UNIVERSITY OF TORONTO
MUNK CENTRE
FOR INTERNATIONAL STUDIES
AT TRINITY COLLEGE

MCIS BRIEFINGS

APRIL 2010

Comparative Program on Health and Society Lupina Foundation Working Papers Series 2007–2009

Edited by
M. Bianca Seaton
and
Sara Allin
Comparative Program on
Health and Society
University of Toronto

MCIS BRIEFINGS

Comparative Program on Health and Society Lupina Foundation Working Papers Series 2007–2009

Edited by

M. Bianca Seaton
and
Sara Allin
Comparative Program on
Health and Society
University of Toronto

Munk Centre for International Studies
University of Toronto

Munk Centre for International Studies
University of Toronto
1 Devonshire Place
Toronto, Ontario, Canada M5S 3K7
Telephone: (416) 946-8900
Facsimile: (416) 946-8915
E-mail: munk.centre@utoronto.ca
Website: www.utoronto.ca/mcis

978-0-7727-0844-1
ISSN 1715-3484

The Munk Centre for International Studies at the University of Toronto seeks to be an internationally recognised leader in interdisciplinary academic research on global issues and to integrate research with teaching and public education. We place special emphasis on the fostering of innovative interdisciplinary knowledge through the exchange of ideas and research among academics as well as the public, private, and voluntary sectors.

COMPARATIVE PROGRAM ON HEALTH AND SOCIETY
Munk Centre for International Studies
University of Toronto
1 Devonshire Place
Toronto, Ontario, Canada M5S 3K7
Telephone: (416) 946-8891
Facsimile: (416) 946-8915
E-mail: cphs.munk@utoronto.ca
Website: www.utoronto.ca/cphs

The CPHS Working Papers Series

The Comparative Program on Health and Society maintains a collection of academic papers which we call our Lupina Foundation Working Papers Series. These works can range from research papers to thought pieces; and from statistical analyses to historical case studies. Our series represents a snap-shot of the work being done by our Lupina Fellows, past and present. Taken together, our Working Papers Series encapsulates the wide-ranging approaches to the study of the social determinants of health. We hope that you will find the individual papers in our series thought-provoking and helpful.

MCIS Briefings

1. *Bound to Follow? US Foreign Policy, International Reactions, and the New Complexities of Sovereignty.* By Louis W. Pauly.
   September 2005 ISBN 0-7727-0825-5.

2. *The Multilateral Agenda: Moving Trade Negotiations Forward.* By Sylvia Ostry.
   November 2005 ISBN 0-7727-0822-3.

3. *The Jerusalem Old City Initiative Discussion Document: New Directions for Deliberation and Dialogue*. By Michael Bell, Michael J. Molloy, John Bell and Marketa Evans.
   December 2005 ISBN 07727-0823-1.

4. *Comparative Program on Health and Society Lupina Foundation Working Papers Series 2004–2005*. Edited by Jillian Clare Cohen and Jennifer E. Keelan.
   January 2006 ISBN 0-7727-0818-5.

5. *Comparative Program on Health and Society Lupina Foundation Working Papers Series 2005-2006*. Edited by Jillian Clare Cohen and Lisa Forman.
   October 2006 ISBN 0-7727-0829-0.

6. *Darfur and Afghanistan: Canada's Choices in Deploying Military Forces*.
   By Ambassador David S. Wright.
   October 2006 ISBN 0-7727-0830-4.

7. *Trade Advocacy Groups and Multilateral Trade Policy-Making of African States*.
   By Sylvia Ostry and Thomas Kwasi Tieku.
   April 2007 ISBN 978-0-7727-0832-8.

8. *Water Diversion, Export, and Canada-U.S. Relations: A Brief History.*
   By Frank Quinn.
   August 2007 ISBN 978-0-7727-8054-6.

9. *Intersubjectivity in Literary Narrative*. By Tomas Kubicek
   October 2007 ISBN 978-0-7727-0834-2.

10. *Comparative Program on Health and Society Lupina Foundation Working Papers Series 2006-2007*. Edited by Jillian Clare Cohen-Kohler and M. Bianca Seaton.
    November 2007 ISBN 978-0-7727-0838-0.

11. *A Model Act for Preserving Canada's Waters*. Canadian Water Issues Council in collaboration with the Program On Water Issues.
    February 2008 ISBN 978-0-7727-0839-7.

    Loi type sur la protection de l'eau au Canada. Conseil sur les questions de l'eau au Canada en collaboration avec le Programme sur les questions de l'eau.
    Février 2008 ISBN 978-0-7727-0840-3.

MCIS Briefings (continued)

12. *The World's First Anti-Americans: Canada as the Canary in the Global Mine*. By Richard Gwyn.
March 2008 ISBN 978-0-7727-0842-7.

13. *Comparative Program on Health and Society Lupina Foundation Working Papers Series 2007-2009*. Edited by M. Bianca Seaton and Sara Allin.
April 2010 ISBN 978-0-7727-0844-1.

Table of Contents

Foreword

The end of the first decade of the millennium was characterized by tremendous global change, as evidenced by the upheaval in financial markets, new leadership of national governments, and innovative developments in science and technology. In the domain of health, things have been equally interesting; the outbreak of pandemic H1N1 influenza, introduction of novel forms of medical treatment for diseases old and new, and inventive strategies for the organization of healthcare, both locally in Canada and internationally, have all provided much food for thought for the researchers affiliated with the Comparative Program on Health and Society (CPHS).

The CPHS, which was founded in 2000, is a dynamic research institute based at the Munk Centre for International Studies at the University of Toronto. Founded and funded by the Lupina Foundation, the CPHS supports innovative, interdisciplinary comparative research on health, broadly defined through our extensive range of fellowships.

In this Munk Centre for International Studies Briefing Series, we present the fourth collection of papers from CPHS Fellows. The thoughtfully written, insightful, and challenging papers in this edition reflect the diverse range of intellectual pursuits and substantive research interests of scholars involved with the CPHS between 2007 and 2009. We hope you will find these papers to be informative about areas of contemporary health research with which you may be currently unfamiliar, and inspiring in their presentation of novel perspectives on topics about which you may already be knowledgeable. Such is the great gift, and unique promise, of a collection of working papers. We are delighted to present these to you for intellectual consideration and personal enjoyment.

The collection begins with papers by Patrick Zylberman and Lisa Forman, who both, in their own way, invite us to consider the impact of particular ways of framing, and naming, issues related to health. In "Neither Certitude nor Peace. How Worst-Case Scenarios Reframed Microbial Threats, 1989–2006," historian Zylberman charts important shifts in the way that Western governments have conceptualized outbreaks of infectious disease, both real and imagined, over the past twenty years. He asserts that the politicization of disease, in particular the discursive move from disease as population health issue to "human security" issue, has paved the way for the use of "worst-case scenario" strategies as a tool to help governments plan for and make decisions about how to deal with microbial threats. The end result, according to Zylberman, has been the deliberate rejection of scientific rationality in favour of historical fantasy. Thus, he argues, governments are no longer taking into consideration the true likelihood of such outbreaks actually taking place in their strategic planning and policy-making activities, but rather remain focused on the worst possible imaginings of what "might" be. Turning our attention to more a tangible health threat is Lisa Forman, who considers the value of framing human health as a legal priority over human property in her paper, "An Elementary Consideration of Humanity? Linking Trade-Related Intellectual Property Rights to the Human Right to Health in International Law." Using access to essential medicines and pharmaceutical patents as a case in point, Forman points out that trade rules ought to be subject to reasonable limitations in order to protect human health and life rather than taking priority, as they have in international legal disputes

of late. She proposes the integration of trade agreements into the broader normative system of international law, and argues that such reforms are legally warranted and socially desirable.

Political scientists Antoinette Handley and Suzanne Hindmarch draw our attention to the pressing social, political, and economic consequences of AIDS in Africa. In "The State That AIDS Makes: The Moral and Political Economy of Epidemic in South Africa," Handley explores how the AIDS epidemic has shaped the political and economic contours of South African society. Arguing that the epidemic and the state have been mutually co-constituting, Handley documents how AIDS has informed the respective roles and responsibilities of public and private sector actors in South Africa over the past two decades. Her paper raises provocative questions about what happens when businesses, rather than the government, become the primary providers of health services to their workers living in a particular society. Hindmarch focuses on the pivotal and unprecedented reframing of HIV as a threat to peace, security, and development in Africa, rather than merely a disease threatening population health and well-being, by the United Nations Security Council in 2000. Hindmarch asks why the UN Security Council chose to consider HIV at all, and suggests that this reframing was part of a broader move towards more expansive understandings of human security in the UN system. Her paper also argues that the ultimate outcome of this reframing for the treatment of HIV by the UN has been "business as usual"—proposed solutions for Africa continue to focus on technical, behavioural, and individual interventions instead of broader structural reform.

Brigit Ramsingh and Regine King tell us the "story behind the story" of two very different types of commissions: The Codex Alimentarius Commission focused on international food safety standards, and the Gacaca Truth Commission on psychosocial healing in post-genocide Rwanda. Ramsingh, an historian, charts the evolution of international food safety standards in the post-war period and explores how a range of different tensions—between international organizations, previously existing standards, and perspectives on what constitutes "correct science"—shaped their development and eventual uptake by different governments. King explores a different range of tensions that affect the uptake of a commission's work, namely those between top-down conceptual frameworks for healing psychosocial trauma in post-conflict situations, and bottom-up approaches. Using the case of post-genocide Rwanda, King argues that while top-down may be seen as more legitimate, it runs the risk of further disempowering and dividing an already vulnerable population; grassroots projects, on the other hand, offer greater potential to contribute to the healing of psychosocial trauma and community building.

Between 2007 and 2009, the CPHS supported the work of several scholars engaged in comparative research focused on the use of or decision making around health services in Canada. For example, Roger Chafe and colleagues ("The Effect of Contextual Differences on Health Resource Allocation Decision Making") examine the role of contextual factors in the politically charged and often difficult resource allocation decisions around health services in three Canadian provinces. They find evidence of variations in the decisions around funding services in three sectors—acute care, diagnostics, and rehabilitation across these jurisdictions—which leads to the conclusion that proposals for reforming resource allocation need to acknowledge the institutional and program-level environments in which they are being employed.

In their paper, "An International Comparison of Priority Setting for Orphan Drugs," Zahava Rosenberg-Yunger and colleagues take a closer look at one particularly complex and controversial resource allocation decision: the funding of drugs that are very expensive but unique in their ability to provide treatment for extremely rare, life-threatening diseases. Rosenberg-Yunger et al. explore the funding considerations and decisions for two such "orphan drugs" across three countries' drug reimbursement committees and explain that there is a great deal of consistency in the value that different decision makers place on evidence, rule of rescue, and equity in their decision-making process about which drugs to fund. Once medicines are approved for funding, such as in Ontario's public drug program, is the use of these medicines equitable across all members of the population? Sara Allin addresses this question in her paper, "Equity in Prescription Drug Use among Older Ontarians." Allin takes a unique methodological approach in her study, examining first the comparability of two sources of information on prescription drug use in Ontario and then modeling the effect of income, education, and other individual factors on the use of publicly funded medicines. She concludes that her two data sources are not easily comparable and that there is a relatively equitable pattern of prescription drug use among the older Ontario population. Her paper draws attention to the challenges of collecting accurate information on medication use, especially among the oldest population groups, and the potential overuse of medications among those with the lowest income.

Turning to the structures and organizations in health care, in "A New Chapter in the Interprofessional Collaborative Literature Developing a Framework to Examine Professional Culture on a Family Health Team," Jennifer Beales and colleagues examine what happens when a diverse group of health care professionals come together to work in a collaborative fashion to provide patient-centred care, as they are increasingly doing in the "family health teams" recently introduced in Ontario. The focus of this work is on professional culture: they develop a framework to help us understand how the customs, practices, and disciplinary perspectives of different health care professionals intersect in practice on these newly formed teams. Given this important shift in Ontario towards patient-centred care and interprofessional collaboration, the work of Beales et al. is valuable in that it sheds light on the complex array of cultural factors that influence collaboration among health care professionals and will ultimately support and sustain their collective goal of providing patient-centred care.

The final two papers in this collection present novel methodological approaches to understanding interrelations among social determinants of health. The complexity of health determinants and inequalities requires sophisticated methodological techniques, Jason Fletcher, Steven Lehrer, and Heather Spielvogle accomplish this beautifully. Fletcher and Lehrer draw on a unique dataset of adolescents in the United States that integrates genetic data with information on social and health information to estimate the casual pathway between poor health and academic outcomes in their co-authored paper, "Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement." Taking an instrumental variables approach that uses relevant genetic markers as instruments, and controlling for family effects, Lehrer and Fletcher report robust evidence of an effect of poor mental health on educational outcomes. These methodological innovations pave the way for further integration of biological and social science

research in order to gain a better understanding of the causal relationships between health and its social determinants. The series concludes with Heather Spielvogle's paper, "Understanding and Addressing Barriers to Adolescent Mental Health Service Retention: An Intervention to Enhance Engagement," which employs an innovative approach to understanding and circumventing barriers to mental health care for adolescents with mental health problems. Spielvogle draws on adolescents' perceptions of the barriers to accessing care to develop a framework for an engagement intervention aimed at encouraging adolescent participation in the therapeutic relationship. This work not only has implications for the delivery of mental health services to children and adolescents but also makes important methodological contributions by incorporating service users' perceptions and perspectives into the analysis.

Acknowledgements

As with all initiatives of the CPHS, the research and publication of these working papers would not have been possible without the generous support and patronage of Margret Hovanec and Peter Warrian of the Lupina Foundation, and the intellectual support and leadership of Janice Stein and Carolyn Tuohy. We are also grateful to Janet Hyer for her guidance and assistance with the preparation of this publication, and to all of the CPHS Fellows and other scholars who provided feedback on previous versions of these papers. Lastly, we extend our most heartfelt thanks to Lisa Forman for encouraging and supporting the publication of these papers.

The Editors

M. Bianca Seaton and Sara Allin

Contributors

Sara Allin, PhD, is a CPHS and CHSRF postdoctoral fellow in the Department of Health Policy, Management and Evaluation, University of Toronto. Sara's PhD is in Social Policy from the London School of Economics and Political Science, and her research interests include equity in health and health care, and comparative health policy.

Zubin Austin, BScPhm, MBA, MIS, PhD, is Associate Professor, Leslie Dan Faculty of Pharmacy, University of Toronto. His current research interests include bridging education programs for foreign-trained professionals seeking licensure in Canada, development of professional identity among pharmacy students, maintenance of professional competency, and the influence of peer groups on learning and professional practice.

Jennifer Beales is a doctoral candidate at the Leslie Dan Faculty of Pharmacy, University of Toronto. She received her undergraduate degree in Sociology from Wilfrid Laurier University and also holds an MA in Sociology from the University of Western Ontario. Her research interests include qualitative methodologies in health science research, interprofessional collaboration on primary health care teams, interprofessional education, occupations and professions, gender and nontraditional professions, and the study of professionalism.

Roger Chafe, PhD, is a CHSRF-CIHR postdoctoral fellow in the Department of Health Policy, Management and Evaluation at the University of Toronto and in the Cancer Services and Policy Research Unit at Cancer Care Ontario. He is also an Adjunct Professor in Dalhousie University's School of Health Services Administration and a Research Associate of the Comparative Program on Health and Society. His research interests include health care decision making, resource allocation, coverage decisions, health policy, and public participation.

Abdallah S. Daar is Professor of Public Health Sciences and of Surgery at the University of Toronto. He is also Senior Scientist and Director, Program on Ethics and Commercialization of the McLaughlin-Rotman Centre for Global Health and Director of Ethics and Policy at the McLaughlin Centre for Molecular Medicine.

Jason M. Fletcher, PhD, is Assistant Professor, Division of Health Policy and Administration, Yale School of Public Health. His research focuses on examining social network effects on adolescent education and health outcomes, combining genetics and social science research, estimating long-term consequences of childhood mental illness, and child and adolescent mental health policy.

Lisa Forman, BA, LLB, MA, SJD, is Assistant Professor, Dalla Lana School of Public Health, and Director, Comparative Program on Health and Society, Munk Centre for International Studies. Her research interests include the right to health in international human rights law, global health and human rights, trade-related intellectual property rights and access to medicines globally, right to health impact assessments, and South African constitutionalism related to health.

Antoinette Handley, PhD, is Associate Professor, Department of Political Science, University of Toronto. Her research interests include comparative development, the political economy of reform in developing countries, African politics (especially Southern Africa), business-government relations, and the political economy of epidemics, especially HIV/AIDS.

Suzanne Hindmarch is a PhD candidate in the Department of Political Science at the University of Toronto. She holds a BA from the University of Alberta and an MA from Dalhousie University. Her research focuses on global health (with a current focus on HIV) and the role of ideas and institutions in shaping policy responses to health challenges.

Regine King, MEd, is a PhD student in the Factor-Inwentash Faculty of Social Work, University of Toronto. She is interested in finding alternative mental health approaches for survivors of massive violence and reconciliation processes for post-conflict reconstruction.

Murat Kristal, MBA, PhD, is Associate Professor of Operations Management and Information Systems, Schulich School of Business, York University.

Steven F. Lehrer, PhD, is Associate Professor, School of Policy Studies, Queen's University, Kingston. His research interests focus on health economics, the economics of education, experimental economics, and empirical microeconometrics.

Douglas K. Martin, PhD, is Associate Professor with the Department of Health Policy, Management and Evaluation and the Joint Centre for Bioethics, University of Toronto.

Doreen Neville, ScD, is the Associate Vice President, Academic, at Memorial University of Newfoundland and an Associate Professor of Health Policy and Health Care Delivery in the Faculty of Medicine. Her research interests include governance, the role of information technology in health and education, and the development and evaluation of health policy.

Christine Papoushek, PharmD, is with the Leslie Dan Faculty of Pharmacy, University of Toronto and the University Health Network (UHN) Department of Pharmacy, Toronto Western Family Health Team, Toronto.

Brigit Ramsingh is a doctoral candidate at the University of Toronto's Institute for the History and Philosophy of Science and Technology (IHPST). As of March 2010, she will be completing her thesis at the Max Planck Institut für Wissenschaftgeschichte (MPIWG) in Berlin. Brigit is also a staff member at the Charité Institut für Geschichte der Medizin as the Programme Coordinator for the European Science Foundation DRUGS Research Networking Program (Standard Drugs and Drugs Standards), which is a comparative project on the history of pharmaceuticals in the twentieth century.

Thomas Rathwell, PhD, is a Professor in the School of Health Services Administration, Dalhousie University, and was Director of the School from 1995 to 2007. He was a founding member and Dalhousie University lead (2001–9) of the Atlantic Regional Training Centre in Applied Health Services Research, a consortium of Atlantic Canada leading universities. His research interests embrace health policy analysis, comparative health care, primary health care, and health system reforms.

Zahava R.S. Rosenberg-Yunger, PhD, is a postdoctoral fellow with the Department of Health Policy Management and Evaluation and the Joint Centre for Bioethics, University of Toronto.

M. Bianca Seaton is a doctoral candidate at the Dalla Lana School of Public Health at the University of Toronto. She holds an MSc in Public Health Sciences (Social Science and Health/Women's Studies) from the University of Toronto, and a BASc in Arts and Science from McMaster University. Her current research interests include the sociology of risk, qualitative methodology, and the relationship between knowledge and practice in hospital-based nursing work.

Heather Spielvogle is a doctoral candidate in the Factor-Inwentash Faculty of Social Work at the University of Toronto. She received a Master's Degree in Social Work from the University of Pittsburgh. Her areas of research include mental health, culturally competent social work practice, therapeutic engagement and alliance, and Internet-based psychotherapy.

Halla Thorsteinsdóttir is with the McLaughlin-Rotman Centre for Global Health at University Health Network/University of Toronto and Dala Lana School of Public Health, University of Toronto.

Patrick Zylberman is professor of the history of health, École des hautes études en santé publique, EHESP School of Public Health, Rennes and Paris. In collaboration with Antoine Flahault, he was charged with the scientific direction of the exhibition "Epidemik" which was on display at the Cité des sciences et de l'industrie, Paris-La Villette, October 2008–January 2010. He is currently working on bioterror and antipandemic scenarios which imposed new images of microbial threats that might affect the management of epidemic crises.

Neither Certitude nor Peace. How Worst-Case Scenarios Reframed Microbial Threats, 1989–2006

Patrick Zylberman

Abstract

Governments are increasingly resorting to scenarios and tabletop exercises, such as TOPOFF 2000, Dark Winter (2001) or Atlantic Storm (2005). Western governments share scenario-based strategies (September 2003 Global Mercury exercise). There is compelling evidence that recent health crises such as the anthrax letters (2001) or SARS (2002–3) have had a more far-reaching impact on emergency preparedness and response ahead of health surveillance and outbreak management in governments' agendas. This paper looks at government attempts to train the citizenry in vigilance, alertness, and response. It explores more specifically how Washington responded to health threats, namely the anthrax letters of October 2001, and why the worst-case scenarios became dominant in the organization of emergency preparedness and response schemes. Another case in point is the smallpox vaccination campaign of 2002–3.

Introduction

At the end of September 2005, the WHO coordinator for influenza was plugging his view of a pandemic risk: if avian flu becomes a human disease, said he, between 2 and 150 million people will die globally (Associated Press 2005). The range was rather broad! A few months earlier, the 1918 pandemic had been lurking in everybody's mind in Geneva. It also informed Anthony Fauci's thoughts when, in February 2006, he addressed a medical audience in Boston about flu risk. (Fauci directs the Bethesda National Institute of Allergy and Infectious Diseases and was a scientific adviser to George W. Bush.) In the United States as well as in Europe, pandemic plans turned out to be more

alarmist than they had been in the past. In May 2005, *Nature* expressed the same panicky feeling (On a Wing 2005), and at about the same time Michael Osterholm, a professor at the Minnesota School of Public Health and a leading expert on infectious threats, published a paper in *Foreign Affairs* entitled "Preparing for the Next Pandemic," in which he offered a scary scenario (2005). A pandemic would trigger a global disaster of the first magnitude, he explained. Fearing a shortage of vaccines, Western countries would at once nationalize their pharmaceutical industries; hospitals would be rapidly overwhelmed; flu would decimate health care personnel, leaving the sick unattended. In the face of a torrent of angry demands by the public for universal vaccination (notwithstanding the shortage of products and the number of adverse events), the armed forces (weakened by flu) would have to protect stockpiles as well as maintain law and order. The economy would suffer an instant meltdown. The full fury of the storm would send fear and terror to dizzying heights. One anxiety would reign supreme: to find staples that had become difficult to collect.

