite uncertain.

Especially, for the effect of time input ( $TH_t$ ), the effects of many of health care activities are ambiguous. Moreover, there is another difficulty which economic models can suggest. Though it is natural that, in economic models, people are assumed to pursue their happiness under cost constraint, as for prevention activities, the concept of cost is hard to estimate. For example, according to epidemiological studies, we know well that cigarette smoking cessation is effective for one's health. However, how can we estimate the cost to quit smoking. For some people, it is costless and for other people, it will need much time.

In general, to estimate the cost of time within household activity is quite difficult. Therefore, household time investment tends to be under-estimated.

#### V. Concluding Comments

In terms of theoretical economics, the distinction between investment and consumption is clear. When, however, we want to go forward to empirical studies, this distinction becomes ambiguous. If we change the viewpoint, it can be said that the concept of investment has come to play too heavy role in economics. Nevertheless, we have to try to further clarify the investment nature of health care activity.

## References

- 1 Becker, Gary.S. "A Theory of the Allocation of Time." Economic Journal 75 (1965): 493-517.
- 2 Cropper M.L. "Measuring the Benefits from Reduced Morbidity," American Economics Review, Papers and Proceedings 71:235-240.
- 3 Dardanoni, Valentino and Adam Wagstaff. "Uncertainty, Inequalities in Health and The Demand for Health." Journal of Health Economics 6 (1987):283-290.
- 4 Folland, Sherman, Allen C. Goodman, and Miron Stano. The Economics of Health and Health Care. New York: Macmillan Publishing Company, 1993.
- 5 Grossman, Michael. The Demand for Health: a Theoretical and Empirical Investigation. New York: National Bureau of Economics Research, 1972a.
- 6 Grossman, Michael. "On the Concept of Health Capital and the Demand for Health," Journal of Political Economy, 1972b, 80(2):223-255.
- 7 Newhouse Joseph P, and Phelps Charles E. "Price and Income Elasticities for Medical Services, in M.Pertman ed. The Economics of Health and Medical Care, Macmillan, London, 1977, 139-161.
- 8 Newhouse, Joseph P. "Cross National Differences in Health Spending What Do They Mean?" Journal of Health Economics 6 (1987): 159-162.
- 9 OECD. OECD Health Systems: Facts and Trends, 1960-1991. OECD, 1993.
- 10 Parkin, David, Alistair McGuire, and Brian Yule. "Aggregate Health Care Expenditures and National income." Journal of Health Economics 6 (1987): 109-127.

# Discussion of Paper by Professor Nishimura

---

*John R. Evans*

Professor Nishimura has presented a thoughtful analysis of an economic model of health care as an investment and as consumption. The model has several interesting features, including an attempt to factor in the contributions of the patient to maintaining health as well as provider investment.

Models as such are useful instruments to analyze assumptions but this model contains one assumption which is particularly worrisome: It implies that there is no productive return from investment in health care of the elderly. This is neither valid nor ethically acceptable. It is a mistake to think of health care for the elderly only in terms of consumption. There is abundant evidence of contributions of the elderly to economic and social productivity and through investments in health care there can be avoidance of cost of expensive health and social services which is a form of potential productivity. This productivity is only realized if the savings are not dissipated in other services.

There is a strong case for health care investment in the elderly to decrease the demand on health and social services. For example, exercise and pharmaceutical therapy to delay osteoporosis; replacement of parts such as hips to maintain independent, pain free mobility or lenses for loss of vision from cataracts; and palliative care in the case of terminal illness such as cancer which would allow individuals to die with dignity out of hospital rather than consume vast resources in the hospital setting. The return on these investments would diminish if the interventions were less effective or the stage of the disorder allowed relatively short improvement in the quality and independence of life. The object of this investment is to maintain or restore function, not to cure disease.