There was nothing new about this. "We have never been more vulnerable," wrote in 1996 Nobel Prize winner Joshua Lederberg, a scientific adviser to Bill Clinton (Lederberg 1996, p. 244). And yet, every emerging virus does not automatically trigger a pandemic. In 1976, the U.S. swine flu virus turned out to set in motion "the epidemic that never was" (Neustadt and Fineberg 1983).[1] The 1997 avian flu in Hong Kong and in Holland never transformed into a human disease. H5N1 does not replicate very well in humans.

Now, even though political and media fever has subsided since 2006, we know it may still be ratcheted up any time. The pages that follow tell how at the end of the twentieth century the language of national security ultimately won out over the multilateral notion of "human security" in framing infectious diseases, and how such a new frame led to a vertiginous descent into fiction, now spreading to numerous sectors of the public health system. Fantastic figures, unwarranted analogies, and, above all, scenarios are significant examples of fiction. At the same time, fiction may be useful to decision makers, for instance, in scenario planning.[2] The implications for governance in health is our topic, but the aim in this paper is rather limited. We introduce the notion of worst-case scenario, look for its origins, and see what is at stake when it is applied to the dissemination of infectious disease.

Reframing Infectious Diseases: "Human Security" or National Security?

So far infectious risk has generally been considered a hindrance by scholars of development. How did it become part of national security? Does an effective, clear-cut link between health and national security really exist? Historian Susan Peterson appears doubtful about answering in the affirmative to that question (Peterson 2007). Linking health to security always makes government more alert to bioterrorist risk (low probability notwithstanding) than, for instance, to tuberculosis (however high its rates of morbidity and mortality). To convince the sceptics may necessitate greater elaboration.

Let's go back to the end of the Cold War. At that point, a directive for a fresh theory of disease emergence combined with a broadened notion of security, now freed from the straitjacket of Cold War military threats (see below, p. 12). A geo-epidemiological and a geopolitical concept, health security integrated natural (e.g., epidemics) and intentional (e.g., bioterrorist) threats (Hamilton and O'Toole 2005). As those threats were likely to be deterritorialized or asymmetrical, they would be hard to perceive. Such "threats without an enemy" constituted what was called a "grey zone."

The notion of a "grey zone" appeared in 1993 contemporaneously with the notion of "emerging diseases." It is not until 2000, however, that disease was cited in a National Security Council (NSC) report and that HIV/AIDS was described as a national security threat by a White House National Security Strategy Report. At what was, in January 2000, the first meeting ever of the National Security Council dedicated to AIDS, Vice President Al Gore stated that the pandemic, having outrun the stage of a health crisis, had become a threat to global security and to the viability of African countries and African economic development. A few months later, President Bill Clinton followed in his vice president's footsteps when addressing AIDS policy in Nigeria.

As a matter of fact, the Americans had been outdone by Commonwealth sociological "brigades" who co-authored the promising new concept of "human security."[3] As early as 1966, Toronto psychologist William E. Blatz had stated that, far from being related to others' aggressive behaviour, one's sense of security was actually contingent on others' cooperation. Thirty years later, Lloyd Axworthy, Minister of Foreign Affairs in the Chrétien government (1996–2000), and George Maclean, a professor of political science at

the University of Manitoba, firmly put the individual at the centre of the web of security. Not the state and its territory but the well-being and the quality of life of the individual were now centre stage. And William Tow, Anthony Giddens, and Martin Shaw showed the pros and cons of globalization, linking their assessment to the impact of global policy on the well-being and security of the individual. In 1994, the UN Development Programme would provide through its *Human Development Report* a formidable outlet for these various ideas.

What is to be protected is not the state but, first of all, the individual. HIV/AIDS threatens the individual in many aspects: economic, health, ecological, social, and political. In short, the notion of human security marked the commencement of what Rasmussen would call the "sociological era" of international security studies (2004). Hunger, oppression, disease: everything could be a threat against human security. The trouble with the notion was precisely that: it was too broad to be really helpful. It was too vague—maybe studiously vague. The coalition supporting the notion—Norway, Canada, and Japan—scored some remarkable success, like the 1997 Ottawa Convention banning landmines. But, pushing aside the security of the state (and military threats), the notion of human security contributed to broadening the gap between health and security issues. In the absence of any common language between health and national security, governments held to the objective of health security while taking no action.[4]

Biosecurity, scenarios, and tabletop exercises such as the Top Officials exercises series (dubbed TOPOFF: 2000–09), *Dark Winter* (June 2001), or *Atlantic Storm* (January 2005), which were developed by the U.S. Department of Justice, the U.S. Department of Homeland Security, and the Johns Hopkins Center for Civilian Biodefense Strategies as preparation against imagined microbial threats, will provide a common language. As used here, fiction means more than narrative—it is a method of training. In scenario playing, people are immersed in an imagined universe. Through this immersion, players can adjust attitudes, cognition, and further acquaintance with problems that must be mastered if officials are to exert full control over the type of social stress associated with massive epidemic crisis. Underlying this adoption of scenarios is the two-pronged assumption that role-playing can facilitate the rational mastery of events whose cause, occurrence, and place are, by definition, unknown, and that preparedness is the mainstay of government response plans (Naylor et al. 2003, pp. 4, 18).

The Embrace of Fiction

The embrace of fiction and epidemic scenarios by the White House took a strange road.

After the Tokyo chemical attack of March 1995,[5] Clinton began reading a host of fictional accounts in which terrorists wield chemical and biological weapons. "Before September 11," wrote the American lawyer and philosopher George Annas, "most procedures for dealing with a bioterrorist attack against the United States were based on fiction. Former President Bill Clinton became engaged in the bioterrorism issue in 1997, after reading Richard Preston's novel "*The Cobra Event*" (Annas 2002, p. 1337).

A so-called bio-thriller, *The Cobra Event* is nothing but a hackneyed horror tale. Here's the pitch. Having been fired by a biotech firm, a disgruntled biologist takes revenge by disseminating in the Lower East Side a genetically modified virus causing horrible symptoms. The clinical picture shows an extremely rare genetic illness, the Lesch-Nyhan disease, a brain disease with violent behavioural troubles like auto-cannibalism plus kidney damage. Within a few days, the attack produces 32 cases. The mortality reaches a frightening 95% (by the way, exactly the same mortality rate caused by Ebola in Preston's previous book, *The Hot Zone* [1994]). Chased by a team of Centers for Disease Control and Prevention detectives and FBI special agents, and bitten by a contaminated rat, the lone terrorist, infected by his own evil virus, dies in a subway tunnel. The tone of the novel is alarming, and there is a lot of blood and gore—which is why it was so popular.

A couple of months later, in April 1998, Clinton convened a group of scientists to discuss bio-threats. Sitting in the Oval Office were Nobel Prize winner Josh Lederberg, geneticist Craig Venter, and biochemist and environmentalist Barbara Hatch-Rosenberg (from the Federation of American Scientists) among others. Some of them had provided Preston with useful scientific advice. And it was Venter who recommended Preston's novel to Clinton. Through Secretary of Navy Richard Danzig, Preston had contact with people in the FBI and in the Pentagon, where emergency terrorist scenarios were made by the dozen, especially since the Atlanta Olympic Games in 1996. Fictional account and fictional strategy increasingly came together. Not the aftermath of 9/11, but the second Clinton mandate was to be the watershed in this process.

Preston's novel seeks to produce the effect of reality at all costs (Sarasin 2006, pp. 85–100). Three appendixes try to convey the "reality behind *The Cobra Event*." A glossary explains the technical terms. The acknowledgements give credit to a long list of people who provided advice and information. A photograph shows the virus. By nesting historical accounts of bacteriological weapons inside a work of fiction, Preston glued fiction onto reality in order for the fictional account to work as a factual account.

By no means was he alone in his presentation of the new brand of terrorists as a sort of BioUnabomber. Historian Walter Laqueur (1996) and Lederberg among others shared the idea. The opening session of the Infectious Diseases Society of America annual meeting, held in September 1997, featured Preston as an invited speaker, along with Donald A. Henderson and Michael Osterholm, two prominent experts on bioterrorism (Cole 2003, p. 126). A close collaborator with D.A. Henderson, Tara O'Toole reviewed *The Cobra Event* in the *Public Health Reports*, the journal of the U.S. Public Health Service. And, believe it or not, Preston had a hearing as an expert in the U.S. Senate in 1998.

Congressmen would not be long in using fiction in the interests of policy-making. In June 2001, a scenario called *Dark Winter*, authored by the Johns Hopkins Center for Civilian Biodefense headed by D.A. Henderson, was convened at Andrews Air Force Base near Washington. It staged the global spreading of a smallpox virus that had been "weaponized" by the Iraqis much as the Soviets had done through their illegal Biopreparat Program, which had been developed despite the 1972 Convention banning biological weapons (Johns Hopkins 2001; O'Toole et al. 2002). Taking advantage of the fuss over *Dark Winter* in Washington, the U.S. Congress voted extra funding for stockpiling smallpox vaccine and domestic response training. Awfully worried, Vice President Richard Cheney tabled the issue of smallpox vaccination at the next meeting of the NSC (Drexler 2002, p. 263).

The Spread of Epidemic Scenarios

As noted above, the use of scenarios and exercises has spread rapidly within health administrations over the past ten years. We have even seen the sharing of scenario-based strategies among Western governments. In September 2003, for instance, Health Canada initiated and monitored *Global Mercury*, an alert and crisis management exercise involving Canada, Mexico, the United States,

France, Italy, Germany, the United Kingdom, Japan, the European Commission, and the World Health Organization.

From the start, fiction was treated as a tool for devising creative policies. One of the most intriguing early instances of epidemic scenarios took place in March 1998 at Blair House, just on the other side of Pennsylvania Ave. facing the White House. Although secret at the time (Blumenthal 2003, p. 657), this episode has been recounted more than once since. Richard Clarke, head of the Counterterrorism Security Group of the NSC, convened a tabletop exercise simulating a cabinet meeting. Officials from the State Department, CIA, FBI, Departments of Justice, Defense, Health and Human Services, and Energy, Federal Emergency Management Agency (FEMA), and Office of Management and Budget sat around a series of U-shaped tables in a ballroom. According to Clarke, exercises like this had been held for years within the NSC, but this one was the first done by the U.S. Cabinet (Clarke 2004, p. 164). The author of the exercise was William A. Haseltine, a geneticist who had founded Human Genome Sciences, a start-up that specialized in the sequencing and patenting of genes, and who had been a contractor for the Clinton administration. The exercise began with a report on a spreading infection in the U.S. Southwest, followed by a second report stating that patients were diagnosed with Marburg or Ebola fever (fairly unrealistic, but who cared?). What was to be done? Huge gaps unfolded in the discussion about logistics, legal authority, or medical care. No one seemed to be in charge. “No Cabinet member knew the answer […] it was clear that there was no plan,” concluded Clarke. People reacted with alarm to this finding. “The officials who emerged from the darkness of the Blair House reception room and piled into their cars were a shaken group,” two National Security Council staffers later reported (Benjamin and Simon 2002, p. 255).

One of the largest emergency drills ever conducted—TOPOFF—took place in May 2000. The U.S. Congress directed the Justice Department and FEMA, together with the National Security Council, to stage a big exercise in order to test the ability of the administration to control consequence management of an attack that used weapons of mass destruction. Costing around $10 million, the scenario was written by a defence contractor, SAIC (Science Application International Corporation). It simulated a mustard gas attack in Portsmouth, New Hampshire, where three thousand government officials had been mobilized, and the dissemination of a plague aerosol within the

premises of the Center for the Performing Arts in Denver, Colorado, where twenty-five hundred officials participated. News alerts and press briefings were covered by a team of pseudo-reporters. The response in Portsmouth went smoothly, but Denver was in chaos. City and state had a fierce argument about a distribution plan for antibiotics, several hospitals reached full capacity, and so did morgues. Tensions arose between public health officials, hospital representatives, and the FBI. The medical system, government offices, and society itself were collapsing (Inglesby et al. 2001; Miller et al. 2001, pp. 270–76; Guillemin 2005, p. 163).

A second edition of this scenario, TOPOFF2, was staged in May 2003, costing $16 million and involving 19 federal agencies, the Red Cross, and the government of Canada for five days. It simulated a radioactive blitz in Seattle and an outbreak of pneumonic plague in Chicago. The results of this exercise were immediately classified (Guillemin 2005, p. 180).[6]

Such scenarios were later dismissed by Centers for Disease Control (CDC) scientists, among others, as full of exaggeration. According to the critics, the contagion rates were far too high and too little emphasis was placed on proven simple ways to curtail epidemics (home care, masks, hand washing, avoiding hospitals where transmission rates would soar). However, as one historian noted, they served well as political rhetoric (Guillemin 2005, p. 164). It is to this political rhetoric that we now turn.

Vulnerability and Preparedness

In the beginning, the Bush administration was content with the Clinton doctrine warmed over. To some extent, the 2002 *National Security Strategy of the United States* parroted Al Gore and Bill Clinton's last vision. Within a few months, however, health would turn into a high security priority (Nye 2002).

Heightened alarm did not fail to profit by the anthrax attacks of the fall of 2001. Between October 4 and November 11, 22 persons were diagnosed with anthrax. Eleven contracted cutaneous anthrax and survived. Among the other 11 who contracted inhalation anthrax, five died. Two batches of contaminated envelopes were mailed from Princeton, New Jersey. The first around September 18 to mass media offices in Boca Raton, Florida, and New York City; a second set around October 8 to Democrat senators Patrick Leahy and Tom Daschle on

Capitol Hill. The dispersal of anthrax spores through the mail put postal workers at risk. Out of the 22 people infected, nine were U.S. postal employees. Out of those nine, two died. Decontamination of the U.S. postal facility at Brentwood took two years. "The anthrax letters," wrote Leonard Cole, "moved concerns about bioterrorism from theory to reality." They helped to create wide recognition of "the country's vulnerability to germ weapons" (2003, p. 239).

Anthrax attacks also induced a period of panic in American policy. The White House, in particular, trembled and felt the approach of doom. "I think the seminal event of the Bush administration was the anthrax attacks," a close aide to the President told Newsweek journalist Jacob Weisberg. Weisberg went as far as to write that "without the anthrax attacks, Bush probably would not have invaded Iraq" (2008, p. 189). Surely some further investigation is needed in order to corroborate such a statement, but it is true that the sense of vulnerability in Washington at that time was at record levels. In the immediate aftermath of 9/11, the Secret Service began monitoring the air inside and outside the White House. Cheney never travelled without a full biohazard protective suit. He spent most of his time in the national security bunker at Raven Rock Mountain near Waynesboro, Pennsylvania, under hundreds of feet of granite. Like Bush, he feared the "second wave." The anthrax letters were just that (Daadler and Lindsay 2003, p. 118).[7] When, on October 23, they were told that a man in a hospital in Florida was believed to have smallpox, it confirmed their impression that a pervasive and peculiar perversity was controlling events. They were relieved to hear that the poor man suffered only from syphilis.

Vulnerability was not a hidden perception. Some sniffed vulnerability in every imported virus. Former Soviet biological scientists were a transcendent threat to America. There was a stir and flutter in the White House. "Quite suddenly the entire enterprise of biology became potentially suspect," according to *Nature* (The End of Innocence 2001, p. 235). And yet, there is perhaps a gulf between a vulnerability and a threat. "A vulnerability," notes Richard Falkenrath, "is a situation of being open to harm; a threat is the known or suspected presence of an actor with the ability, will and motive to inflict harm…in general, vulnerabilities make poor proxies for threats." For example, a British nuclear strike could devastate the United States—but who is going to believe that Great Britain plans such an attack (Falkenrath et al. 1998, p. 11)?

Assuming anthrax attacks did effectively signal a turning point in the U.S. officials' mindset, to judge the new policy of preparedness in the glare of the rough and tumble of politics would nonetheless be mystifying.

To begin with, preparedness used to characterize exercises staged by officials in order to test personnel and procedures. At the end of the 1990s, a shift towards strengthening epidemiological surveillance and more proactive alert systems occurred. Five elements characterize preparedness: planning (response plans), detection (surveillance), identification (laboratories), neutralization (quarantine, vaccines, antibiotics), and rehabilitation. As a result, public health was put in motion on a wider scale. Programs traditionally regarded as separate—social policy, health care, environment, defence, public health itself—were now seen as closely related fields. Bioterrorist threats came to be considered as part and parcel of emerging infectious risk; and, at the same time, response plans had to take various components of civil society on board and stick to a multisector approach as numerous levels of government, administration, and expertise would be involved in crisis and consequence management.

There was nothing especially novel about these ideas. It has been said that the Bush administration "intensified Clinton-era politics already in place for national security and defense against biological weapons" (Guillemin 2005, p. 167). Rightly so. Bush—and Cheney—leaned heavily on TOPOFF2000 and *Dark Winter* (June 2001) (Weisberg 2008, pp. 189–90). Throughout the crisis, the CDC and local health services drew on drugs (ciprofloxacin) stockpiled by the Clinton administration. Clinton had followed biological security issues with the closest personal attention since 1998. His administration had shown concern even earlier, since around 1996. Besides human security, infectious diseases were framed to fit into national security concern for the sake of bioterrorism response. Men from the government like William Cohen (Secretary of Defense) or Fred Iklé (Under Secretary of Defense for Policy), Louis Freeh (the director of the FBI), former director of the CIA James Woolsey, strategists like Joseph Nye, all heard warning bells. "These are not the alarming voices of the past," wrote Richard Falkenrath (who later participated in authoring the 2003 Homeland Security Act [Guillemin 2005, p. 237]), "not the lunatic fringe for whom the sky has always been falling. These are respected mainstream national security officials and experts" (Falkenrath et al. 1998, p. 4).

And so a dash of alarm regarding nuclear-biological-chemical threat was allowed to insert itself into the moral universe of "human security." It may be that Susan Peterson draws too clear-cut a divide between human security and national security. Both notions actually constitute a "grey zone" that will keep rising around microbial threats like mist from a marsh.

Fiction Reaches its Climax: Worst-Case Scenarios

According to Rosner and Markowitz, this alarmist surge in biological security policy thinking permanently altered the political and cultural life of America (2006). But did it really?

Academia has recently started to turn its attention to worst-case scenario: witness the penultimate book by Cass R. Sunstein, *Worst-Case Scenarios* (2007). A constitutional lawyer, Sunstein joined Harvard in 2008. He was a law professor at the University of Chicago for years, and a colleague of Barack Obama. In January 2009, the President named him administrator of the Office of Information and Regulatory Affairs, a key part of the federal government aimed at implementing the laws that Congress passes (Shear 2009, p. A12).

In his book, Sunstein focuses on what he calls the "excessive concern with worst-case scenarios" to which, he says, the Bush government (but also Al Gore in relation to climatic change) was "attuned" (2007, pp. 37, 267, 276). In this anxiety prone ideology riveted upon perspectives of catastrophe and disaster, one key instance is Dick Cheney's "one percent doctrine," which regards a worst-case scenario as a certitude, even though it cannot be attributed to a distinct probability.[8] The irony, considering the fuss the United States has made over EU environmental legislation, is that this doctrine embodies in fact a super-precautionary principle (ibid., p. 123): even if a terrorist attack is wholly improbable, it must nevertheless be prevented. One has to be thoroughly prepared for the worst eventualities, no matter their probability. Another irony is that this doctrine originated not in some executive leeway but in federal courts, which in the 1980s judged that governmental agencies should take into account worst-case scenarios in environmental disasters (oil spills), a doctrine Reagan fiercely rejected at the time.

Now, to view this doctrine as bred in anxiety alone would be misleading. Worst-case scenarios were nurtured in two independent

fields narrowly connected to our subject: military planning and the theory of emerging viruses.

The end of the Cold War completely upset the U.S. Defense Department notion of risk. From "a known and clear danger" (the Soviet threat), risk entered the grey zone of "uncertainty." Military planners did not bother anymore exploring the enemy's interests, intents, and capabilities. Any risk, however improbable or remote, was now considered worthy of attention. Most importantly, the "uncertainty hawks" viewed the strategic environment as unpredictable. More than uncertainty (a known threat to which no probability can be assigned), more than risk (a known threat to which a distinct probability can be assigned), and even more than ignorance (the character and probability of the threat remain unknown), we are dealing here with deliberate rejection of probability. Bending to realism has become but an amicable grimace: strategic thinking, it was said, would once and for all free itself from the "tyranny of the plausible" (Conetta and Knight 1998, pp. 32–34).

In contrast to strategic planning (or prediction), scenario planning is about giving people a more piercing gaze capable of seeing and decoding weakly articulated, emerging realities. Admittedly, disruptive change is hard to quantify. Scenario planning stands apart from probabilities, and that is one big contrast to prediction and projection. As said in the presentation of the *U.S. National Planning Scenarios* developed by the Department of Homeland Security in 2006 to address all-hazard incidents, which include terrorism, natural disasters, and health emergencies, "The scenarios are not intended to be exhaustive or predictive. Nor do they address every possible situation, but in combination, they illustrate the tasks and capabilities required to respond to a wide range of major events" (U.S. Dept. of Homeland Security n.d.). Now, the Bush administration went further afield. Nothing fostered the flight from plausibility more than the doctrine of pre-emptive action—some might call it "prophylactic aggression"—forged by Dick Cheney in 2001. As Jacob Weisberg has rightly noticed, under Cheney's relentlessly organizing hand, this central doctrine went to the root of the global geopolitical strategy. Informed by a Hobbesian concept of geopolitics and alarm at America's Achilles heel in relation to biological threats, the doctrine gathered together all the threads in the worldviews of the neocons, Cheney, and Bush (2008, pp. 197–99). Preparedness, in particular, dovetailed nicely with Bush's warning at the West Point graduation

ceremonies in June 2002, that "if we wait for threats to fully materialize, we will have waited too long."

Another opener, quite different in character, was the theory of emerging viruses. In 1967, Surgeon General William H. Stewart exclaimed, "The chapter of infectious diseases is closed." However, worrying signs soon destroyed this naive optimism. There was the first appearance of Ebola hemorrhagic fever in Zaire in 1976, and in the same year, Legionnaire's disease at the American Legion convention in Philadelphia. And especially, there was the identification of the first cases of AIDS in 1981–83, a previously unknown "new" disease. With the sudden increase in multiresistant tuberculosis in New York City in 1985–91 and fears concerning the spread across Europe of resistant forms of the bacillus from Russia, the price of complacency, which had held sway up to that point, became clear. The entire discourse (concepts and metaphors) surrounding the fight against infectious disease began changing. In the wake of Pasteur and Koch, a "war on microbes" was pursued. This doctrine has rendered immeasurable service (public health, sanitation), but the increase in cases of emerging or re-emerging diseases since 1970 has changed our view of things. Today, we understand that social and cultural factors play an important role alongside biological determinants. "Microbial circulation," wrote virologist Stephen Morse who coined the very notion of emerging viruses at the first conference on the topic held in Washington in 1989, "is the main factor of virus emergence. Human beings are often complicit of this process of circulation" through deforestation, population movement, urban crowding, refugee camps, wars, and even climate change.

We will refrain in this paper from expanding on the historical account of the theory of emerging diseases, whose immediate antecedent was the "ecological point of view" upon the relations between man and infectious diseases (Burnet and White 1972, pp. 1–21; Fenner 1970). Two years after the 1989 conference, the Institute of Medicine (IoM) of the U.S. Academy of Science established a committee for the study of emerging disease. In 1992, the IoM warned the medical community about the re-emergence of infectious diseases. "By no means are we prepared to respond to a highly virulent influenza pandemic that cannot fail to happen," wailed Joshua Lederberg (Institute of Medicine 1992; Stone 1992, p. 540). In 2003, the IoM "Forum on Emerging Infections" (in which personalities in the economy and politics joined scientists) was renamed the "Forum on Microbial Threats," thereby

institutionalizing the link between emergence and threat. To this pair, Don Henderson later added bioterrorism.

Far from drifting into absurdity, the worst-case scenario is a well-developed concept. Rhetoric has become doctrine. Less seismic than chaos, worst-case scenario refers to some kind of event that is not irrational in itself as much as unpredictable, since it is altogether impossible to locate in space, time, and character (Rosset 1971, pp. 74, 98–99). Epidemics, accidents, and criminal intention combine various components of fear: chance (unknown probability), perversity of nature (brutal mutation of a virus, apparently decent people suddenly turning out to be a criminal), and non-existent entity (preparation to fight against "a disease that does not exist"). One example fits this definition perfectly: pandemic influenza. For, considering this threat, our predictive capabilities are very limited in relation not only to its "where" and "when," but also to the antigenic composition of the virus or to its pathogenicity. Ours is an impossible situation where we are forced to rely on uninterrupted vigilance, surveillance, and preparedness *in the absence of any epidemic*, while lacking any means to determine when and where the menace will finally materialize.

The worst aspect of worst-case scenarios is that they ultimately defy one to set down any general method for preparedness.

The difference with "risk society" is of the highest consequence. In a "risk society," social or technological systems ultimately provided the tools to repair an accident or contingency. By contrast, in a worst-case scenario, all occurrences are acts of God, whose biological (for instance, viral mutation or reassortment) or social (for instance, terrorism) consequences are not susceptible of prediction. "Risk society" never gave up hope for an overarching rationality, a precautionary principle, or international cooperation in the form of the Kyoto Protocol or the Chemical Weapons Convention of 1992, for instance. "Threat society," on the other hand, flatly did. Health security now has to work to counter the brutal and savage attacks of unknown viruses and jihadists. As a result, figuring out a protective shield looks like fighting a battle against desperate odds.

A Case in Point: The Smallpox Vaccination Campaign, 2002–3

Immediately after the first wave of anthrax letters in October 2001,

Cheney came to fear something worse: an attack involving smallpox. Vaccination against smallpox had been stopped in the United States in 1972, and, as immunity appears to last less than thirty years, the vast majority of the U.S. population was therefore susceptible. The Vice President believed that Iraq's potential to launch such an attack necessitated the universal vaccination of the American population. On hearing this, Don Henderson was greatly alarmed. Henderson was Bush's scientific adviser and the head of the Office of Emergency Preparedness in the U.S. Health and Human Services Department. He tried hard to persuade Cheney not to set such a terrible idea in motion. Henderson was sceptical about Iraq working with smallpox, and he described to Cheney not only the deep ignorance among doctors (smallpox was no longer taught in the medical curriculum[9]), but above all the reactions, normal and abnormal, to vaccination and the number of casualties to be expected from an inoculation program (one or two deaths out of every million people). Henderson thought he had persuaded Cheney. But, impervious to the obvious, Cheney decided the entire country had to be vaccinated. Fortunately, 24 hours later, this plan was called off. Bush had overruled his vice president (Weisberg 2008, pp. 192–93). Another program had been decided on: to immunize five hundred thousand health care workers and first responders. This program was announced on December 13, 2002. As Barbara Tuchman said, "error multiplies, never retreats" (1984, p. 383). And this program would encounter a lot of problems.

In the beginning, local health services were not opposed. California, Arizona, and Connecticut showed interest. (It must be said, though, that most of the cooperative states were red states.) Yet the campaign would soon be abruptly stopped. On March 2003, after twenty or so adverse events had occurred, a number of states suspended vaccination. The number of inoculated persons dropped, and the numbers never recovered. In May, after George Bush declared "mission accomplished," the immunization campaign's numbers dropped a second time. One year later, only 39,579 civilians (plus 631,000 armed forces personnel) had been vaccinated. Short of candidates, the campaign stopped spontaneously. Among various hurdles, vaccine safety had been the most critical to its fate. But its political underpinnings had badly hit the government initiative, too. Washington had never said smallpox vaccination was linked to the odds-on war against Saddam Hussein. Yet Julie Gerberding (director of the CDC), Tommy Thompson (HHS Secretary), and Will Orenstein

(head of the national vaccination program at the CDC), not to mention senators Judd Gregg (R-New Hampshire) and Bill Frist (R-Tennessee), disseminated allusions to such link. Furthermore, the vast majority of public health specialists, CDC included, strongly favoured ring vaccination. The CDC was uncomfortable with this tiresome campaign, which, according to William Bicknell, a professor of public health at Boston University, had been imposed on it by the White House.[10]

The reader might say that no matter how you slice it, it's only media hype in addition to folly under full sail. No doubt it was, if only because the vaccination campaign awakened some sleeping dogs, among which was the burning issue of the legitimacy of the state. The campaign was supposed to set in operation preparedness through a method—administrative coercion—squarely opposed to one of the bases of what Philip Bobbitt calls the "market state" (by which he means the latest form of the long evolution of the Western state), namely, more individual options and less coercion by law. Bush and Cheney tried, in addition, to liquidate the tiresome business of transparency and communication, and this was odious to many. The White House behaved as if it wanted to put Humpty Dumpty back on the wall—that is, to protect the state first, and individual (human) security second. This was almost unanimously rejected.

And in the end, there was the issue of expertise. Since the 1990s, smallpox has ranked in the United States among the most dangerous germs. In June 2003, the European Medicines Evaluation Agency (EMEA) classed it as the highest level of terrorist risk, and France deemed it a "plausible" menace, although no serious survey supports the probability of its deliberate reintroduction. Here lies the critical point. How probable is the risk? A paper that went online on December 19, 2002, in the *New England Journal of Medicine*, supported the preventive vaccination of health care workers, although it denied any value to universal immunization. Vaccinating first responders was supposedly justified because of the probability of an attack, a probability that seems '"plausible" to the authors of the paper (Bozzette et al. 2003). According to the Advisory Committee on Immunization Practices (ACIP)—which strongly opposed the smallpox vaccination campaign—the National Council of Vaccination had been informed in 2002 that "the risk was low but not zero." How could we put a figure on this, asked the committee (Baciu et al. 2005, p. 88)?

It is a pity that no historian was present at the meeting of the ACIP. He or she could have set forth for the distinguished gentlemen of the committee the outstanding facts of the swine flu scare of 1976. This virus was similar to the 1918 virus (H1N1). Some raised the alarm, among them David J. Sencer, the then director of the CDC. Yet no actual epidemic came, not the faintest sign of any gathering storm appeared. How then could people put a figure on a "vague possibility"? Weak expertise was easily outdone by public opinion—the U.S. Congress, the mass media—which propagated recollections of the "Spanish" flu. Assistant Secretary of Health Theodore Cooper, himself a doctor, recalled the bad old days when his father was a kid in Pennsylvania, where in 1918 soldiers buried *en masse* victims of the flu. And suddenly, memory true to the past was putting bright colours on what was only fiction or fantasy. A past reality gave the present unreality poster paints. The past had prefigured the present—and this prefiguring gave flesh and blood to what was nothing yet but a mind event.

Universal vaccination was adopted on the grounds of emotion more than science, and then was halted quickly because of 532 cases and 25 fatalities supposedly related to vaccination. In 1976, as in 2002, probability was only of marginal relevance to the decision to launch a mass vaccination campaign. "Expertise counted for something," wrote historians of the swine flu, "but only in support of a key subjective element." In the absence of any objective probability, the images of the "Spanish" flu that people recalled (or believed they recalled) still won the day (Neustadt and Fineberg 1978, pp. 87–92). Historical imagination—or historical fantasy—had got the better of statistics and probability in the decision-making process.

It has been said that throughout the swine flu episode, science was the central brain and controlling power during two brief moments only: the identification of the virus and the confirmation of an increasing risk of Guillain-Barré linked to vaccination.[11] "Everything else was political" (Dowdle 1997, p. S71). True, 1976 was an election year, and Gerald Ford already had enough on his shoulders without wanting to seem indifferent to the biological security of the American people. Moreover, as of the next year, the federal government would start elaborating anti-pandemic plans. Washington would thus take charge of influenza, along the same lines it had appropriated measles, polio, and rubella in the past. An "epidemic that never was" thus resulted in institution building if nothing else. This process of broadening

appropriation by the federal government of disease control was repeated under Clinton's and George W. Bush's presidencies for different kinds of risk. When we add to this development that political casuistry (subjective probability and memory) increasingly got the better of technical casuistry (objective probability and history), we can understand how, at the end of the day, worst-case scenarios became embedded into the politics of microbial threats and soaked through government and expertise like melted butter.

Notes

1. Neustadt and Fineberg (1983) was previously published as Neustadt and Fineberg (1978).

2. Kahn underscored this use of fiction (1962, pp. 143–45, 155–58, 172–75).

3. See the bibliography in Brower and Chalk (2003, pp. 4–5). See also Waters (2001).

4. See the 2001 interview with Joshua Lederberg in McBride (2003, pp. 130–34).

5. On March 20, 1995, the Aum Shinrikyo Sect disseminated sarin (aka GB), a nerve agent, in plastic bags placed in five different subway trains converging on the centre of the city of Tokyo. The intoxicating vapour resulted in 12 deaths, 1,000 wounded, and more than 5,500 hospitalizations (Olson 1999; Tucker 1999, p. 317).

6. As the series went on, TOPOFF3 ($21 million) would be staged in April 2005, TOPOFF4 ($25 million) in October 2007, and TOPOFF5 in April 2009.

7. After 9/11, many in Washington speculated about a second, more devastating, possibly biological terrorist attack. After having tried to spread confidence in America's preparedness during the first half of September, the Bush administration became more concerned. When, on October 4, Secretary of Health Tommy Thompson informed George W. Bush that the death of photo editor Robert Stevens, from Boca Raton, Florida, was caused by anthrax, some of the President's aides said "they saw fear in his eyes for the first time since 9/11" (Thompson 2003, 92). Cheney vented his fears of a biological attack by al-Qaeda on PBS Network a few days before the anthrax letter was opened in Tom Daschle's office on Capitol Hill (ibid., p. 113).

8. In case of a catastrophic event, a 1% probability is as good as any higher probability: it is actually equal to a certitude (Sunstein 2007, p. 1).

9. According to an experiment staged in Pittsburgh in February 2000, only one doctor out of 17 was able to diagnose smallpox when he saw symptoms and patient's photos (Smithson 2001; Smithson is a nonproliferation scholar at the Henry L. Stimson Center, Washington).

10. Quoted in Dyer (2003, p. 1010). In September 2002, CDC favoured ring vaccination (U.S. Centers for Disease Control 2002).

11. The Guillain-Barré syndrome (one case out of one million people inoculated) brings about a paralysis of lower and upper limb and often of the trunk. The paralysis can be fatal in case respiratory muscles are hit. This correlation between vaccination and Guillain-Barré syndrome has been questioned (Freedman and Stark 1999).

References

Annas, G.J. 2002. Bioterrorism, Public Health, and Civil Liberties. *New England Journal of Medicine* 346 (17): 1337–42.

Associated Press. 2005. WHO: Impossible to Estimate Pandemic Deaths. 30 September.

Baciu, A., et al. 2005. *The Smallpox Vaccination Program. Public Health in an Age of Terrorism*. Washington, DC: National Academies Press.

Benjamin, D., and S. Simon. 2002. *The Age of Sacred Terror*. New York: Random House.

Blumenthal, S. 2003. *The Clinton Wars*. New York: Farrar, Strauss & Giroux.

Bozzette, S.A., et al. 2003. A Model for a Smallpox-Vaccination Policy. *New England Journal of Medicine* 348 (5): 416–25.

Brower, J., and P. Chalk. 2003. *The Global Threat of New and Reemerging Infectious Diseases. Reconciling U.S. National Security and Public Health Policy*. Arlington, VA: Rand Science and Technology.

Burnet, F.M., and D.O. White. 1972. *Natural History of Infectious Disease*. 4th ed. Cambridge: Cambridge University Press.

Clarke, R.A. 2004. *Against All Enemies. Inside America's War on Terror*. New York: Free Press.

Cole, L.A. 2003. *The Anthrax Letters. A Medical Detective Story*. Washington, DC: Joseph Henry Press.

Conetta, C., and C. Knight. 1998. Inventing Threats. *The Bulletin of the Atomic Scientists* 54 (2): 32–34.

Daadler, I.H., and J.M. Lindsay. 2003. *America Unbound. The Bush Revolution in Foreign Policy*. Washington, DC: Brookings Institution Press.

Dowdle, W.R. 1997. The 1976 Experience. *Journal of Infectious Disease* 176, Suppl 1: S69-72.

Drexler, M. 2002. *Secret Agents. The Menace of Emerging Infections*. Washington, DC: Joseph Henry Press.

Dyer, O. 2003. UK Government Denies that Smallpox Vaccination Plans Have Fallen Short of Target. *British Medical Journal* 327 (1 November): 1010.

The End of Innocence. 2001. *Nature* 414 (6861): 235.

Falkenrath, R., R.D. Newman, and B.A. Thayer. 1998. *America's Achilles' Heel: Nuclear, Biological, and Chemical Terrorism and Covert Attack*, BCSIA Studies in International Security. Cambridge, MA: The MIT Press.

Fenner, F. 1970. The Effects of Changing Social Organisation on the Infectious Diseases of Man. In *The Impact of Civilisation on the Biology of Man*, ed. S.V. Boyden, 48–76. Toronto: University of Toronto Press.

Freedman, D.A., and P.B. Stark. 1999. The Swine Flu Vaccine and Guillain-Barré Syndrome: A Case Study in Relative Risk and Specific Causation. *Evaluation Review* 23 (6): 619–47.

Guillemin, J. 2005. *Biological Weapons. From the Invention of State-Sponsored Programs to Contemporary Bioterrorism*. New York: Columbia University Press.

Hamilton, D., and T. O'Toole. 2005. Facing up to the Bioterror Threat. *International Herald Tribune*, European edition, 31 January.

Inglesby, T.V., R. Grossman, and T. O'Toole. 2001. A Plague on Your City: Observations from TOPOFF. *Clinical Infectious Disease* 32: 436–45.

Institute of Medicine. 1992. *Emerging Infections: Microbial Threats in the United States*. Washington, DC: National Academy Press.

Johns Hopkins Center for Civilian Biodefense, Center for Strategic and International Studies, ANSER, and Memorial Institute for the Prevention of Terrorism. 2001. Dark Winter. Bioterrorism Exercise, Andrews Air Force Base, 22–23 June, final script.

Kahn, H. 1962. *Thinking about the Unthinkable*. New York: Horizon Press.

Laqueur, W. 1996. Postmodern Terrorism. *Foreign Affairs* 75 (5): 24–36.

Lederberg, J. 1996. Infection Emergent. *Journal of the American Medical Association* 275 (17 January): 243–45.

McBride, D. 2003. *Bioterrorism: The History of a Crisis in American Society*, vol. 2. New York: Routledge.

Miller, J., S. Engelberg, and W.J. Broad. 2001. *Germs. Biological Weapons and America's Secret War*. New York: Simon & Schuster.

Naylor, D., et al. 2003. *Learning from SARS. Renewal of Public Health in Canada, a report of the National Advisory Committee on SARS and Public Health, October 2003*. Ottawa: Publications Health Canada.

Neustadt, R.E., and H.V. Fineberg. 1978. *The Swine Flu Affair. Decision-Making on a Slippery Disease*. Washington, DC: U.S. Department of Health, Education, and Welfare.

———. 1983. *The Epidemic That Never Was: Policy-Making and the Swine Flu Scare*. New York: Vintage Books.

Nye, J. 2002. Health Turns into a Security Priority. *International Herald Tribune*, European edition, 2 September.

Olson, K.B. 1999. Aum Shinrikyo: Once and Future Threat. *Emerging Infectious Disease* 5 (4): 513–16.

On a Wing and on a Prayer. 2005. *Nature* 435 (7041): 385–86.

Osterholm, M.T. 2005. Preparing for the Next Pandemic. *Foreign Affairs* 84 (4): 24–37.

O'Toole, T., M. Mair, and T.V. Inglesby. 2002. Shining Light on "Dark Winter." *Clinical Infectious Disease* 34: 972–83.

Peterson, S. 2007. Human Security, National Security, and Epidemic Disease. In *HIV/AIDS and the Threat to National and International Security,* ed. R.L. Ostergard, Jr., 36–63. London: Palgrave.

Rasmussen, M.V. 2004. 'It Sounds Like a Riddle': Security Studies, the War on Terror and Risk. *Millennium* 33 (2): 381–95.

Rosner, D., and G. Markowitz. 2006. *Are We Ready? Public Health since 9/11*. Berkeley: University of California Press.

Rosset, C. 1971. *La Logique du pire*. Paris: Presses universitaires de France.

Sarasin, P. 2006. *Anthrax. Bioterror as Fact and Fantasy*. Cambridge, MA: Harvard University Press.

Shear, M.D. 2009. Obama to Name Lawyer Friend to Regulatory Affairs Position. *The Washington Post*, 8 January, A12.

Smithson, A. 2001. A Review of Federal Bioterrorism Preparedness Programs from a Public Health Perspective. The U.S. House Committee on Energy and Commerce, Subcommittee on Oversight and Investigations, 10 October.

Stone, R. 1992. IoM weighs in on Microbial Threat. *Science* 258: 540.

Sunstein, C.R. 2007. *Worst-Case Scenarios*. Cambridge, MA: Harvard University Press.

Thompson, M.W. 2003. *The Killer Strain. Anthrax and a Government Exposed*. New York: HarperCollins.

Tuchman, B.W. 1984. *The March of Folly. From Troy to Vietnam*. New York: Ballantine Books.

Tucker, J.B. 1999. Bioterrorism: Threats and Responses. In *Biological Weapons: Limiting the Threat*, ed. J. Lederberg, 283–320. Cambridge, MA: The MIT Press.

U.S. Centers for Disease Control. 2002. Smallpox Response. Plan and Guidelines. Draft 3.0-9/21/2002.

U.S. Department of Homeland Security. n.d. Strengthening National Preparedness: Capabilities-Based Planning. Fact Sheet.

Waters, S. 2001. *Bibliography on Human Security Prepared by the Harvard Program on Humanitarian Policy and Conflict Research* (August).

Weisberg, J. 2008. *The Bush Tragedy*. New York: Random House.

An Elementary Consideration of Humanity? Linking Trade-Related Intellectual Property Rights to the Human Right to Health in International Law[1]

Lisa Forman

Abstract

This paper explores methods of achieving linkage in international law between the international right to health and the rules in the World Trade Organization (WTO) Agreement on Trade-Related Intellectual Property Rights (TRIPS). The paper assesses the location of the right to health within international law's accepted hierarchies. It argues that despite the lack of explicit recognition of the right to health within these categories, their content nonetheless suggests that international law places reasonable limits in protection of human health and life on all human conduct, including trade and commerce. The paper turns to explore the utility of interpretive legal methods for achieving recognition of the priority of human rights to life and health within existing WTO law and dispute settlement processes. It concludes that raising the priority of health within WTO law requires a substantive reordering of the normative priorities which drive trade rules, including by advancing recognition of health within international law's accepted hierarchies and advancing integration of TRIPS into the broader normative system of international law.

Introduction

In this paper, I explore methods of achieving linkages between the right to health in international law and the World Trade Organization (WTO) Agreement on Trade-Related Intellectual Property Rights (TRIPS). I argue that linkage between trade and human rights law is inherent within international law itself, which prescribes universal

interpretive methods for treaties in the Vienna Convention on the Law of Treaties. I also explore the idea that international law places a universal minimum duty on all states to abstain from gross violations of human rights to life and health, including through a set of prioritized duties including jus cogens (peremptory norms), obligations ergo omnes (state duties owed "to all"), and duties in the Charter of the United Nations. The implication of these duties is that trade (like all human conduct) is subject to reasonable limitations in protection of health and life, including in relation to medicines, a key health care intervention. I do not suggest that health is explicitly recognized within international law's prioritized norms, but rather that health as an interest could appropriately fall within their conceptual domain. I suggest that the absence of explicit recognition of health in these norms is more reflective of the de facto neglect of health as a social right within international legal theory and practice than of any categorical or substantive basis for its exclusion. While international human rights law does not currently recognize the right to health within these prioritized norms, it does prohibit gross violations of human rights. I argue that this legal duty is reflected in provisions throughout WTO agreements which limit trade where doing so is necessary in order to protect human life and health.

This paper therefore advances arguments grounded in international legal theory and praxis about what the law should be. The aim of these arguments is to attempt to raise the priority of health both within and outside the human rights domain. This effort is particularly apposite given that existing interpretations of TRIPS by WTO dispute settlement panels have prioritized trade and intellectual property rights at the expense of collective health needs. I suggest that raising the comparable or enforceable priority of health in these rules requires a substantive reordering of the normative priorities which drive trade rules, and that argumentation for the elevated priority of health within international law's existing hierarchies provides a practical strategy for achieving this goal, including within decision making by WTO dispute settlement panels and domestic governments.