In the earlier discussion of economic development and health status, I was particularly impressed with the report of Richard Wilkinson on the promptness of decline of life expectancy in the UK with widening income distribution after 1985. The very rapid expression of that decline in life expectancy suggests dramatic biological consequences or self-destructive behaviour. Professor Takeuchi and Peter Nicholson, among others, gave an encouraging outlook for the return to prosperity following the transition from a mass production, material and energy intensive economy to the post-industrial information economy. They left the impression that the "jobless economy" might be a relatively short stage of transition. I would like to urge some caution about their optimism for two reasons. First, I believe the growth in economic prosperity will not necessarily be reflected in growth in employment. Secondly, the benefits of that economic growth are likely to be concentrated in a smaller proportion of the countries in the world and of the individuals within those countries. There is a high probability, therefore, that a large number of people will not share in the economic recovery or the non-economic benefits of being employed. The disparities in wealth generation and lack of employment opportunities may result in much longer and deeper social dislocation with the negative impact on health described by Professor Wilkinson. This should motivate us to look for preventive measures and other interventions which may diminish this impact and provide the support structure not available through conventional employment.

In this context and in response to the appeal of Dr. Mustard and Dr. Frank to explore the "relationship between economic growth, prosperity and high quality social environments," I should like to refer to Robert Putnam's studies which suggest that economic development is strongly influenced by the quality of social organizations in the communities. Historical reviews in Italy suggest that communities did not become civic because they were rich, but rather became rich because they were civic. Furthermore, Putnam's recent study of twenty new powerful regional governments in Italy shows wide differences in success in promoting investment in economic development, job training and innovation, reasonable environmental standards and observance of the law. Success was not related to party politics, affluence or prosperity. The best predictors of success were strong traditions of civic engagement as reflected by voter turnout, newspaper readership and active membership in churches, community organizations, choral and literary societies, sports clubs, Lions Clubs, etc., social and political networks that are organized horizontally not hierarchically. Putnam described these aspects of civic engagement as social capital, a type of capital that is not depleted by use; indeed, it is strengthened and renewed through use.

In a society that is racked by social change, unemployment, large differences in cultural background and which is leaving young children, adolescents, youth and older people out of the mainstream to an increasing degree, it is fundamental that we match our efforts in economic development with efforts to achieve more effective social organization. There is considerable evidence from the type of studies reported at this conference of the importance of social networks in sustaining health. It is urgent that we go beyond diagnosis of determinants of health and identify interventions which will bring the many excluded groups back into the mainstream of our society. We will not have the money to buy their inclusion, nor would that be successful because of the dependency it would create. We must look, therefore, for new innovative methods whereby people to people relationships are reestablished at local and community level. Moving from diagnosis to intervention in the development of social capital is every bit as pressing an issue for the well being of our society as the critical investments in renewal of economic prosperity.

# “Difference of Definition and Concept of Health between Western and Eastern Medicine”

---

*Kazuhiko Atsumi*

**A**t the present state, there are big differences in definition and concept of health between western and eastern medicine.

In western medicine, definition of health has not been clarified. An individual's health is determined as a condition without pathological findings by chemical and physical examinations which are used in present-day clinical medicine. Namely, the definition of health is a non-pathological condition, but not a completely healthy one.

On the other hand, definition of health in eastern medicine has actually been broader. The condition of health is not only physically well but also psychologically noble.

In western medicine, a disease is determined statistically on the basis of cell pathology. However in eastern medicine, a disease is determined individually on the basis of personal constitution and is considered as a deviation from the healthy condition.

In western medicine, science is based on diagnosis and therapy of diseases, but in eastern medicine, human experience and understanding of the universe are the basis of health.

In general, some mental phenomena and human behaviours cannot be understood with ready-made science. This is indeed the limitation of western medicine.

Furthermore, dualism is a western way of traditional thinking; however, monism is an eastern way. Therefore, eastern people have long been considering that physical health is correlated more closely with mental condition than western people have.

Finally, from the viewpoint of the difference between western and eastern medicine, I would like to suggest that we frame our discussion on health on the basis of eastern medicine.