The paper proceeds by, first, exploring the conflict between human rights and trade related to pharmaceutical patents; second, examining the location of health within the established superior norms within international law contained in peremptory norms, ergo omnes duties, and article 103 of the UN Charter; third, exploring TRIPS and its jurisprudence in light of international law on treaty interpretation,

important changes in state agreement, and practice relating to medicines; and finally, exploring the implications of the foregoing legal analysis for rights-based reforms of TRIPS.

1. Conflicts between the Right to Health and Trade-Related Intellectual Property Rights

Legal protections of the right to health and trade-related intellectual property rights relate to similar subject matter (namely, access to medicines) and impose obligations that appear to be contradictory, since their respective norms appear to point to incompatible choices between their respective duties (ILC 1996, para.2; Hestermeyer 2007, p. 175). On the one hand, TRIPS requires all WTO members to provide twenty-year exclusive patent protection on pharmaceuticals (TRIPS, article 28.1.a and b), which serve to raise drug prices and restrict domestic policy options for accessing affordable medicines. On the other hand, the right to health in international law requires governments to provide access to affordable medicines as a minimum core duty under the international right to the highest attainable standard of health (ICESCR, article 12.1; UN CESCR 2000, para. 47).

This is a conflict faced by a majority of countries globally: Approximately 84% of all WTO members (128 countries) have ratified the ICESCR (UN OHCHR 2008a) and are therefore bound by both TRIPS and ICESCR duties. The remaining WTO members who are not parties to the ICESCR will certainly be bound under health rights in article 24.1 of the Convention on the Rights of the Child, which, with 193 ratifications, holds an effective universality (albeit that its rights and duties only apply to children) (UN OHCHR 2008b). Thus, almost all WTO members must balance treaty or legislative right to health duties with TRIPS duties, and this has practical implications for how government policy-makers and legal adjudicators of either treaty interpret and implement these duties.

The problem of conflicting treaty duties is not novel to TRIPS and the right to health. Since 1945, the field of international law has exploded, with specialized regimes emerging in multiple issue areas, each with separate rules, institutions, and practices (ILC 2006, para. 8). Similarly, international treaties have proliferated: 6,000 multilateral treaties were concluded in the twentieth century, around 1,800 of which are general treaties that all states can participate in (ILC 2006,

para. 7 footnote 10). The proliferation of treaties and specialized legal regimes that appear to be siloed from each other and international law have led to growing concerns about the fragmentation and increasing incoherence of international law (ILC 2006; Koskenniemi and Leino 2002; Pauwelyn 2003; Benvenisti and Downs 2007).

As a result, the International Law Commission (ILC), a body set up by the United Nations to codify and develop international law, initiated a four-year investigation into the issue. In 2007, it released a report which emphasizes that no specialized regime, including the WTO, operates outside of international law (ILC 2006, para. 13.a). Martti Koskenniemi, the chair of the project and author of the report, argues for an application of the principle of systemic integration so as to link functional areas to a deeper normative idea in international law, so that the "common good of humankind [is] not reducible to the good of any particular institution or regime" (ILC 2006, para. 480). Koskenniemi suggests that systemic integration between functional areas of international law can be achieved in two primary ways: first, because all bodies of law must respect hierarchically superior norms in international law, and, second, because all international law is linked through treaty interpretation in the Vienna Convention on the Law of Treaties, a legal treaty which establishes the framework and interpretive methods that all international treaties are subject to. The idea of hierarchically superior norms and the interpretative approach provided in the Vienna Convention suggest both an analytical framework and strategic approach to addressing the conflict between TRIPS and the right to health. The remainder of the paper will explore both approaches as potential solutions to achieving linkages between human rights and TRIPS.

2. The Right to Health and International Law's Superior Norms

The right to health in international human rights law is the primary basis on which to seek the prioritization of health within international law more generally. Yet human rights in general are not universally viewed as prioritized norms within international law. International law only recognizes three primary sources of hierarchically superior norms: jus cogens or peremptory norms, obligations ergo omnes, and article 103 of the Charter of the United Nations. The following section explores each category and how they may be understood to relate to the right to health.

2.1 Jus Cogens and Peremptory Norms

Jus cogens are the only norms within international law recognized as superior to all others. Article 53 of the Vienna Convention on the Law of Treaties defines peremptory norms as those norms accepted and recognized by the international community of states as a whole as norms from which no derogation is permitted, and which can be modified only by a subsequent norm of general international law having the same character. Few acts are seen in this way, although international consensus has coalesced around a core list of prohibited acts, including slavery, genocide, torture, and racial discrimination (Dugard 2001, p. 40; Brownlie 1979, p. 513; ILC 2001, pp. 283–84).

From an international law perspective, these norms have two primary functions. The first, as article 53 goes on to specify, is to void any treaty that conflicts with such norms. The second function is to bind any state to compliance, irrespective of whether it agrees with the rule or not. This aspect of the norms runs contrary to the traditional view of international law as a consensual order and has generated considerable controversy (Weil 1983, pp. 430 and 441; D'Amato 1990; Caplan 2003, p. 741). The non-consensual nature of these norms is nonetheless somewhat attenuated by the Vienna Convention's definitional requirement that peremptory norms are recognized as such by the international community of states. Yet, requiring political consensus to identify peremptory norms also points to the circuitous logic of their definition: Peremptory norms are supposed to be binding irrespective of consent, albeit that their content is determined by consent (Koskenniemi 1989, p. 283). The requirement of consent also suggests more inherent limitations, since contestation, *realpolitik*, and political or cultural bias may prevent political consensus being reached about new jus cogens norms or about classifying conduct as falling under existing norms (Meron 1986, p. 4) (the lack of a concerted political response to genocide in Darfur provides a cogent contemporary example of the latter). I am not suggesting that existing peremptory norms are wrong, but, rather, that this short list cannot possibly reflect the priority of all human rights norms, nor exhaust international law's prohibitions on similarly wrongful acts. The absence of health-related prohibitions from the list is therefore not a prohibitive bar to placing health within a hierarchy of protected human rights interests within international law in the future. On the contrary, I argue that the circumscribed list of accepted peremptory norms is itself reflective of normative priorities within international

law that could reasonably be interpreted to implicitly include health-related interests.

Indeed, as the International Law Commission pointed out in its commentary to the Vienna Convention's provisions on peremptory norms, it is not the form but the particular nature of the subject matter that gives norms the character of jus cogens (*Reports of the ILC* 1966). This recognition is evident in the way that the International Court of Justice has recognized that these norms are derived from interests that are "fundamental" or relate to "elementary considerations of humanity" (*Corfu Channel* 1949, p. 22) or "intransgressible principles of international law" (*Legality of Nuclear Weapons* 1996, para. 79). The "fundamental" nature of these values is suggested in the prohibitions on slavery, genocide, and torture, acts that grossly violate human rights to dignity, life, and bodily integrity to the extent that such violations negate the essential content of all human rights.

These recognized prohibitions hold two primary implications relevant to the present inquiry. The first is that human rights to life, dignity, and bodily integrity reflect prioritized and "fundamental" interests within international law. If this is the case, aspects of the right to health would appropriately fall within the ambit of recognized prioritized human rights within international law, since the right to health is intimately connected to these (and other) rights. As the United Nations Committee on Economic, Social and Cultural Rights indicates, "[h]ealth is a fundamental human right indispensable for the exercise of other human rights" (UN CESCR 2000, para. 1).

Health and life in particular are intimately interlinked, a relationship well recognized within international legal theory and practice. For example, while the right to life in article 6.1 of the International Covenant on Civil and Political Rights is formulated as imposing a duty to prohibit arbitrary deprivations, this right has been interpreted as imposing a positive obligation on states to increase life expectancy especially in eliminating epidemics (UN Human Rights Committee 1982, para. 5). As a result the Human Rights Committee has required states to include health-related data including in relation to pregnancy and child-related deaths of women in their regular treaty reports (UN Human Rights Committee 2000, para. 10). Several regional and national judicial fora have similarly been willing to read positive health-related obligations into the right to life (*Villagran Morales* 1999, para. 144; *Paschim Banga Khet* 1996; *Glenda Lopez* 1997).

Indeed, some academic commentators and courts suggest that the right to life is itself a jus cogens (Parker and Neylon 1989, pp. 431–32; *Street Children Case* 1999, p. 139).

It seems unlikely, however, that the right to life could be designated as a jus cogens in toto. For example, while the arbitrary deprivation of human life is almost universally viewed as unlawful (Finnis 1980), such acts would likely only rise to the level of jus cogens or international crime if committed on a mass scale or with the requisite intent to constitute genocide or a crime against humanity. This implication is made explicit in the Rome Statute of the International Criminal Court, which specifies that killing will only constitute genocide if accompanied with an "intent to destroy, in whole, or in part, a national, ethnical, racial or religious group" (article 6).

This latter provision points to the second major implication of the acts prohibited under jus cogens, namely that any gross violations of these and other human rights would be similarly prohibited. International legal doctrine and practice suggest that, at a particular scale and intent, all human rights violations become gross and therefore peremptory. This idea is supported in several authoritative legal sources: For example, the American Law Institute, a respected codifier of international law, has identified the broad category of gross violations of internationally recognized human rights as a peremptory norm (American Law Institute 1987, section 702). This recognition is similarly reflected in the United Nations "1503" mechanism, which allows individual complaints about gross human rights violations to be made against any country, irrespective of any specific treaty duties that a country may hold (ECOSOC Resolution 1970, para. 1.a). These categories imply a universal prohibition against gross violations of human rights which are not dependent on treaty ratification.

What constitutes a gross violation would be factually dependent: A violation could be gross because of its scale or order of magnitude, such as for example, a legally enforced system of apartheid (Crawford 2002, pp. 245–47). However, a violation could also be gross based on "the intensity of the violation or its effects" (ibid.) even if perpetrated against only one person. For example, the prohibition on slavery would be violated by enslaving only one person.

Thus, both accepted peremptory norms and the categorical prohibition of any gross human rights violation suggests a hierarchical importance for certain human rights. This seems fairly logical within

the human rights context—since a person is far more egregiously harmed by being enslaved than by, for instance, losing their right to vote.

As a recognized human right and an interest intimately connected to human life, liberty, and worth, individual rights to health surely can be grossly violated. The alternative is to suggest that autocrats can cause the deaths of their populations with impunity by depriving them of access to medicines instead of killing them with guns. In this regard, it is notable that the Rome Statute of the International Criminal Court includes, within the definition of crimes against humanity, extermination through depriving people of access to medicines (article 7). The implication is that even without the intent to exterminate, acts which deprive populations of access to medicines with the intent to cause mass death could sit comfortably within the prohibition on gross human rights violations and the hierarchy of prioritized rights it infers.

Therefore, the fact that the right to health does not appear within accepted peremptory norms does not suggest that this category lacks relevance to health. Indeed, the absence of health-related norms could be understood instead as reflecting the skewed and culturally specific nature of those norms identified as peremptory (Simma and Alston 1988–89, p. 94; Charlesworth and Chinkin 1993, p. 68). As Simma and Alston suggest

> it must be asked whether any theory of human rights law which singles out race but not gender discrimination…and which finds no place for a right of access to primary health care, is not flawed in terms both of the theory of human rights and of United Nations doctrine. (Simma and Alston 1988–89, p. 95)

The lack of doctrinal and jurisprudential development of the right to health within domestic and international law has certainly contributed to the de facto inferior status of social rights despite international human rights law's rhetorical commitment to the indivisibility of all human rights. In recent years this neglect has considerably abated, with moves to delineate the scope and content of the right to health (UN CESCR 2000). The legal and political significance of this right has also been elevated through the creation of a permanent post for a UN Special Rapporteur on the Right to Health (UN Commission on Human Rights 2003). At the same time, there has been a considerable increase in domestic enforcement of the

right to health (Hogerzeil et al. 2006; Forman 2008), and in June 2008, a long awaited individual complaints procedure for violations of rights under the International Covenant on Economic, Social and Cultural Rights was adopted by the Human Rights Council and was opened for state signature in March 2009 (UN Human Rights Council 2008).

This growth in the legal clarity and enforceabilility of the right to health contradicts earlier contentions that social rights cannot be grossly violated to rise to the level of jus cogens (Sinclair 1984, p. 217). The legal development of the right to health (and social rights in general) may therefore contribute to a more coherent integration of these rights into other aspects of international human rights theory, including, over time, jus cogens. Moreover, while state consensus is required for the emergence of new jus cogens, popular opinion and "the dictates of public conscience" can influence and help to form state opinion (Meron 2000, p. 83). Advancing arguments that the right to health could appropriately fall within the category of a peremptory norm may therefore assist this process by coalescing popular opinion and thereby influencing state practice. Even in the absence of these kinds of developments, the jus cogens category remains relevant since it suggests that the right to life (and, ergo, related aspects of the right to health) should be viewed as prioritized interests within international law that should be favoured when balanced against certain competing legal interests.

2.2 Obligations Ergo Omnes (Duties "Owed to All")

The idea of prioritized interests is similarly reflected in the legal doctrine of obligations ergo omnes, in other words, state duties owed to the "international community as a whole." This concept was created by the International Court of Justice in the *Barcelona Traction* case, where the Court suggested that certain rights are so important that all states can be held to have a legal interest in protecting them, and that these include duties arising from the "basic rights of the human person" (*Case Concerning the Barcelona Traction Light and Power Company* 1970, paras. 33–34). The idea of obligations ergo omnes is functionally related to peremptory norms, since both concepts refer to prioritized human interests. The difference, however, is that peremptory norms prohibit the violation of these interests, whereas obligations ergo omnes speak only of duties regarding the interests themselves. Thus, rather than constituting non-derogable norms like jus cogens, obligations ergo omnes confer a general

standing on all states to make claims in the event of a violation (Byers 1997, p. 230).

While the Barcelona dictum suggests that ergo omnes duties arise from "basic" human rights, academic commentary has shifted to viewing human rights in totality as imposing ergo omnes obligations (Meron 1986, p. 13; Seiderman 2001, p. 133; Institut de Droit International Annuaire 1989, p. 338). Nonetheless, as the discussion around jus cogens suggests, the extension of universal duties of protection with regard to human rights is likely to extend not to all violations, but rather only to those considered serious or gross, or which relate to "basic," elementary, or fundamental human interests.

Henry Shue provides a workable definition of what could be considered to fall within this realm, by articulating the notion of "basic rights." He suggests that such rights would include rights to security, liberty, and subsistence—in other words, those rights we consider to constitute "everyone's minimum reasonable demands upon the rest of humanity" (Shue 1980, pp. 18–22). This idea is well reflected in human rights theory, which designates certain rights in total and aspects of others as non-derogable (Shelton 2002, p. 330). The UN Economic, Social and Cultural Rights Committee applies a similar concept to the right to health, which is viewed as having an essential minimum core that cannot be limited without very stringent justification, and which includes available, accessible, and affordable essential medicines (UN CESCR 2000, paras. 43–44). To the extent that human life *in toto* may depend on accessing medicines, the priority of this right within human rights more generally is implied.

The demarcation of prioritized rights and core elements of rights does not simply denote their *a priori* importance, but also suggests an approach to how they should be balanced with competing interests. International human rights treaties such as the International Covenant on Civil and Political Rights provide for the justifiable limitation of certain human rights to protect collective interests such as national security, public health, or the rights of others. International law has further developed the criteria for such restrictions, indicating that they must be both necessary and proportional, meaning that restrictions should be the least restrictive alternative for achieving a particular aim (UN ESC 1984).

Thus international law prioritizes certain human interests by peremptorily prohibiting their gross violation, and by placing

reasonable limits on political, economic, or social conduct that renders human rights meaningless. This idea is practically animated in the emergence of a prohibition of slavery, a legal system of property in which human life was subordinated to the commercial interests of private slave owners and the countries that supported and benefited from the slave trade. The prohibition of slavery not only suggests that human life and equal worth are prioritized values within international law, but also implies that international law limits trade and commerce to the extent that they grossly violate the worth and existence of human health and life. This basic principle is reflected within the WTO's founding document, including article XX of the General Agreement on Trade and Tariffs, which allows states to limit GATT duties to ensure open markets when doing so is necessary to protect human health and life. As I discuss below, while this clause does not necessarily provide effective protection against trade restrictions affecting human life, its inclusion within both the 1948 and current trading regime does suggest recognition that the priority of health and life places reasonable limits on trade.

2.3 Article 103: Charter Duties Regarding Global Health

The final area of international law where one could locate a prioritized value for the right to health is in the Charter of the United Nations, which established the UN and which all UN member states ratify as a binding treaty. In the Charter, states pledge to meet the UN's primary objectives of maintaining peace and security, solving international problems of an economic, social, cultural, or humanitarian nature, and promoting human rights (articles 1.1, 1.3, 55, and 56). Article 103 states that Charter duties take precedence over all other treaty duties, a primacy reiterated in decisions of the International Court of Justice (see for example, *Military and Paramilitary Activities* 1986). This article has been viewed as prioritizing human rights duties over other treaty obligations, since, under the Charter, UN member states pledge to take action to achieve universal respect for human rights. This is, however, a point that academics are divided on, since references to human rights in the Charter are sparse, and the general understanding of human rights in 1945 might not have incorporated all human rights subsequently protected. The inclusion of human rights duties within the Charter nonetheless suggests that states have undertaken some kind of duty regarding human rights, and the quibble can therefore only be about the nature of that duty rather than its existence. Thus,

all that can be said with any precision is that this uncertain duty holds a priority status over other treaty duties and would likely extend to health as an indivisible part of human rights.

The Charter is far more explicit about state duties regarding health. For example, in article 55.b, UN member states explicitly pledge to take action to achieve solutions of international health problems, and it seems clear that such duties will set aside other treaty rules to the extent that they conflict with article 103. Identifying these duties might then depend on what is defined as an international health problem. These duties are almost certain to cover disease threats with international dimensions (such as pandemic disease threats or the spread of resistant strains of extensively drug resistant tuberculosis). Access to medicines falls within the ambit of these Charter duties to the extent that resolving these threats may require the provision of medical treatments.

The other point to emphasize is that the explicit purpose of the UN Charter is to maintain peace and security, suggesting that the reason that Charter duties are prioritized over others is because of their importance to collective wellbeing (ILC 2006, para. 36). The implications of Charter duties for the present inquiry are therefore twofold: first, they indicate that health is prioritized within international law because of the global harm of not doing so; second, they explicitly indicate that states hold international responsibilities to protect against global health problems. This latter idea is reflected in the UN Security Council's resolution in 2000 recognizing AIDS as a threat to global security (Hindmarch 2010). The recognition of HIV/AIDS as a threat to security gives currency to the idea that health (including access to medicines for global health threats) is a prioritized collective interest that implicates powerful Charter duties capable of subordinating competing obligations under other treaties.

It remains in question whether the recognition of health as a prioritized norm within international law could be achieved within the WTO, where health impacts are viewed, if at all, as externalities to trade policies, rather than as their primary objectives.

3. TRIPS and Rights: Achieving Linkage through an Interpretative Approach

The Vienna Convention on the Law of Treaties establishes the framework and interpretive methods that all international treaties are

subject to. I turn now to explore the potential for this interpretive approach to achieve linkages with the right to health by, first, overviewing the Vienna Convention's provisions on treaty interpretation; second, exploring the implications of these provisions for interpreting trade-related intellectual property rights at the WTO; third, overviewing WTO jurisprudence on TRIPS in light of this provision; and, fourth, identifying subsequent state agreements and practice related to TRIPS that could affect its interpretation.

3.1 The Vienna Convention on the Law of Treaties

The Vienna Convention's primary provision on treaty interpretation is contained in article 31, which specifies that treaties must be interpreted in good faith according to the ordinary meaning of treaty terms in their context and in the light of a treaty's object and purpose (article 31.1). Adjudicators can ascertain this context from the text of a treaty (including its preamble and annexes) as well as any agreements or instruments relating to the treaty made between parties (article 31.2). They can also take into account subsequent state agreement and practice between parties regarding treaty interpretation (article 31.3) as well as "any relevant rules of international law applicable between the parties" (article 31.3.c).

The broader intent of these rules is to give meaning to treaties that accords with that agreed to during the drafting of a treaty and subsequent practice. The specific relevance of article 31 to WTO adjudication is well established in law and practice: the WTO Understanding on Rules and Procedures Governing the Settlement of Disputes mandates that WTO dispute settlement panels interpret trade agreements according to customary rules of interpretation of public international law (article 3.2). WTO dispute settlement panels have interpreted this WTO provision to require reference to article 31 of the Vienna Convention (*U.S.—Gasoline* 1996, pp. 15–16; *U.S.—Carbon Steel* 2002, paras 61–62; *India—Patents* 1998, para. 45). Article 31 of the Vienna Convention therefore requires that TRIPS be interpreted in light of its provisions in their entirety, as well as subsequent agreement and practice and other relevant rules of international law. As the following sections illustrate, subsequent agreement and practice around TRIPS may hold important interpretive value in relation to health concerns under the agreement.

However, at a legal and normative level, the aspect of article 31 most likely to enable incorporation of the right to health within TRIPS is

article 31.3.c, which authorizes reference to "any relevant rules of international law applicable between the parties." This article most obviously links disparate treaties, and is viewed as animating the principle of systemic integration whereby contextual interpretation of treaties takes account of a broader normative environment within international law (ILC 2006, pp. 420–23; McLachlan 2005).

In practice, international tribunals are using article 31.3.c as a linkage device between disparate bodies of international law, including the International Court of Justice (*Case Concerning Oil Platforms* 2003, para. 41), the European Court of Human Rights (*Bosphorus* 2005; *Al-Adsani* 2001; *Fogarty* 2001; *McElhinney* 2001), and the Iran-U.S. Claims Tribunal (*Esphahanian* 1983). Moreover, a WTO tribunal has referred to this section, albeit only in a footnote, as the basis for seeking additional interpretive guidance of article XX of the General Agreement on Trade and Tariffs according to general principles of international law (*U.S.-Shrimp* 1998, para. 181).

Whether this section could be used to take account of the right to health would, however, depend on whether an adjudicative body considered this right as "a relevant rule applicable between the parties." Nonetheless, it is doubtful whether this article alone provides a solution to systemic fragmentation (Higgins 2006, p. 804). A recent international case illustrates why the problem of linkage is unlikely to be resolved simply by formal recognition of other rules of international law. In the 2005 *MOX Plant* case heard by the International Tribunal for the Law of the Sea, the United Kingdom invoked three separate treaties and institutional procedures: the United Nations Convention on the Law of the Sea (UNCLOS), the Convention on the Protection of the Marine Environment of the North-East Atlantic (OSPAR Convention), and the European Community and Euratom Treaties within the European Court of Justice (EC/EURATOM Conventions). While the case clearly raised questions of which issue area should be determinative in the case, the UNCLOS tribunal addressing this issue also noted that

> even if the OSPAR Convention, the EC Treaty and the Euratom treaty contain rights or obligations similar to or identical with the rights set out in [the UNCLOS], the rights and obligations under these agreements have a separate existence from those under [the UNCLOS]. (*MOX Plant* 2005, para. 50)

This meant that different institutions applying the same rules could come up with quite different results, given "differences in the

respective context, object and purpose, subsequent practice of parties and *traveux preparatoires*" (ILC 2006, para. 12). The implication therefore is that "the meaning of legal rules and principles is dependent on the context in which they are applied. If the context, including the normative environment, is different, than even identical provisions may appear differently" (ibid.). This problem is exemplified in international law, where specialized regimes respond to different functional and technical requirements, providing entirely different normative environments for the interpretation of even ostensibly similar rules (ILC 2006, para. 15). The *telos* (or ultimate end) of each regime will therefore fundamentally influence how rules are interpreted. Assuring that TRIPS will be interpreted in accordance with the right to health may therefore require more than simply referring to other formal international legal rules, and may instead require a considerable shift in the normative priorities underlying the trade regime. Nonetheless, the approach to interpretation laid out in the Vienna Convention provides an interpretive approach which may enable the prioritization of TRIPS even on its own terms.

3.2 Interpreting TRIPS in Light of the Vienna Convention

As its title suggests, a primary purpose of an agreement on trade-related intellectual property rights is to protect intellectual property rights within all WTO member states. The agreement achieves this goal succinctly, requiring members to provide patents which confer exclusive rights for twenty-year periods. The context of the agreement also is reflective of its purpose: TRIPS is an annex to the Marrakesh Agreement, which established the WTO, and this context locates TRIPS firmly within a legal system driven by the *telos* of free trade. This ambition is explicit in the very first lines of the preamble of TRIPS, which indicate that its primary object is to protect intellectual property rights in the context of free trade while ensuring that intellectual property rights enforcement does not itself become a barrier to trade. This recognition reflects the somewhat contradictory inclusion of TRIPS within the WTO, since the WTO aims to reduce market barriers and remove protectionist measures, while TRIPS (and intellectual property) is protectionist and can create barriers to trade by limiting imports and exports (Frankel 2006, pp. 373–74). This point is somewhat moot since the existence of TRIPS is itself illustrative of an acceptance of intellectual property rights as a

reasonable limit on trade. However, this argument does suggest that excessive intellectual property rights protection would fall foul of the agreement to the extent that it interfered with free trade (ibid., p. 374).

A third objective is reflected in the preamble, namely the need to balance intellectual property rights protection as private rights with public policy objectives, and the recognition that least developing countries (LDC) have special needs for maximum flexibility. These provisions were inserted in the agreement at its drafting at the urging of developing countries, which feared the negative impacts of TRIPS on public welfare and wished to highlight the importance of the public policy imperatives underlying national intellectual property rights systems (Adede 2003, p. 28; Barbosa et al. 2007, pp. 93–94).

The need to balance intellectual property rights protection with public policy is iterated most explicitly in articles 7 and 8. Article 7 indicates that the objective of TRIPS is that intellectual property rights promote innovation to the mutual advantage of end users and producers of intellectual property, in a manner conducive to social and economic welfare and a balance of rights and obligations. Article 8 is explicit that, as governments implement TRIPS, they "may adopt measures necessary to protect public health and nutrition." The language of mutual advantage, public health, and social welfare provide explicit links for right-consistent interpretations cognizant of the public welfare implications of intellectual property restrictions on drug access. However, the provisions are silent on what this entitles governments to do, and only indicate the proviso that these measures should be consistent with TRIPS. Yet the implication is that these measures should include the provisions TRIPS itself famously provides in exclusions, exceptions, and limitations which implicitly recognize the need to balance intellectual property rights with social welfare interests. For example, members can exclude certain inventions from patentability to protect life or health, make limited exceptions to exclusive rights, authorize use without consent (through compulsory licensing), limit rights to prevent anti-competitive measures, and parallel import cheaper patented medicines (TRIPS articles 27.1, 30, 31, 40, and 6). These provisions therefore seem to reflect the balance sought within the agreements: namely, to protect intellectual property rights without unduly restricting either free trade or public welfare. From the perspective of the Vienna Convention on the Law of Treaties, these are the objects and purpose which must be balanced by treaty interpreters.

3.3 TRIPS and the WTO Dispute Settlement Procedure

Since 1995, 25 TRIPS-related complaints have been lodged at the WTO dispute settlement units. Thirteen of these cases (over 50%) were settled, while eight went to panel reports and three to the appellate body. These eight cases comprise the entire body of the WTO dispute settlement body's jurisprudence on TRIPS, and only four addressed pharmaceutical patents (see Table 1). These TRIPS cases have been criticized as largely interpreting the object, purpose, and context of TRIPS in favour of protecting intellectual property rights, and giving little weight to arguments about public welfare (Barbosa et al. 2007, p. 99; Howse 2000; Frankel 2006, pp. 373–74; Shankar 2002). This approach is evident in the first two cases decided on TRIPS, where WTO panels upheld challenges by the U.S. and E.U. against what they alleged was India's insufficient protection for patents during its transition period to full implementation of TRIPS. In the first *India—Patents Case*, which dealt with the U.S. complaint, the Appellate Body upheld the panel's finding that India's failure to provide mailbox protection and exclusive marketing rights during its transition period violated articles 70.8 and 9 of TRIPS. Despite explicit reference to article 31 of the Vienna Convention, the Appellate Body found that TRIPS' primary object and purpose as deduced from the preamble were "to take into account, *inter alia*, 'the need to promote effective and adequate protection of intellectual property rights'" (*India—Patent Protection* 1997, para. 57).

Table 1: TRIPS Jurisprudence

India—Patent Protection, 1997
India—Patent Protection, 1998
Canada—Patent Protection of Pharmaceutical Products Case, 2000
U.S.—Section 110(5) of the U.S. Copyright Act, 2000
Canada—Term of Patent Protection, 2000
United States—Section 211 Omnibus Appropriations Act of 1998, 2002
European Communities—Protection of Trademarks and Geographical Indications for Agricultural Products and Foodstuffs, 2005

This emphasis on intellectual property rights to the exclusion of other interests is reiterated in the 2000 *Canada-Pharmaceuticals* case, where a WTO panel interpreted the limited exception provision in such a

narrow way that intellectual property rights were protected as a primary objective, and public interest arguments made under articles 30, 7, and 8 were effectively ignored. This decision has been criticized as reflecting a one-sided and narrow interpretation of the object and purpose of TRIPS, and interpreting TRIPS "largely from the perspective of intellectual property rights holders, abstracting from competing social interests, and reducing considerably the range of regulatory diversity permitted under TRIPS" (Howse 2000, p. 494). As other commentators have suggested, this approach is in conflict with international law's approach to treaty interpretation since it focuses on a single provision rather than reading the whole treaty together (Barbosa et al. 2007, p. 102).

3.4 The Doha Declaration, TRIPS Amendment, and State Practice on AIDS

It is doubtful in light of subsequent legal and political events whether WTO panels today would adopt a similarly restrictive approach to interpreting TRIPS. A primary intervening variable is the Doha Declaration on TRIPS and Public Health, passed at the urging of developing countries at the Doha round of WTO trade negotiations in November 2001 (WTO 2001). These countries sought legal clarification of their right to use TRIPS flexibilities given political limitations of previous efforts to do so (WHO 2002). The Doha Declaration recognizes the agreement of member states that TRIPS does not and should not prevent members from taking measures to protect public health, and that TRIPS should be interpreted and implemented in a manner supportive of WHO members' right to protect public health and, in particular, to promote access to medicines for all, including through using TRIPS flexibilities to the full (WTO 2001, para. 4). The Doha Declaration explicitly seeks to raise the priority of articles 7 and 8, indicating that customary rules of interpretation require reading TRIPS provisions in light of its object and purpose, particularly its objectives and principles (the official titles of articles 7 and 8 within TRIPS) (WTO 2001, para. 5.a). The Declaration also notes the need to revise the compulsory licensing provision to allow production for export (WTO 2001, para. 6). A decision to do so was made by the WTO in 2003 and will likely become a formal amendment to TRIPS itself (WTO General Council 2003).

The Doha Declaration is not a formal amendment of TRIPS and has no specific legal status within WTO law (WTO 2001, p. 43). However, it

could be construed as a subsequent agreement that should guide the interpretation of a treaty as envisaged in article 31.3.a of the Vienna Convention (Frankel 2006, p. 400). This would certainly found arguments that state intentions require panels to recognize health as an interest within TRIPS at least of equal weight to intellectual property rights. The Doha Declaration's use of the terms "right to protect health and promote access to medicines" also provides a framework to support references to similar rights in international law. This suggestion has precedent in WTO law: in the *U.S.-Shrimp* case, a WTO panel used several international environmental treaties and non-binding instruments to interpret a key term in article XX (*U.S.-Shrimp* 1998). Yet, panels interpreting TRIPS have not been particularly welcoming of social welfare concerns, and it is unclear from the jurisprudence what approach they would take to recognizing the right to health. What is clear is that a panel's approach today would likely be quite different given political and legal developments over the last few years with regards to TRIPS.

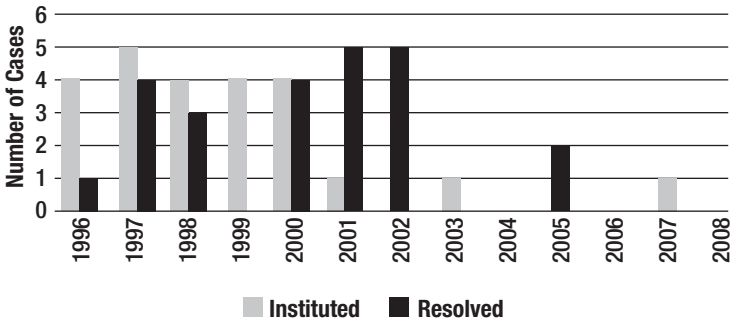
In addition to the Doha Declaration, there have been very important changes in state practice around TRIPS and AIDS treatment, which has become accepted as a legitimate reason for limiting intellectual property rights. This acceptance is reflected in growing issuances of compulsory licenses for AIDS medicines, including by Malaysia, Indonesia, Zambia, Zimbabwe, Mozambique, Brazil, and, most recently, Thailand (Love 2007). However, it remains unclear whether these gains can be expanded out of the HIV/AIDS or pandemic context to apply to public health needs more generally, as recent disputes over Thailand's compulsory licensing of cancer drugs suggest. These changes in state practice around the use of TRIPS flexibilities are legally significant since the Vienna Convention indicates that treaties can be interpreted by reference to subsequent state practice to assess agreements regarding interpretation. This practice will be relevant to adjudication should another case on TRIPS reach the WTO's dispute settlement panels.

However, this outcome appears increasingly unlikely. In the first place, the dispute settlement process (including in relation to TRIPS) has been used primarily by developed countries (Leitner and Lester 2007, p. 165). In ten years of dispute resolution, less developing countries have brought only one case and have never been the respondents in WTO disputes (ibid., pp. 167–68). While these trends are not determinative of future use of the mechanism by developing countries,

they nonetheless likely reflect structural barriers to using the mechanism as a strategic venue for developing countries, suggesting the limitations of this venue for achieving health aims.

At the same time, there has been a significant drop in the use of WTO dispute settlement for TRIPS cases. The majority of TRIPS cases were instituted between 1996 and 2000, with only three cases instituted after 2000, none of which related to pharmaceutical patents (see Figure 1). Instead of relying on dispute settlement at the WTO to protect against lenient application of intellectual property rights, the U.S. uses its own unilateral sanctions mechanism to enforce these rights. Moreover, developed countries are directly negotiating bilateral and regional free trade agreements with developing countries, which impose even stricter intellectual property rights than contained in TRIPS. For example, around 60 countries are currently bound by bilateral or regional free trade agreements negotiated by the U.S. (Forman 2006). These stricter intellectual property rights have the strategic effect of ratcheting these standards globally, since the non-discrimination requirement of WTO law requires that standards given to one country be given to all (Drahos 2005, p. 7). None of these factors present a bar to a developing country instituting a complaint at the WTO. However, the implication is that, for the moment, strategic approaches aimed at integrating rights to health standards into TRIPS would be better focused elsewhere than the WTO dispute settlement panels.

Figure 1: TRIPS Cases 1996–2008



4. Implications for Reform

What then are the implications of the legal analysis on hierarchy and interpretation for raising the primacy of health in the interpretation of TRIPS provisions and assuring reference to rights standards at the

WTO? The Doha Declaration and the effective carve out of AIDS medicines from the application of TRIPS provisions seem to animate the de facto priority of life and health interests within international politics and law. The promotion of these interests by social actors and governments ensured that TRIPS provisions were curtailed to the extent that they conflicted with these interests. These gains were not achieved through adjudication but through social action, and this suggests that legal strategies alone are not necessarily optimal approaches for achieving linkages between rights and trade.

The potential limitation of legal strategies is reinforced by the declining use of the WTO dispute settlement panels, suggesting that efforts to achieve linkage may be most effectively aimed at domestic policy-making. Achieving a pragmatic linkage between health and pharmaceutical patents in the minds of policy-makers may acculturate a policy response whereby intellectual property rights are not implemented or negotiated without reference to their health impacts or the consequent requirement of using the legal flexibilities in TRIPS to mitigate these impacts. This outcome is surely a lesson that can be drawn from the very title of TRIPS itself, which is an agreement to relate trade and intellectual property rights, essentially forcing a marriage between the two despite the fact that they have no natural association. Certainly, a primary implication of the legal analysis is that it should directly influence how TRIPS is applied, and support the most expansive use of TRIPS flexibilities possible.

However, this approach has important limitations beyond the clear difficulties that still persist in using flexibilities. The first is that using the legal flexibilities will never resolve the systemic impact of TRIPS on global production and export of generic medicines, which have become strictly confined to licensing and export authorized under TRIPS. Whether this systemic impact can be mitigated depends on the extent to which major generic producers like India can maintain the policy space they have carved out to ensure generic production. Nonetheless, expanded use of legal flexibilities will never fully contain the global impact of this agreement on access to generic medicines. A more insidious danger is that relying on exceptions effectively proves the rule and reinforces the legitimacy of TRIPS itself. In other words, arguing that exceptions should be made for AIDS, diabetes, or other narrowly defined exceptions simply reinforces the general rule that drug patents in poor countries should be respected and extended for all other cases. This is not ground that should be ceded, and, indeed,

arguments that TRIPS should be excised from WTO are forthcoming not simply from treatment advocates, but from prominent economists like Joseph Stiglitz and Jagdish Bhagwati (Stiglitz and Charlton 2005, pp. 141–46; Bhagwati 2004, p. 182). Bhagwati, an ardent supporter of free markets, describes the inclusion of TRIPS at the WTO as a result of "pharmaceutical and software companies muscling their way into the WTO to turn it into a royalty-collection agency simply because the WTO can apply trade sanctions" (Bhagwati 2004, p. 182). This critique suggests that this is an important arena in which human rights and economic arguments can cohere and suggests more broadly that efforts to achieve reform should not simply aim for the lenient application of TRIPS, but should rather argue that for some and perhaps all countries, it should not apply at all.

Another danger of the interpretive approach is that little is achieved by inserting formal rules into a system if those rules continue to be seen as foreign to the system. Legal systems cohere to their internal logics and priorities, and trend towards compliance with those logics and priorities. Achieving counter-systemic objectives is not impossible, but will be exceptional. This certainly seems to be the case in the WTO cases which trend, as they should, towards achieving the WTO's systemic objectives of free trade. The implication is that formal linkages to rights through interpretation are unlikely to achieve public health objectives, and that inserting rights into TRIPS will be ineffective without a normative recognition of health as a comparable or superior interest. Formal rules will always be deprioritized in service of deeper normative priorities unless they can be seen to be relevant to them, and the confluence of rights and trade on avoiding excess intellectual property rights seems to provide important framing for this outcome. This requires advancing towards a new *telos* for the global trading system in which trade and social values can be upheld as synergistic rather than competing aims (Cho 2005, p. 674).

Assuring a more appropriate balancing of health and trade interests is not dependent, however, on establishing the legal priority of health over trade within international law's hierarchically superior norms. As Koskenniemi suggests in the Fragmentation Report, in relation to article 103 duties, "the importance of the notion—like the importance of *ergo omnes* obligations—may lie less in the way the concepts are actually 'applied' than as signals of argumentative possibilities and boundaries for institutional decision-making" (ILC 2006, p. 409). The drawback to arguing that health is a prioritized value based on its

putative hierarchical status within international law is, of course, that if these arguments fail, then so does the argument for value. Engaging in analytic distinction regarding the hierarchical status of health as a human right within international law may therefore unnecessarily raise the bar for protecting the right to health (and ergo all human rights), by pinning its priority to the putative legality of the right to health within jus cogens, obligations ergo omnes, or article 103 of the UN Charter. The inquiry into hierarchy cannot therefore be the sole strategy for raising the priority of health within trade. Instead, these rules suggest the importance of shifting adjudicative interpretation of trade rules towards recognition of the broader international law framework. As Howse and Teitel suggest, the normative environment of international law "presupposes that there is a minimum substantive normativity inherent in the international legal order, a kind of foundation or floor, grounding the aspirations and efforts of the international legal system," and that the preservation of human life and health can be understand to comprise that floor (Howse and Teitel 2007, p. 10).

Ultimately these arguments may most effectively be used to promote a political agenda regarding health as a value that seeks to influence the state interests, preferences, and practices that lead to the creation of new international legal rules. The argument regarding superiority is therefore a normative and pragmatic exercise in shifting lex lata (the law as it should be) to lex ferenda (the law as it is).

Achieving this goal requires reconceiving human rights as providing a normative framework for resolving trade-offs between trade and health interests in light of the normative importance of human rights values, and pre-existing approaches to balancing rights within human rights law and theory according to principles of necessity and proportionality (Lang 2007, p. 379).

Conclusion

This article overviews strategic methods of achieving linkage between the right to health and trade rules within international law. However, these approaches to linkage do not propose the integration of human rights into trade rules. This implies the one-way insertion of rights into TRIPS or the WTO, a strategic approach which is not simply overly narrow but potentially counterproductive. A formal recognition of rights at the WTO is unlikely to raise their priority unless they are linked to the deeper normative priorities of international law, of which

the right to health forms an integral part. The integration therefore should be not of rights into TRIPS but of TRIPS into the broader normative system of international law. This argument holds important implications for the formulation and interpretation of the substantive provisions of TRIPS and for its enforcement.

This argument finds important support in international law, as this paper seeks to illustrate. Yet, the legal exegesis necessary to support this claim is itself illustrative of the moral imbalance within international politics: There is something absurd in having to go to such great lengths to establish that human life should be worth more than property. Having to do so defies common sense. Yet this is not the common sense encoded into TRIPS, which relegates health protection to a non-essential exception to a property right. Historical shifts around slavery, women's suffrage, and colonialism suggest that systemic changes are possible, and that we should not restrict ourselves to pragmatism at the expense of transformation. The aim of reform should therefore be to achieve the integration of the priority of health into the very fabric of economic and political life, not simply because it serves important instrumental ends, which it does, but because we do not choose to live in a world where elementary considerations of humanity can be so easily and pragmatically ignored.

Note

1. I am grateful for helpful comments on earlier drafts from Antoinette Handley and Asher Alkoby. Please address correspondence to the author at lisa.forman@utoronto.ca.

References

Abbott, Frederick M. 2002. Study Paper for the British Commission on Intellectual Property Rights on the WTO TRIPS Agreement and Its Implications for Access to Medicines in Developing Countries. 14 February 2002.

Adede, Adronico O. 2003. Origins and History of the TRIPS Negotiations. In *Trading in Knowledge: Development Perspectives on TRIPS, Trade and Sustainability* edited by Christophe Bellmann, Graham Dutfield, and Ricardo Melendez-Ortiz. London: Earthscan Publications.

Agreement on Trade-Related Aspects of Intellectual Property Rights, Annexure 1C to the Marrakesh Agreement Establishing the World Trade Organization, signed in Marrakesh, Morocco on 15 April 1994.

Al-Adsani v. United Kingdom. 2001. 123 ILR 24.

American Law Institute. 1987. *Restatement (Third). The Foreign Relations Law of the United States*. St. Paul, Minn.: American Law Institute Publications.

Barbosa, Denis Borges, Margaret Chon, and Andres Moncayo von Hase. 2007. Slouching towards Development in International Intellectual Property. *Michigan State Law Review* 71–141.

Benvenisti, Eyal, and George W. Downs. 2007. The Empire's New Clothes: Political Economy and the Fragmentation of International Law. *Stanford Law Review* 60 (2): 595–632.

Bhagwati, Jagdish. 2004. *In Defense of Globalization*. New York: Oxford University Press.

Bosphorus Haya Yollart ve Ticaret Antonim Sirkett (Bosphorus Airways) v. Ireland [GC]. 2005. No 45036 98, ECHR.

Brownlie, Ian. 1979. *Principles of Public International Law*. Oxford: Clarendon Press.

Byers, Michael. 1997. Conceptualising the Relationship between *Jus Cogens* and *Ergo Omnes* Rules. Nordic Journal of International Law 66 (2–3): 211–39.

Caplan, Lee M. 2003. State Immunity, Human Rights and *Jus Cogens*: A Critique of the Normative Hierarchy Theory. *American Journal of International Law* 97 (3): 741–81.

Case Concerning the Barcelona Traction, Light and Power Company, Limited (Belgium v. Spain) (Second Phase). 1970. I.C.J. Reports.

Case Concerning Oil Platforms (Iran v. United States of America). 2003. 42 ILM 1334.

Case Concerning Questions of Interpretation and Application of the 1971 Montreal Convention Arising from the Aerial Incident at Lockerbie (Libyan Arab Jamahiriya v. United States of America, Judgment of 27 February 1998.

Caves, R.E, M.D. Whinston, and M.A. Hurwitz. 1991. Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry. *Brookings Papers on Economic Activity*. Special Issue. 1–62.

Charlesworth, Hilary, and Christine Chinkin. 1993. The Gender of Jus Cogens. *Human Rights Quarterly* 15 (1): 63–76.

The Charter of the United Nations, 26 June 1945, Can. T.S. 1945 No. 7 (entered into force on 24 October 1945).

Cho, Sungjoon. 2005. Linkage of Free Trade and Social Regulation: Moving Beyond the Entropic Dilemma. *Chicago Journal of International Law* 5 (2): 625–74.

Corfu Channel Case (United Kingdom v. Albania). 1949. I.C.J. Reports.

Correa, Carlos. 2002. Public Health and Intellectual Property Rights. *Global Public Policy* 2 (3): 261–78.

Crawford, James. 2002. *The International Law Commission's Articles on State Responsibility. Introduction, Text and Commentaries*. Cambridge: Cambridge University Press.

D'Amato, Anthony. 1990. It's a Bird, It's a Plane, It's Jus Cogens! *Connecticut Journal of International Law* 6 (1): 1–6.

Drahos, Peter. 2005. An Alternative Framework for the Global Regulation of Intellectual Property Rights. *Austrian Journal of Development Studies* 21: 44–68.

Dugard, John. 2001. *International Law—A South African Perspective*, 2nd ed. Lansdowne: Juta.

ECOSOC Resolution 1503 (XLVIII), 1970, and Sub-Commission resolution 1 (XXIV), 1971.

Esphahanian v. Bank Tejarat. 1983. 2 Iran-U.S. Claims Tribunal 157.

Finnis, John. 1980. *Natural Law and Natural Rights*. New York: Oxford University Press.

Fogarty v. United Kingdom. 2001. 123 ILR 54.

Forman, Lisa. 2006. Trading Health for Profit: Bilateral and Regional Free Trade Agreements Affecting Domestic Property Rules on Intellectual Property Rules on Pharmaceuticals. In *The Power of Pills: Social, Ethical, and Legal Issues in Drug Development, Marketing, and Pricing* edited by Jillian C. Cohen, Udo Schuklenk, and Patricia Illingsworth. London: Pluto Press.

———. 2008. Justice and Justiciability: Evaluating Right to Health Jurisprudence in South Africa. *Journal of Medicine and Law* 27: 661–83.

Frankel, Susy. 2006. WTO Application of "the Customary Rules of Interpretation of Public International Law" to Intellectual Property. *Virginia Journal of International Law* 46 (2): 365–431.

Glenda Lopez v. Instituto Venezolano de Seguras Sociales. 1997. 487-060401. Supreme Court of Venezuela, Constitutional Chamber.

Hestermeyer, Holger. 2007. *Human Rights and the WTO: The Case of Patents and Access to Medicines*. Oxford: Oxford University Press.

Higgins, Rosalyn. 2006. A Babel of Judicial Voices? Ruminations from the Bench *International and Comparative Law Quarterly* 55 (4): 791–804.

Hindmarch, Suzanne. 2010. The Implications of Treating HIV as a Security Threat in Africa: An Analysis of the United Nations Security Council Debate on HIV. In *Comparative Program on Health and Society Lupina Foundation Working Papers Series 2007–2009*, edited by M. Bianca Seaton and Sara Allin. Toronto: Munk Centre for International Studies.

Hogerzeil, Hans V., Melanie Samson, Jaume Vidal Casanovas, and Ladan Rahmani-Ocora. 2006. Is Access to Essential Medicines as Part of the Fulfilment of the Right to Health Enforceable through the Courts? *The Lancet* 368 (9532): 305–11.

Howse, Robert. 2000. The Canadian Generic Medicines Panel: A Dangerous Precedent in Dangerous Times. *Journal of World Intellectual Property* 3 (4): 493–507.

Howse, Robert, and Ruti G. Teitel. 2007. Beyond the Divide: The Covenant on Economic, Social and Cultural Rights and the World Trade Organization. Occasional Paper No. 30, Friedrich Ebert Stiftung.

Institut de Droit International Annuaire. 1989. International Affairs of State, 63.

International Covenant on Civil and Political Rights G.A. res. 2200A (XXI), 21 U.N. GAOR Supp. (No. 16) at 52, U.N. Doc. A/6316 (1966), 999 U.N.T.S. 171, entered into force 23 March 1976.

International Law Commission. 1996. Conclusions of the Work of the Study Group on the Fragmentation of International Law: Difficulties Arising from the Diversification and Expansion of International Law. In *Yearbook of the International Law Commission*, 2006, vol. 2, part 2.

———. 2001. Draft Articles on State Responsibility, Commentary on Article 40, paras. 4–6 in Official Records of the General Assembly, Fifty-sixth Session (A/56/10).

———. 2006. Fragmentation of International Law: Difficulties Arising from the Diversification and Expansion of International Law: Report of the Study Group of the International Law Commission - Finalized By Martti Koskenniemi. U.N. Doc. A/CN.4/L.682, 13 April 2006.

Jackson, John H. 2006. Reflections on the Possible Research Agenda for Exploring the Relationship between Human Rights Norms and International Trade Rules. In *International Trade and Human Rights: Foundations and Conceptual Issues* edited by Frederick M. Abbott, Christine Breining-Kaufmann, and Thomas Cottier. Ann Arbor: University of Michigan Press.

Keck, Margaret, and Kathryn Sikkink. 2000. *Activists beyond Borders: Advocacy Networks in International Politics*. Ithaca, New York: Cornell University Press.

Kinney, E.D., and B.A. Clark. 2004. Provisions for Health and Healthcare in the Constitutions of the Countries of the World. *Cornell International Law Journal* 37: 287.

Koskeniemmi, Martti. 1989. *From Apology to Utopia: The Structure of International Legal Argument*. Cambridge: Cambridge University Press.

———. 1997. Hierarchy in International Law: A Sketch. *European Journal of International Law* 8 (4): 566–82.

Koskenniemi, Martti, and Paivi Leino. 2002. Fragmentation of International Law? Postmodern Anxieties. *Leiden Journal of International Law* 15 (3): 553–80.

Lang Andrew T.F. 2007. Re-thinking Trade and Human Rights. *Tulane Journal of International and Comparative Law* 15 (2): 335–413.

Legality of the Threat or Use of Nuclear Weapons Case, Advisory Opinion. 1996. I.C.J. Reports.

Leitner, Kara, and Simon Lester. 2007. WTO Dispute Settlement 1995–2006—A Statistical Analysis. *Journal of International Economic Law* 10 (1): 165–79.

Love, James. 2007. Recent examples of the movements on the use of compulsory licenses around the world, quoted in the Ministry of Public Health and the National Health Security Office Thailand, Facts and Evidences on the 10 Burning Issues Related to the Government Use of Patents on Three Patented Essential Drugs in Thailand. http://www.moph.go.th/hot/White%20Paper%20CL-EN.pdf.

Marrakesh Agreement Establishing the World Trade Organization, 15 April 1994, 1867 U.N.T.S. 154; 33 I.L.M. 1144 (1994).

McElhinney v. Ireland. 2001. 123 ILR 73.

McLachlan, Campbell. 2005. The Principle of Systemic Integration and Article 31(3)(c) of the Vienna Convention. *International and Comparative Law Quarterly* 54 (2): 279–320.

Meron, Theodor. 1986. On a Hierarchy of International Human Rights. *American Journal of International Law* 80 (1): 1–23.

———. 2000. The Martens Clause, Principles of Humanity, and Dictates of Public Conscience. American Journal of International Law 94 (1): 78–89.

Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. U.S.), Jurisdiction and Admissibility, 1986. I.C.J. REP. 14 (Judgment on merits 27 June).

MOX Plant Case, Request for Provisional Measures Order (Ireland v. the United Kingdom). 2005. International Tribunal for the Law of the Sea, ILR vol. 126.

Parker, Karen, and Lyn Beth Neylon. 1989. Jus Cogens: Compelling the Law of Human Rights. *Hastings International and Comparative Law Review* 12 (2): 411–63.

Paschim Banga Khet Mazdoor Samity v. State of West Bengal. 1996. 4 SCC 37.

Pauwelyn, Joost. 2003. *Conflict of Norms in Public International Law: How WTO Law Relates to Other Rules of International Law*. Cambridge: Cambridge University Press.

Prosecutor v. Anto Furundžija, Judgment of 10 December 1998, Case No. IT-95-17/1, Trial Chamber II, 121 ILR (2002) pp. 260–62.

Sanchez, Michelle Ratton, Barbara Rosenberg, and Gregory C. Shaffer. 2008. Winning at the WTO: The Development of a Trade Policy Community within Brazil. Paper presented at the inaugural conference of the Society of International Economic Law (SIEL).

Scherer Frederick M. 2000. The Pharmaceutical Industry. In *Handbook of Health Economics*, vol. 1, edited by A.L. Culyer and J.P. Newhouse, 1322–24. New York: Elsevier Science.

Seiderman, Ian D. 2001. *Hierarchy in International Law: The Human Rights Dimension*. Antwerp: Intersentia.

Sen, Amartya. 1999. *Development as Freedom*. New York: Anchor Books.

Shanker, Daya. 2002. The Vienna Convention on the Law of Treaties, Dispute Settlement System of the WTO and the Doha Declaration on the TRIPS Agreement. *Journal of World Trade* 36 (4): 721–72.

Shelton, Dinah. 2002. Hierarchy of Norms and Human Rights: Of Trumps and Winners. *Saskatchewan Law Review* 65 (2): 301–31.

Shue, Henry. 1980. *Basic Rights, Subsistence, Affluence and U.S. Foreign Policy*. Princeton: Princeton University Press.

Simma, Bruno, and Philip Alston. 1988–89. The Sources of Human Rights Law: Custom, Jus Cogens and General Principles. *Australian Year Book of International Law* 12: 82–108.

Sinclair, Ian. 1984. The *Vienna Convention on the Law of Treaties*, 2nd ed. Manchester: Manchester University Press.

Stiglitz, Joseph E., and Andrew Charlton. 2005. *Fair Trade for All: How Trade Can Promote Development*. Oxford: Oxford University Press.

Street Children Case (Morales v. Guatemala), Judgment of 19 November 1999, Inter-American Court of Human Rights (Ser. C) No 63.

The Rome Statute of the International Criminal Court, UN Doc. A/CONF.183/9 of 17 July 1998, corrected by procès-verbaux of 10 November 1998, 12 July 1999, 30 November 1999, 8 May 2000, 17 January 2001 and 16 January 2002 (entry into force 1 July 2002).

United Nations Commission on Human Rights. 2003. The right of everyone to the enjoyment of the highest attainable standard of physical and mental health: Report of the Special Rapporteur, Paul Hunt, submitted in accordance with Commission resolution 2002/31. UN Doc. No. E/CN.4/2003.

United Nations Committee on Economic, Social and Cultural Rights. 2000. General Comment No. 14 (2000): The Right to the Highest Attainable Standard of Health (article 12 of the International Covenant on Economic, Social and Cultural Rights), U.N. Doc. E/C.12/2000/4, 11 August 2000.

United Nations Convention on the Elimination of All Forms of Discrimination Against Women, 18 December 1979, U.K.T.S. 1989 No. 2, 19 I.L.M. 33 1980 (entered into force 3 September 1981).

United Nations Convention on the Rights of the Child, 20 November 1989, U.K.T.S. 1992 No. 44, 28 I.L.M. 1448 1989 (entered into force 2 September 1990).

United Nations Economic and Social Council, UN Sub-Commission on Prevention of Discrimination and Protection of Minorities. 1984. Siracusa Principles on the Limitation and Derogation of Provisions in the International Covenant on Civil and Political Rights. U.N. Doc.E/CN.4/1984/4.

United Nations Human Rights Committee. 1982. General Comment 6: The Right to Life. U.N. Doc. HRI/GEN/1, 30 July 1982.

———. 2000. General Comment 28: Equality of Rights between Men and Women. UN GAOR 2000, U.N. Doc. A/55/40, Annex VI.

United Nations Human Rights Council. 2008. Optional Protocol to the International Covenant on Economic, Social and Cultural Rights. A/HRC/8/L.2/Rev.1/Corr.1, 2-18 June 2008.

United Nations International Convention on the Elimination of Racial Discrimination, 7 March 1966, 660 U.N.T.S. 195, 5 I.L.M. 352 1966 (entered into force 4 January 1969).

United Nations International Covenant on Economic, Social, and Cultural Rights, 16 December 1966, 993 U.N.T.S. 3.

United Nations Office of the High Commissioner for Human Rights. 2008a. Ratifications and Reservations to the International Covenant on Economic, Social and Cultural Rights, 18 April 2008, available at http://www2.ohchr.org/english/bodies/ratification/3.htm (accessed on 16 July 2008).

———. 2008b. Ratifications and Reservations to the International Convention on the Elimination of Racial Discrimination, 21 April 2008, available at http://www2.ohchr.org/english/bodies/ratification/2.htm (accessed on 16 July 2008).

United States-Import Prohibition of Certain Shrimp and Shrimp Products. 12 October 1998. WT/DS58/AB/R.

Vienna Convention on the Law of Treaties, 1155 U.N.T.S. 331, 8 I.L.M. 679 (entered into force 27 January 1980).

Villagran Morales et al. v. Guatemala. 19 November 1999. Inter-American Court of Human Rights Series C, No. 63.

Weil, Prosper. 1983. Towards a Relative Normativity in International Law? *American Journal of International Law* 77 (3): 413–42.

World Health Organization. 2002. Implications of the Doha Declaration on the TRIPS Agreement and Public Health. Health Economics and Drugs Series No. 12, WHO/EDM/PAR/2002.3.

World Trade Organization General Agreement on Tariffs and Trade. 1994. 1867 UNTS 187; 33 ILM 1153.

World Trade Organization. 1994. Marrakesh Agreement Establishing the World Trade Organization, Apr. 15, 1994, *Annex 2, Understanding on Rules and Procedures Governing the Settlement of Disputes*.

———. Understanding the WTO: Members and Observers, available at http://www.wto.org/english/theWTO_e/whatis_e/tif_e/org6_e.htm.

World Trade Organization, Ministerial Conference fourth session, Doha, 9–14 November 2001. Declaration on the TRIPS Agreement and Public Health, adopted on 14 November 2001, WT/MIN(01)/DEC/2 (WTO, Doha Declaration).

WTO Appellate Body Report. 1996. United States—Standards for Reformulated and Conventional Gasoline, WT/DS2/AB/R, 20 May 1996.

———. 1998. EC—Measures Concerning Meat and Meat Products, WT/DS26, 48/AB/R, 16 January 1998.

———. 1998. U.S.—Import Prohibition of Certain Shrimp and Shrimp Products, WT/DS58/AB/R, 12 October 1998.

———. 2001. European Communities—Measures Affecting Asbestos and Asbestos-Containing Products, WT/DS135/AB/R, 5 April 2001.

WTO Appellate Body Reports. 1998. Australia—Measures Affecting Importation of Salmon, WT/DS18/AB/R, 20 October 1998.

———. 2000. Canada—Term of Patent Protection, WT/DS170/AB/R, 18 September 2000.

———. 2002. United States—Section 211 Omnibus Appropriations Act of 1998, WT/DS176/AB/R, 2 January 2002.

WTO General Council. 2003. Implementation of paragraph 6 of the Doha Declaration on the TRIPS Agreement and Public Health: Decision of the General Council of 30 August 2003. WT/L/540, 1 September 2003.

WTO Panel Report. 2000. Korea—Measures Affecting Government Procurement, WT/DS163/R, 19 June 2000.

WTO Secretariat. 1996. *United States—Standards for Reformulated and Conventional Gasoline* WT/DS2/AB/R, 20 May 1996.

———. 1997. *India—Patent Protection for Pharmaceutical and Agricultural Chemical Products*, WT/DS50/AB/R, 19 December 1997.

———. 1998. *India—Patent Protection for Pharmaceutical and Agricultural Chemical Products*, WT/DS79/R, 24 August 1998.

———. 1998. *India—Patents, EC*, WT/DS50/AB/R, 24 August 1998.

———. 2000. *Canada—Patent Protection of Pharmaceutical Products Case*, WT/DS114/R, 17 March 2000.

———. 2000. *U.S.—Section 110(5) of the U.S. Copyright Act*, WT/DS160/R, 15 June 2000.

———. 2002. *U.S.—Carbon Steel,* WT/DS213/AB/R/Corr.1.

———. 2005. *European Communities—Protection of Trademarks and Geographical Indications for Agricultural Products and Foodstuffs,* WT/DS290/R, 15 March 2005.

The State That AIDS Makes: The Moral and Political Economy of Epidemic in South Africa

Antoinette Handley

An epidemic that threatens to prematurely bring to an end the lives of one fifth or more of a country's population will change that society. But…the nature of such change and its extent are not easily forecast.

Hein Marais, *Buckling: The Impact of AIDS in South Africa*

A recent print advertisement for a South African bank featured a pair of hands cupping the proverbial crystal ball. The text read: "You can't predict the future. But with us you can price it." This might seem a peculiar observation with which to open a paper on the South African AIDS epidemic. It is less so if you consider the AIDS epidemic not purely as a biomedical event or as a social tragedy, although it is both of these things, but also as a massive economic shock to the South African political economy, a shock which presents actors with all manner of risks and uncertainties, some of which have been anticipated—and even priced.

It has become commonplace that South Africa's epidemic is one of the world's most acute (UNAIDS 2008)—yet, under President Thabo Mbeki, the country's relatively well-resourced and newly democratized government was shockingly unresponsive to the crisis. By contrast, important sections of the country's private sector began to develop their own in-house responses to the epidemic, in some instances, providing anti-retroviral drugs (ARVs) to their workers.[1] The South African case thus provides an important test of how significant shocks—such as an epidemic of this type—might serve to (re)construct the roles and responsibilities of the public and private sector respectively.

We have long known that existing political and economic conditions structure the particular features and trajectories of epidemics in

individual societies. This is true too for the contemporary AIDS epidemic. Too seldom, however, do we reverse the causal arrows and consider how, in turn, AIDS may shape the political and economic contours of an affected society. This paper addresses the mutually constitutive relationship between the epidemic on the one hand and the South African political economy on the other, focusing in particular on the responses of business and the state up until the end of 2008. Put more succinctly (and with apologies to Charles Tilly [1992]) I consider how, in AIDS-affected South Africa, the state makes epidemic, and epidemic makes the state. In particular, I am concerned with the latter, viz. how the epidemic is shaping the South African political economy, and the apportionment of responsibility for the health and welfare of both workers and other citizens.

The question of whose responsibility it is to respond to the crisis remains contested. Tony Trahar, chief executive of Anglo American, even as he was announcing in 2002 his firm's intention to roll out ARVs to Anglo's workers, insisted that "the state should come to the party in terms of providing these drugs" (Schoofs 2002). In the absence of an effective response from the public sector, there was an unspoken tussle over whose job it was to manage the economic fallout from the epidemic and to provide for the welfare needs of the country's population. The outcome of this struggle is, as yet, not clear (and its outlines have shifted under the new government of Jacob Zuma)—but the nature of that struggle in Mbeki's South Africa forms the subject of this paper. Finally, I will address how the apparently competing narratives of the political and moral economy of AIDS informed the responses of the public and private sector respectively.

But first, for those unfamiliar with South Africa's AIDS epidemic, some background information: the epidemic emerged in South Africa in the 1980s as a disease affecting predominantly white men in the country's gay community. Early warnings that the syndrome might threaten the broader population were largely ignored and in the late 1980s, a group of Malawian miners tested positive for the virus. By the early 1990s, low levels of prevalence had been detected within the broader South African population. The country passed the 5% prevalence mark in the mid-1990s, at almost exactly the same time as the country's transition to democracy. Since then, prevalence levels have soared and have yet to abate. Today, South Africa beats out Nigeria and India for the title of the country with the largest number of people living with HIV in the world (despite its population being less than one-third the size of Nigeria's and one twenty-fifth the size of India's) (UNAIDS 2008).

A. The State Makes Epidemic

A good deal of academic work has examined how the course of HIV/AIDS in a particular setting is shaped by social and political factors often quite specific to the society in question (Barnett and Whiteside 2002; Kalipeni, Craddock, et al. 2004). For example, we now understand much better how poverty and inequality facilitated the spread of the virus in developing countries (Farmer 1999; Stillwaggon 2000; Hope 2001; Sosklone and Shtarkshall 2002; Whiteside 2002; Iliffe 2006). In South Africa, as elsewhere, the nature of the economy and of the society's political and social cleavages has shaped the course of the epidemic in distinctive ways. South Africa's racialized patterns of poverty and inequality and, in particular, patterns of labour migration associated with the country's mining industry provided crucial vectors of transmission for the virus. The Medical Research Council's Mark Lurie vividly describes the unintended consequences of the migrant labour system: "If you wanted to spread a sexually transmitted disease, you'd take thousands of young men away from their families, isolate them in single sex hostels and give them easy access to alcohol and commercial sex. Then, to spread the disease around the country, you'd send them home every once in a while to their wives and girlfriends. That," he concludes, "is basically the system we have" (quoted in Schoofs 2001).

The connection between the society and the epidemic is not just about vectors of transmission or the especial vulnerability of the poor, however. It also comprises ways of thinking, discourses, and political identities that can contribute to the spread of the disease and hobble our responses to it. As Hein Marais argues (2005, p. 110), AIDS-related stigma is "hardly an aberration" but instead embodies categories used and value judgments made every day by South Africans. Lieberman and Gauri make a related point when they argue that the way in which South Africans draw political boundaries has meant that they tend to regard AIDS always as a problem for another and distinct group, say, for "gays," or "blacks," or squatters. Consequently, "South African political actors—high-level government leaders, church leaders, ordinary citizens—have portrayed HIV/AIDS as a problem for 'them,' not 'us'" (Gauri and Lieberman 2006, p. 64). Along with class, gender, sexual orientation, and conceptions of morality, racial and ethnic cleavages are central to how the virus is perceived and to how it spreads.

Society Shapes the Epidemic as the Body Shapes the Virus

The interaction between the virus itself and an individual human body serves as an illuminating metaphor for the mutually constitutive relationship between the epidemic and society. That is, we can best understand how an epidemic progresses through a society by considering how the virus enters into and mutates within an individual human body. Those who study the course of the virus in the human body tell us that the way the virus mutates in each individual human body is unique to the specificities and vulnerabilities of that body.[2] Iliffe describes the passage of the virus inside the body as follows (2006, p. 5): "Like all viruses, HIV has no life of its own but is a parasite of cells, drawing its life from theirs. …[T]he virus becomes attached to certain types of cells…that activate the body's immune system." In a similar way, but on a massively larger scale, the epidemic too tracks and exploits the susceptible contours, cleavages, and characteristics of the system (or, in this case, human society) that it enters (Barnett and Whiteside 2002, p. 455). The path of the disease through society, then, closely resembles the path of the virus in the human body: they both map onto existing pathways, rendering dangerous conditions that an otherwise healthy body would shrug off. I noted above, for example, how the Southern African migrant labour system served as an important vector of transmission. These migratory passages inadvertently provided a route for the virus to travel with the infected miners across the region to their rural homesteads.

However, the body in turn impacts on how the virus will mutate. To cite Iliffe again (2006, p. 5), he explains that "the virus enters a cell and integrates its genetic information into its host's, *using the cell's life to reproduce itself*" (italics mine). He continues: "The reproduction process is prone to error…The combination of speed and error in reproduction means that HIV mutates at about one percent per year, or a million times faster than is normal in evolution." This adaptive characteristic of the virus explains in part why it so difficult to develop a cure. Similarly, the epidemic mutates as it progresses, exploiting and further undermining a society's political weaknesses and capacity deficiencies, all the while changing and impacting the society in important ways. This "moving target" nature of the epidemic can make it a formidable foe for even the best public health systems—and it may mean that generic international policy responses will not be equally effective across different countries. Both the virus inside an individual body and the epidemic in society mutate and shift their shape depending on the idiosyncrasies of their host.

Given the nature of the virus and its interactions with the body politic, what kind of public and private sector is the epidemic making and how have the dual context of the country's recent transition to democracy and its conversion to a relatively conservative set of macroeconomic policies shaped these outcomes? In order to answer this question, we need to explore how, as one recent newspaper headline put it, the social and moral tragedy of AIDS is being recast as an economic risk, and hence as "an accounting item" (Bridge 2002).

B. Epidemic Makes the State: AIDS and Social and Economic Outcomes in South Africa

There is a substantial historical literature describing how past epidemics have shaped social and political outcomes in a number of societies, both over the long term and in a more concentrated time span (McNeil 1998; Cohn 2002; Armus 2003; Clark 2007). The idea, then, that significant demographic and socio-economic shocks, such as those embodied by devastating epidemics, impact affected societies is nothing new. Given that AIDS primarily affects men and women of prime working and childbearing age, and given also the high prevalence levels of the Southern African epidemic, it seems reasonable to assume that the AIDS epidemic in South Africa too will have a substantial and discernible impact on the nature and functioning of the country's political economy. Indeed, in the 1990s, a number of analysts predicted the direst consequences for democratic politics and economic growth in the worst affected states (de Waal 2003a; de Waal 2003b). Researchers at the Brookings Institution, for instance, argued that the impact of AIDS on the economy would effectively be like "running Adam Smith in reverse" (cited in Whiteside, McPherson, et al. 2001, p. 3; Dawes 2004). Some 20 years into the South African epidemic then, it seems timely to enquire whether these outcomes have come to pass.

Without wanting to downplay the human tragedy of the epidemic, the first observation that one can make with respect to South Africa is how many of the more alarmist predictions have failed to materialize—as yet anyhow.[3] We have not seen complete social breakdown or major political and economic collapse. Instead, what we have seen is "more of the same," that is, an exacerbation of underlying characteristics and tendencies of the political economy.[4] It is as if AIDS serves as a distorted looking glass, reflecting back at the viewer an exaggerated, occasionally grotesque image of South Africa's failings and weaknesses.

In what follows, I highlight three aspects of the impact of the epidemic. First, I argue that the patchwork nature of the health care being provided to South Africans and their limited access to ARVs both reflect and deepen the country's underlying inequalities. Similar inequalities and unevennesses are evident in the private sector response, which I consider in turn. Here I spotlight the leading role played by the insurance industry in modeling and forecasting the projected impact of the epidemic and in shaping the private sector response. Finally, I consider how the political and the moral economy have intersected in apportioning roles, responsibilities, and responses to the epidemic from the public and private sector respectively.

AIDS and the South African Economy

It is harder than one might anticipate to determine precisely how AIDS is currently shaping the broader political economy, and how it might do so in the future.[5] Analysts began trying to model the epidemic's impact in the late 1990s (see, for example, Whiteside, McPherson, et al. 2001; Bell, Devarajan, et al. 2003) and initially there was little agreement on what the economic impact of the epidemic was likely to be. Most analysts assumed that the epidemic would adversely affect economic growth, but it proved a complicated task to calculate by how much—or even to reach agreement on the precise mechanisms by which the epidemic would impact the economy (Dawes 2004).[6] How diverse populations and a range of both public and private institutions would respond to epidemic—when and how they would change their behaviour—added to the complexity of the problem.

Nonetheless, there has been no shortage of attempts to map the macroeconomic impact of the epidemic. In South Africa, economists, actuaries, and those based in the country's home-grown insurance industry have been prominent in these efforts (Quattek 2000; Arndt and Lewis 2001; Bureau of Economic Research 2001; Dixon, McDonald, et al. 2001; Metropolitan 2005; Bureau for Economic Research 2006). The efforts of the latter should not be that surprising: given their role as effective underwriters of myriad pension plans and medical insurance schemes, the insurance industry had a direct and tangible interest in calculating, as precisely as possible, the costs and uncertainties threatened by the epidemic. Moreover, insurers not only needed to understand these risks for themselves, but they had an incentive to ensure that the firms and households that they insured also understood those risks and took some measures to reduce them.

Even in the absence of injunctions from the state, these exercises, along with workplace prevalence tests, served as an "early warning" of future risks and their costs. The nature of those warnings, however, was partial, as were the private sector responses.

To the layperson, what is perhaps most notable about these models is that a number of them suggest that the economic impact of the epidemic will not be as dire as was originally expected. Indeed there are those who argue that, as a direct result of the epidemic, GDP per capita in South Africa might actually rise![7] Such a conclusion seems not only monstrous but also implausible—until one considers more closely the demographics of the South African epidemic.

The South African AIDS epidemic has had its most direct and negative impact predominantly on the poor, the unskilled, and the unemployed. On the face of it, this might blunt the shock to the economy in two ways. First, for some time, prevalence levels in South Africa were concentrated disproportionately among the lower socio-economic strata, specifically within the ranks of the unskilled and unemployed. This is important not simply because the unskilled are poor and therefore, on an individual basis, not significant consumers of goods and services. It is important also because in an economy like South Africa's, which suffers from an extreme oversupply of unskilled labour, they are relatively easily replaced, raising the horrifying prospect that large numbers of unskilled could effectively vanish from the political economy with little discernible effect except, arguably, a decline in the unemployment rate.

Continuing this line of thinking, the South African economy suffers from an extreme shortage of skilled labour. Skilled workers contracting the virus would pose a more urgent and direct threat to the economy—and sections of the private sector mobilized to neutralize this threat, bringing me to the second reason why analysts might assume that the socio-economic profile of the epidemic would mute its economic impact. Because of their access to workplace programs, private health care, and medical insurance, those scarce and valuable skilled workers are, for the most part, far better able to access effective medical treatment. The introduction of anti-retroviral medicines has meant that those able to access these drugs could extend their lives and their ability to keep working—and hence, their ability to keep consuming, paying their mortgages, and contributing to the overall functioning of the economy.

Bayer and Oppenheimer effectively describe these "unevennesses" in the health care system (Oppenheimer and Bayer 2007, p. 145). Those in unionized or white-collar jobs (i.e., those covered by some sort of medical aid provision or private health insurance) are much better off, as are those formally employed by large, institutionalized employers like the mines or the banks who elected to develop workplace AIDS programs. Those not lucky in these ways must take their chances with the public health care system,[8] a system not only battling with the usual resource and capacity constraints common to all developing countries, but one that in South Africa was, at least until December 2007, hobbled also by the fierce political battles raging over the meaning and treatment of HIV/AIDS.

It may well be that this conclusion (that the epidemic will have little impact on the formal economy) is built on a series of faulty assumptions. Not only does this logic ignore the complex interdependencies that exist between the formal and informal sectors of the economy, it also overlooks the shifting demographics of the epidemic: prevalence levels have recently begun to rise within the ranks of semi-skilled and skilled workers too (Dynes 2002). Regardless of the validity or not of its assumptions, however, the chilling implications of this line of thinking have informed the thinking and planning of a range of important economic actors. It has, however, also had to contend with the claims and contradictions of the moral economy, and to engage the broader question that Tania Li poses: "Whose responsibility is it to attend to the welfare of surplus populations?" (2009, p. 1229).

Responding to AIDS in South Africa's Neo-liberal Era: Whose Job Is It Anyway?

Before the South African cabinet agreed in 2003 to roll out ARVs in the public health system, the AIDS landscape was dotted with only small "islands of access" to care: clinical trials provided limited access to select treatments for participants, as did the clinics and initiatives run by international NGOs and the limited number of employer-provided programs (Oppenheimer and Bayer 2007, p. 157). The situation following the cabinet's 2003 decision was, for years, not nearly as dramatically different as one would have hoped, as the roll out of ARVs was staggered, and delayed by the need to set up accredited Service Points and train staff to run these. By 2007, less than one-third of those with advanced HIV were receiving access to the necessary drugs (UNAIDS 2008, p. 271).

There is nothing new about the absence of universal access to health care in South Africa. Health care provision in that country has long been a mix of public and private, and access to both has been mediated by both race and class. In the last decades of apartheid, private health care grew substantially, increasingly funded by medical aid provisions, often subsidized by employers. The public system had, of course, long been racially divided and unequally apportioned under apartheid, as all state services were.

Following the advent of democracy, the public health care system was restructured and unified, officially granting all of those who needed it the right to access public health care. (Crucial here was a provision in the new South African constitution which requires the state not only to protect and implement a range of political rights, but to do the same for select social and economic rights.) Unfortunately, the progressive constitution and policy-making preceded the readiness of the public hospitals to receive all those who now flocked to them and the result was—and continues to be—an overstretched and under-resourced public health sector.

As the public sector was de-racialized, so too was the private sector: "private health care facilities were opened to Black elites and, later, as Black labor organizations bargained for medical benefits, to workers" (Oppenheimer and Bayer 2007, p. 19). Admission to the world of private health care now became contingent on one's economic class—and in particular on one's employment status. Thus, to use Marais' formulation, we witness "the diffusion of market ethics through society" (Marais 2005, p. 66).

C. The Political and Moral Economy of HIV/AIDS

Neo-liberalism is the label most commonly assigned to recent attempts to institutionalize the spread of "market ethics" to South Africa in the post-apartheid era. Although its advocates present its prescriptions as part of a natural economic order and not as historically contingent, in South Africa the logic of the market confronts and engages also with the oftentimes contradictory demands of the moral economy. We need to explore how shades of both the moral and the political economy speak to the responsibilities of the state and private sector respectively in their response to the epidemic, but we will find that the interactions and congruencies are not always where we expect them to be.

The idea of the moral economy was famously articulated by the historian E.P. Thompson, in his discussion of food riots in eighteenth-

century England. According to Thompson, the moral economy assumed "definite, and passionately held, notions of the common weal" (1971, p. 79), and was grounded in "a consistent traditional view of social norms and obligations, of the proper economic functions of several parties within the community" (ibid.). The moral economy, then, poses a profound challenge to the political economy view of how individuals and societies can and should operate. Rather than accepting the claim that the market is the appropriate mechanism for allocating resources in society, the moral economy posits instead "a noneconomic, moral universe of solidarity and the right to subsistence" (Booth 1994, p. 654). The moral or "embedded" economy as it is also called, is one which is "'submerged' in social relations" (ibid., p. 653) and reflects the social values of the community. In such an economy, discussions about almost any substantive issue are "inseparable from ethical questions of the good life" (ibid., p. 654).

By contrast, the political economy is epitomized in a set of ideas often associated—not always accurately—with the writings of Adam Smith. This economy, Thompson explains, "entailed a de-moralising of the theory of trade and consumption...By 'de-moralising' it is not suggested that Smith and his colleagues were immoral or were unconcerned for the public good. It is meant, rather, that the new political economy was *disingested of intrusive moral imperatives…*" (1971, pp. 89–90, italics mine). Instead, "the natural operation of supply and demand in the free market would maximize the satisfaction of all parties and establish the common good" (ibid., p. 90).

Some analysts see the distinction between the political economy and the moral economy as time-bound, with the dominance of the moral economy being eclipsed by the ascendance of the political economy and the development of capitalism—or at least of a market economy—in Western Europe: "before modernity, the securing of human livelihood had no separateness—no boundary line that marked it out as distinct from the enveloping society's institutions and values" (Booth 1994, p. 653). The Great Transformation, according to Karl Polanyi (1944), changed all of that: "The 'nature of things' which had once made imperative, in times of dearth, at least some symbolic solidarity between the rulers and the poor, now dictated solidarity between the rulers and 'the Employment of Capital'" (Thompson 1971, p. 131). This produced "an atomized society in which the interdependency of individuals was not mediated through political, social or religious institutions but via the market and contract. The noneconomic bonds of the world before modernity were supplanted by those forged from economic self-interest" (Booth 1994, p. 657).

However, the reality is that this process of transformation from one economic mode to another is never truly complete or definitive. Adam Ashforth's studies of witchcraft describe the ongoing workings of the moral economy in twenty-first-century South Africa (2002; 2005). In the scarcity, insecurity, and upheaval that accompanies neo-liberal transformations, as well as those wrought by the epidemic itself, the moral economy continues to inform how people perceive and understand the epidemic, and the role of their society's political and economic institutions.[9]

The Response of Business to the HIV Epidemic

In the public debates over how to respond to the AIDS epidemic, small firms have barely featured, largely for the reason that few small firms appear to have any positive programs in place to deal with the epidemic (Wadula 2004). One commentator argued that small firms "think about the epidemic like [they think] about the weather" (Kotze 2006), that is, it is big and important but not really something that they can control. By contrast, some of South Africa's large mining companies[10] have attracted significant media attention because of their relatively far-reaching responses to the epidemic. They have since been joined by other large firms, including some of the country's banks, and companies like BMW and South African Breweries. Their responses include employer-funded workplace prevention programs and voluntary counseling and testing sites; at their most comprehensive, some have also made free access to ARVs available for their staff (and, in some instances, for the partners and immediate dependents of their staff too). These programs have not proven to be as expensive in practice as was initially anticipated,[11] but what is noteworthy is that, at the times that firms chose to take on these responsibilities, their assumption was that the costs of these programs would be significant. Not all large firms have responded in this laudable way. The construction industry, for example, presented a striking contrast, preferring often to simply replace sick workers. For those firms that have responded constructively, however, it is worthwhile to enquire why they have done so and how they regard their own role in this regard.[12]

Arguably, there are four "logics" that motivate firms to develop a proactive and broadly constructive approach to the epidemic. These logics frequently overlap but are also sufficiently distinct that they may be conceptually separated. First, firms might regard the issue as

one which is politically important to their future, for example by affecting their relationship with the unions. South African business has a long tradition of having to navigate a politically complex environment, being dominated by white South Africans but needing to ensure a long run (and multi-racial?) future for capitalism in South Africa (Whiteside 2006). These considerations led important sectors of business to support political reform in the last decades of apartheid. Arguably, the AIDS epidemic poses at least as great a threat to the long-term future of business as apartheid did, yet business leaders have done little to openly challenge the government's inaction on this issue. Instead, they have quietly sought to develop, in conjunction with the unions, a set of their own in-house responses.

What such political motivations assume, of course, is a second logic, namely, that AIDS needs to be dealt with by firms as a core business issue, one which threatens not just the long-term political sustainability of capitalism, but also their profits and investments in the short and medium term. According to the Minister of Minerals and Energy, up to 30% of all miners in South Africa may be HIV-positive (cited by Nattrass 2007, p. 139), and there is significant evidence that the epidemic is "badly affecting" the profitability of many of the country's mines (*Citizen* 2004b). Arguably, threats to profitability provide the most potent motivation to firms to react to the crisis. Brian Brink, champion of the ARV program at Anglo American, stressed the need to convince business executives that "'this is a risk that we can quantify, that we can manage and fully understand.' Then, you find the attitudes changing" (Oppenheimer and Bayer 2007, p. 187). Gavin Churchyard, a colleague of Brink's who also worked with Anglo American, highlighted the importance of research on cost-effectiveness in making the case for ARV provision within Anglo American. "Within business," he remarked, "bottom lines, that's what talks" (ibid.). This is the kind of logic which appeals to the chief executive officer of a firm, and it represents the logic of the political economy at its most basic level: a series of cost-benefit risk calculations.

On a day-to-day basis, however, the people in large firms who are thinking hardest about AIDS are often not those in the boardroom, but the staff of the Human Resources (HR) departments, which raises a third logic evident in firms' responses. The job of HR departments is to protect their companies' investments in human capital. Consequently, as disability claims and sick leave requests began to rise in large firms, HR was frequently the division that first noticed the

direct impact of the epidemic.[13] Keith Bryer, a spokesperson for BP, justified his company's decision to provide ARVs to its staff as follows: "any well-run company is interested in keeping its staff fit for work" (Altenroxel 2002).

The final logic is perhaps the dimension which comes closest to the moral economy, namely, to see responding to the issue as part of a firm's corporate social responsibility. This view is encapsulated in the view of one business executive that "a well-managed company needs to know about and to understand and to care about the kind of society it operates within" (Leroni 2006). While it is easy to be cynical about decision making by large firms, interviews with those who have made the decisions to adopt workplace programs reveal a mix of motivations at work: sure, it makes good business sense to respond to the epidemic, but there is also a strong sense that it is simply the right thing to do. This is captured by Anglo American's Brian Brink: "Everybody keeps asking us, 'What's it going to cost?' as if it's a cost-benefit analysis. The decision [to provide ARVs] was not based on economics—I mean, clearly we looked at the economics—but far more important a thing was the social imperative, simply that people are dying. You cannot stand by and watch this epidemic taking root in this country and do nothing" (Oppenheimer and Bayer 2007, p. 191). Nonetheless, it is probably also fair to say that firms would be unlikely to do the right thing if it cost too much or if it did not make good business sense.

The Response of the State to the HIV/AIDS Epidemic

Are any of the same logics in play when it comes to decision making by the public sector over how to respond to the epidemic? Popular accounts of the South African government's failure to respond effectively to the epidemic have tended to focus on a single figure to account for this outcome, then President Thabo Mbeki, scapegoating him as both extraordinary and irrational in his treatment of the issue. While there is no question that Mbeki bears major responsibility for many of the failings of the public sector response, it is important to contextualize his response and to note the strains of denialism that run far wider than one individual. After all, Mbeki's outlook on the epidemic was supported by no less than two health ministers, both of whom were medical doctors, and by key organs of the party including its highest decision-making body, the National Executive Committee (NEC) (Brand 2000),[14] and the ANC Youth League (Nattrass 2007,

p. 98). Analysts have long commended the depth and quality of leadership within the ANC, yet many key figures were notably quiet in the public debates about AIDS in the 1990s. Yes, there was a cabinet revolt in 2003 (Kindra 2003) and there were also a few courageous individuals within the ANC who opposed Mbeki's denialism or spoke out about AIDS in their own families,[15] but these incidents were singular enough to stand out, and what dissent there was emerged surprisingly late.

The broad appeal of the denialist position lies in its articulation of sentiments that emerge from the moral economy, a direct response to the racialized and unequal functioning of the political economy in which most South Africans live and work. In their defence of Mbeki, for example, the ANC's NEC argued that the Congress "would not be stampeded into precipitate action by pseudo-science, an uncaring drive for profits or an opportunistic clamour for cheap popularity" (African National Congress 2002; cited in Nattrass 2007, p. 101). A potent set of motifs in denialist discourse identify with the poor and the marginalized, with Africa and with ordinary people, and set denialists up against the powerful worlds of profit taking,[16] multinational companies, Westerners, and medical experts.

Sentiments associated with denialism are more widely held than is commonly noted; they appear even amongst the ranks of those who have subsequently become prominent AIDS activists. The National Union of Mineworkers, for example, suppressed early research in the 1980s on the sexual behaviour of mine workers for fear it would feed into stereotypes of black promiscuity (Oppenheimer and Bayer 2007, p. 193; see also Friedman and Mottiar 2005). Zackie Achmat, now perhaps the country's foremost activist in the struggle for access to ARVS, was himself not immune to denialist critiques in the early days of the epidemic. He concedes that, as a gay, black man schooled in the leftist tradition, he suspected Western governments of hyperbole when he first heard about AIDS (Power 2003, p. 57). "It seemed far-fetched," he recalled, "that a disease would conveniently kill fags, prostitutes, drug users, *and* blacks. Doesn't it sound like propaganda?"

Across the subcontinent, there has been, and continues to be, great sensitivity about the connection between HIV prevalence levels and Western perceptions of black sexuality. Uganda's President, Yoweri Museveni, is now lionized for leading his country's impressive response to the epidemic. However Putzel (2004) describes the sensitivity of the Museveni government to early indications of alarm

about the epidemic coming from white and foreign observers. Much of the appeal of the denialist position across sub-Saharan Africa derives from explicitly moral reactions to a history of racism, economic exploitation, and marginalization. Far from being irrational and uncaring, the denialist discourse, as Nattrass (2007, p. 88) points out, suggests that "Mbeki was concerned about negative normative judgements embedded in the mainstream understanding of AIDS."

It is important then to situate denialist discourse—as it manifests in South Africa at least—as a reaction to longstanding racist tropes about Africa as a hotbed of contagion and disease, and of Africans themselves as sexually promiscuous and even subhuman, close to primates (Nattrass 2007, p. 87; Squire 2007, p. 9). The insensitive way in which the debate over the origins of the virus in central Africa occurred and the connections drawn to ape forms of the virus can only have spurred the denialist impulse. As Squire notes, "in such mythological accounts, there is a liminal place in 'the heart of Africa, where boundaries between…humans and animals, break down" (2007, p. 25). In all of his speeches and writings on the subject, Mbeki was clearly reacting against racist tropes about Africa, Africans, and supposed "African" sexuality.[17]

Mbeki and his supporters reacted also to a history of Western medicine that has not always been seen as sympathetic or responsive to the health concerns of Africans and blacks. The history of racist population control policies and of the role played by the white medical profession in the mines and in apartheid prisons and security forces made South Africa ripe for the reception of conspiracy theories about the malignity of Western medicine and science (Butler 2005, p. 604). Recall accounts of the biological weapons program developed by the apartheid regime for use against those fighting for democracy (Fassin 2007).[18] Jean Comaroff observes how "HIV revivifies scarcely suppressed memories of the violence and medical neglect of times past, jibing with enduring legacies of scientific racism, material extraction, and technological dependency" (2007, p. 202). Indeed, a press statement issued from the office of then President Mbeki explicitly compared ARVs to "biological warfare of the apartheid era." The statement continued: "Our people are being used as guinea pigs and conned into using dangerous and toxic drugs" (cited in Power 2003, p. 63).

In terms of this logic, Medecin sans Frontieres' (MSF) pilot project to administer ARVs in the Cape could be attacked as a hostile, foreign plot. MSF director Eric Gemaere recalled that "the worst attack that

we got was when we started with triple therapy in May 2001, where …a spokesperson of the ANC wrote in the newspaper…that we were fighting with biological warfare. He was implying that we were a modern version of Wouter Basson, who was the Dr. Mengele here during apartheid, who was doing experiments to…sterilize Black people" (Oppenheimer and Bayer 2007, p. 180). Again, we should note that it was not just Mbeki or officials close to him who were suspicious of the project; these sentiments resonated strongly throughout the community. When MSF first set up in Cape Town, a nurse from the neighbourhood visited the clinic to investigate for herself what was going on. "If they are using Black people as guinea pigs," she commented, "as a staunch member of the ANC, I had to know" (Oppenheimer and Bayer 2007, p. 181).[19] The serious side effects and nastiness of conventional medical treatments only reinforced this mistrust (Nattrass 2007).

The willingness of people to embrace local or alternative treatments is again revealing of the twinned mistrust of biomedicine and capitalism. Nattrass describes how AIDS patients are confronted with two competing paradigms: "the scientific paradigm which presents HIV infection as an incurable disease requiring lifelong medication and the managements of side effects; and a curative paradigm offering spiritual, 'natural' and 'alternative' treatments" (2007, p. 163). The Treatment Action Campaign (TAC) understands how damaging the maintenance of these separate, competing paradigms can be—and that they pose what is essentially a false choice. The TAC has thus had to learn how to navigate between the broad popular appeal of denialism and the need to give South Africans accurate information about the virus and the epidemic. To the extent that it has succeeded in this, the TAC has played a role that is both educational and mobilizing. While the broad impact of its programs on the South African population has undoubtedly been slighter than that of the state, it has equipped members of South Africa's most marginalized communities with medical and technical knowledge, as well as fostering their ability to secure better services from their state. Crucially, the TAC enshrined in law the state's obligation to provide at least some of those services by taking the minister of health to court over the government's failure to provide a prevention of mother-to-child transmission program. As Friedman and Mottiar point out (2005, p. 528), in this way the TAC "implicitly strengthens democratisation by ensuring not only that people are able to claim their rights but also that they are better able

to participate as democratic citizens." Robins similarly suggests that AIDS activism in South Africa is creating "new forms of citizenship" (2006). Unfortunately, as I suggest below, these new citizenships are manifold, unequal, and not uniformly positive.

Conclusion

The starting point of this paper was the proposition that AIDS poses severe and potentially transformative challenges to South African society and that, some 20 years into the crisis, we ought to be able to discern the outlines of those challenges. Implicit in this assumption are questions about the extent to which the South African political economy can effectively adapt itself to the challenges posed by HIV/AIDS. All indications thus far are, for the private sector at least, that it can and probably will. The SA Institute of Chartered Accountants, for example, pushed for companies listed on the Johannesburg Stock Exchange to include in their annual financial reports data pertaining to the impact of AIDS on their staff and customers, and the steps they are taking to deal with the epidemic.[20] In other words, the particular logic of capitalism in South Africa in the post-apartheid/HIV era is hard at work, figuring out how to factor in and cost the risks and dilemmas associated with the pandemic (Temkin 2002). As argued earlier, these market-based calculations are not necessarily inconsistent with moral considerations, but such considerations are clearly separate and secondary.

The public sector, too, is grappling with the epidemic. At first glance, it might appear that government decision making—at least under Mbeki's presidency—was effectively paralyzed by considerations coming out of the moral economy. Again, however, we would be mistaken to assume that political economy and moral economy logics are unrelated. This is evident in the approaches adopted by Mbeki's government to macroeconomic policy on the one hand, and to the epidemic on the other. Investors were inclined to approve the fiscal conservatism of Mbeki's government, its aversion to budget deficits, and its determination to avoid significant levels of international borrowing. However, international observers were unable to understand Mbeki's apparently wilful perversity on the subject of AIDS.

In a persuasive polemic, Simon Barber suggests that, far from being irreconcilable, the same impulse informed both of these policy directions, namely, the sentiment that "for centuries Africa's been told

what to do; from now on, we decide for ourselves" (2002). In this view, the South African government's fiscal rectitude was not so much a wholesale conversion to neo-liberal values as evidence of the Mbeki cabinet's determination to protect South African sovereignty and autonomy in the international political economy, without the humiliating and intrusive conditionalities attached to World Bank and IMF lending. This impulse might be captured in a phrase used by the economist Gerry Helleiner and others: the radical case for conservative macroeconomics (Helleiner 2008).

If this is true, and if what the Mbeki government prioritized was financial self-sufficiency, then AIDS must have loomed as a particularly nightmarish threat to South Africa's standing and autonomy. As numerous observers have pointed out, once the scale of the South African epidemic became clear, the projected costs of providing comprehensive access to ARVs threatened to bankrupt the state and undermine the hard-won gains of years of economic reform. Particularly in the early days of the epidemic, when drug costs were still extremely high, the only response that would not have financially crippled the state was rationing of access to treatment. And, as Barber points out (2002), it just so happens that "there [was] a system already in place [to do this]. It's called the market." He elaborates: "Employers will pay to keep alive those infected with HIV who are needed for production. This will spare government not only the politically troublesome triage but also the need to abandon sound economic management and become a serf of the donors." When we examine health care provision in AIDS-affected South Africa, this was effectively the model that had been arrived at.

Of course, societies, like epidemics, are not static. And both are political. The impact of the South African epidemic has begun to be felt within the ranks of the country's more skilled strata of labour, including crucial personnel like police officers, nurses, and teachers. Politically, too, much has changed. Since 2008, Mbeki has been replaced as leader of the ruling party and president of South Africa by Jacob Zuma, and the much-hoped-for shift in public policy on HIV/AIDS appears to be underway. Accordingly, it is now resource and capacity constraints that are emerging as the obstacles to an effective public sector response. It remains to be seen then whether and how these changes will affect the broader outlines of public and private sector responses to the epidemic.

In the same way that the mutation of the virus is shaped by the individual human host within which it reproduces, so, too, has South Africa's AIDS epidemic been profoundly shaped by the contours of South African society. Conversely, we can also see how AIDS is beginning to reshape that society in turn. I have argued in particular that who falls ill and who dies from the HI virus have informed the society's response to the epidemic, and both the epidemic and responses to it are shaping and moulding the kind of state and the kind of private sector that South Africa has. The responses of these institutions embody a complex plaiting of the country's moral and political economies in often unexpected ways. The consequences will play out for decades yet to come.

Acknowledgements

The School of Public and Development Management at the University of the Witwatersrand, Johannesburg, and its director, Francis Antonie, provided me with visiting fellow status and an office in 2006, and a SSHRC Institutional Grant from the University of Toronto funded preliminary research. Two years later, a fellowship from the Comparative Program on Health and Society at the Munk Centre for International Studies, University of Toronto, provided me with crucial time and space to write this paper. Helpful comments on this paper were received from Lisa Forman, Gerry Helleiner, Evan Lieberman, Nicoli Nattrass, and James Orbinski. I am most grateful for all of this assistance.

Notes

1. Indeed, one survey found that "South African companies are doing better than some of their international counterparts in responding to the HIV/AIDS crisis" (Katzenellenbogen 2003).

2. Alex de Waal uses the inverse metaphor in de Waal (2001).

3. This caveat is epitomised by the title of Alex de Waal's latest book on the subject, *AIDS and Power: Why There Is No Political Crisis—Yet* (2006).

4. See de Waal (2006) for a powerful elaboration of this argument.

5. I must concede a real conceptual difficulty here, namely, that the transition to democracy occurred just as the epidemic was beginning to take hold in broader South African society. Similarly, the staggering dimensions of the epidemic also began to become clear at about the same time as the government was adopting GEAR, a conservative macroeconomic framework. Given this simultaneity, it is extraordinarily difficult to identify the causal arrows, to neatly "disembed" the epidemic from this wider context and sort out what individual factor is causing what outcome. Rather than trying to do so, I instead focus on how these factors interact.

6. National Treasury, for example, has argued that the country's growth would be cut by an average of 0.23% a year because of the epidemic.

7. Similar logic is expressed in an article by Grawitsky that speculates that AIDS could limit the net rise in total unemployment (2000).

8. Tania Li argues that "the key to their predicament is that their labour is surplus *in relation* to its utility for capital" (italics in original). Her argument here is about landless populations in Southeast Asia, but it applies equally well to the unskilled in South Africa (2009, p. 1210).

9. As Booth reminds us (1994, p. 662), there is no neat dividing line between world of the political economy and the world of the moral economy: "all economies…are moral economies."

10. Of course, the mines are very significantly affected by the epidemic. A range of sources agree that prevalence levels run to about one-third of all workers in this subsector. See for example *Citizen* (2004a).

11. There are a number of reasons for this. Drug costs have come down very significantly. This could have been predicted. What was perhaps less predictable was how few workers would make use of employer-funded access to ARVs. See Fraser (2004), Kahn (2005). Similarly low rates of take-up were found in the program being run by the South African National Defence Force (Radebe 2005).

12. See also Handley (2009).

13. One survey of five companies found that disability claims almost doubled over a three-year period because of HIV-related illness (Deane 2003).

14. There was some debate within the ANC's NEC, but in the Committee's public statements, it long supported Mbeki's position.

15. For accounts, see Matisonn (2001), Trengrove-Jones (2001), Kindra (2003), Michaels (2003), Squire (2007).

16. It should be noted that, among many black South Africans, mistrust of capitalism is at least as strong as mistrust of biomedicine and of the apartheid state; indeed, in the minds of many, capitalism was closely associated with the latter.

17. Even those opposing the denialists display some of these tendencies. For example, Blade Nzimande, the Secretary General of the South African Communist Party, lambasted "those *white* Americans who say there is no HIV" (Fela 2002, italics mine).

18. See also references in VirusMyth web site 2002. A version of this paper, which some argue was authored by Mbeki himself, was circulated at the ANC annual conference in 2002. This author was not able to find the document on the ANC's official website, but at the time of writing it was still posted at one of the leading denialist websites.

19. She subsequently became a strong supporter of MSF and the TAC.

20. Via the King II report, this measure has since been adopted for those companies listing on the Exchange's Social Responsibility Index.

References

African National Congress. 2002. Lend a Caring Hand of Hope: Statement of the National Executive Committee of the ANC.

Altenroxel, L. 2002. Business Goes It Alone on AIDS Drugs. *Star* (Johannesburg), August 13.

Armus, D., ed. 2003. *Disease in the History of Modern Latin America: From Malaria to AIDS*. Durham, N.C.: Duke University Press.

Arndt, C., and J.D. Lewis. 2001. The HIV/AIDS Pandemic in South Africa: Sectoral Impacts and Unemployment. *Journal of International Development* 13 (4): 427–49.

Ashforth, A. 2002. An Epidemic of Witchcraft? The Implications of AIDS for the Post-apartheid State. *African Studies* 61 (1): 121–43.

———. 2005. *Witchcraft, Violence, and Democracy in South Africa*. Chicago: University of Chicago Press.

Barber, S. 2002. Market Forces Key to AIDS Policy. *Business Day* (Johannesburg), February 5.

Barnett, T., and A. Whiteside. 2002. *AIDS in the Twenty-first Century: Disease and Globalization*. Pietermaritzburg, South Africa: Palgrave Macmillan.

Bell, C., S. Devarajan, et al. 2003. The Long-run Economic Costs of AIDS: Theory and Application to South Africa. Policy Research Working Paper Series 3152, World Bank.

Booth, W.J. 1994. On the Idea of the Moral Economy. *American Political Science Review* 88 (3): 653–67.

Brand, R. 2000. ANC Backs Mbeki in Dispute over HIV/AIDS. *Star* (Johannesburg), October 4.

Bridge, S. 2002. AIDS Risk May Become an Accounting Item. *Star* (Johannesburg), July 4.

Bureau of Economic Research. 2001. The Macro-economic Impact of HIV/AIDS in South Africa. Stellenbosch, South Africa: University of Stellenbosch Bureau for Economic Research.

———. 2006. The Macroeconomic Impact of HIV/AIDS under Alternative Intervention Scenarios (with Specific Reference to ART) on the South African Economy. Stellenbosch, South Africa: University of Stellenbosch Bureau for Economic Research.

Butler, A. 2005. South Africa's HIV/AIDS Policy, 1994–2004: How Can It Be Explained? *African Affairs* 104 (417): 591–614.

Citizen (Johannesburg). 2004a. Harmony says AIDS Could Kill 33% of Staff, October 6.

———. 2004b. AIDS is Hitting Mining Profits, Says Study, November 26.

Clark, G. 2007. *A Farewell to Alms: A Brief Economic History of the World*. Princeton: Princeton University Press.

Cohn, S.K. 2002. The Black Death: End of a Paradigm. *The American Historical Review* 107 (3): 703–38.

Comaroff, J. 2007. Beyond Bare Life: AIDS, (Bio)politics, and the Neoliberal Order. *Public Culture* 19 (1): 197–219.

Dawes, N. 2004. Assessing the AIDS Catastrophe. *Business Day* (Johannesburg), March 23.

de Waal, A. 2001. AIDS-related National Crises: An Agenda for Governance: Early-warning and Development Partnership, available at http://www.justiceafrica.org (accessed 22 August 2005).

———. 2003a. A Disaster with No Name: The HIV/AIDS Pandemic and the Limits of Governance. In *Learning from HIV and AIDS*, edited by G. Ellison, M. Parker, and C. Campbell. Cambridge: Cambridge University Press.

———. 2003b. How Will HIV/AIDS Transform African Governance? *African Affairs* 102 (406): 1–23.

———. 2006. *AIDS and Power: Why There Is No Political Crisis—Yet*. London: Zed Books.

Deane, N. 2003. Disability Claims Rise and Productivity Falls. *Mail and Guardian* (Johannesburg), September 26–October 2.

Dixon, S., S. McDonald, et al. 2001. AIDS and Economic Growth in Africa: A Panel Data Analysis. *Journal of International Development* 13 (4): 411–26.

Dynes, M. 2002. Gold Fields' Estimate of AIDS Costs Highlights Economic Costs of Epidemic. *Star* (Johannesburg), April 9.

Farmer, P. 1999. *Infections and Inequalities: The Modern Plagues*. Berkeley: University of California Press.

Fassin, D. 2007. *When Bodies Remember: Experiences and Politics of AIDS in South Africa*. Berkeley: University of California Press.

Fela, M. 2002. Men and Women of Steel Take Up Arms in AIDS War. *Star* (Johannesburg), April 30.

Fraser, J. 2004. Anglo's HIV Drive Gets Off to Slow Start. *Business Day* (Johannesburg), June 1.

Friedman, S., and S. Mottiar. 2005. A Rewarding Engagement? The Treatment Action Campaign and the Politics of HIV/AIDS. *Politics and Society* 33 (4): 511–65.

Gauri, V., and E.S. Lieberman. 2006. Boundary Institutions and HIV/AIDS Policy in Brazil and South Africa. *Studies in Comparative International Development* 41 (3): 47–73.

Grawitsky, R. 2000. AIDS Could Limit Rise of Joblessness. *Business Day* (Johannesburg), February 15.

Handley, Antoinette. 2009. Business and the Governance of HIV/AIDS in South Africa. In *Governance of HIV/AIDS: Making Participation and Accountability Count*, edited by Sophie Harman and Franklyn Lisk, 101–23. New York: Routledge.

Helleiner, G. 2008. Informal communication with author. Toronto.

Hope, K.R. 2001. Africa's HIV/AIDS Crisis in a Developmental Context. *International Relations* 15 (6): 15–36.

Iliffe, J. 2006. *The African AIDS Epidemic: A History*. Oxford: James Currey.

Kahn, T. 2004. Unspent HIV Millions "To Be Rolled Over." *Business Day* (Johannesburg), June 1.

———. 2005. Low Uptake of AIDS Drugs at Anglo Mines. *Business Day* (Johannesburg), April 22.

Kalipeni, E., S. Craddock, et al., eds. 2004. *HIV and AIDS in Africa: Beyond Epidemiology*. Oxford: Blackwell Publishing.

Katzenellenbogen, J. 2003. S.A. Firms Lead the Way in AIDS Policy. *Business Day* (Johannesburg), December 1.

Kindra, J. 2003. AIDS: Ministers Revolt. *Mail and Guardian* (Johannesburg), August 15.

Kotze, C. 2006. Personal interview with author. Johannesburg. South Africa: Independent consultant.

Leroni, B. 2006. Personal interview with author. Johannesburg, South Africa: Group corporate affairs executive, MassMart.

Li, Tania Murray. 2009. To Make Live or Let Die? Rural Dispossession and the Protection of Surplus Populations. *Antipode* 41 (6): 1208–35.

Marais, H. 2005. Buckling: The Impact of AIDS in South Africa. Pretoria, South Africa: Centre for the Study of AIDS, University of Pretoria.

Matisonn, J. 2001. Mbeki Faces AIDS Revolt from Women. *Sunday Independent* (Johannesburg), November 18.

McNeil, W.H. 1998. *Plagues and Peoples*. New York: Anchor Books and Doubleday.

Metropolitan. 2005. *Live the Future: HIV and AIDS Scenarios for South Africa: 2005–2025*. Bellville, South Africa.

Michaels, J. 2003. ANC MP Speaks Out in Defence of AIDS Activists. *Star* (Johannesburg).

Nattrass, N. 2007. *Mortal Combat: AIDS Denialism and the Struggle for Antiretrovirals in South Africa*. Scottsville, South Africa: University of KwaZulu-Natal Press.

Oppenheimer, G.M., and R. Bayer. 2007. *Shattered Dreams? An Oral History of the South African AIDS Epidemic*. Oxford: Oxford University Press.

Polanyi, K. 1944. *The Great Transformation*. Boston: Beacon Press.

Power, S. 2003. Letter from South Africa: The AIDS Rebel. *New Yorker,* May 19, 54–67.

Putzel, J. 2004. The Politics of Action on AIDS: A Case Study of Uganda. *Public Administration and Development* 24 (1): 19–30.

Quattek, K. 2000. The Economic Impact of AIDS in South Africa: A Dark Cloud on the Horizon. *Konrad Adenauer Stiftung Occasional Papers*. Johannesburg, South Africa: ING Barings.

Radebe, H. 2005. SANDF Puts Best Foot forward in War on AIDS Pandemic. *Business Day* (Johannesburg), August 10.

Robins, S. 2006. From "Rights" to "Ritual": AIDS Activism in South Africa. *American Anthropologist* 108 (2): 312–23.

Schoofs, M. 2001. Migrant Labour System Perpetuates Spread of AIDS in the Gold Mining Industry. *Star* (Johannesburg), June 27.

———. 2002. Giant Anglo American Will Supply Free Anti-AIDS Drugs to Workers. *Wall Street Journal*, August 7.

Sosklone, V., and R.A. Shtarkshall. 2002. Migration and HIV Prevention Programmes: Linking Structural Factors, Culture, and Individual Behaviour—An Israeli Experience. *Social Science and Medicine* 55 (8): 1297–1308.

Squire, C. 2007. *HIV in South Africa: Talking about the Big Thing*. New York: Routledge.

Stillwaggon, E. 2000. HIV Transmission in Latin America: Comparison with Africa and Policy Implications. *South Africa Journal of Economics* 68 (5): 985–1011.

Temkin, S. 2002. JSE-listed Groups to Report on AIDS. *Business Day* (Johannesburg), August 16.

Thompson, E.P. 1971. The Moral Economy of the English Crowd in the Eighteenth Century. *Past and Present*, no. 50: 76–136.

Tilly, C. 1992. *Coercion, Capital, and European States, AD 990–1992*. Cambridge, Mass.: Blackwell.

Trengrove-Jones, T. 2001. Speak or Be Damned. *Sowetan* (Johannesburg), May 17.

UNAIDS. 2008. *Report on the Global AIDS Epidemic*. Geneva: UNAIDS.

Virus Myth web site. 2002. Castro Hlongwane, Caravans, Cats, Geese, Foot and Mouth and Statistics. Retrieved 13 March 2008, available at http://www.virusmyth.com/aids/hiv/ancdoc.htm.

Wadula, P. 2004. Small Business Faces Big AIDS Dilemma. *Business Day* (Johannesburg), April 28.

Whiteside, A. 2002. Poverty and HIV/AIDS in Africa. *Third World Quarterly* 23 (2): 313–32.

———. 2006. Personal interview with author. Durban, South Africa.

Whiteside, A., M. McPherson, et al. 2001. The Economic Impact of HIV/AIDS in Southern Africa. Washington DC: Brookings Institution.

The Implications of Treating HIV as a Security Threat in Africa: An Analysis of the United Nations Security Council Debate on HIV

Suzanne Hindmarch

Abstract

This paper examines the United Nations Security Council's historic debate in January 2000 on HIV as a threat to peace, security, and development in Africa. It follows the trajectory of the debate from an initially expansive endorsement of human security, through contestation (particularly along North-South lines) about the meaning and policy implications of positioning HIV as a security threat, to the ultimate adoption of Resolution 1308, which focused more narrowly on HIV prevention and education programs for peacekeepers. While recognizing that Resolution 1308 expanded the role of the Security Council in significant ways, the paper contends that the debate also reaffirmed traditional meanings of security, especially those relating to sovereignty, and, by focusing on prevention programs for peacekeepers, foreclosed potential Council endorsement of more comprehensive responses to HIV in Africa (and globally), notably, right to health approaches.

Introduction

In January 2000, the United Nations Security Council held a debate on the impact of HIV on peace, security, and development in Africa, marking the first time in United Nations history that the Security Council had treated a disease, rather than violent conflict, as a threat to peace and security (UNSC 2000c). The debate culminated with the historic, unanimously passed Security Council Resolution 1308 in July 2000, which called for, among other things, the development of HIV education and prevention programs for UN peacekeepers (UNSC 2000d). Subsequent to that debate, the Council has overseen the

Department of Peacekeeping Operations (DPKO) implementation of these programs, which has involved collaboration with the Joint United Nations Programme on HIV/AIDS (UNAIDS) to develop an increasingly comprehensive HIV prevention program for peacekeepers and uniformed services of all member states (UNAIDS 2005), and held regular open-forum updates to track its progress. The Security Council president, too, has issued numerous statements reaffirming this work and encouraging further action from member states and other stakeholders (see for example UNSC 2001c; UNSC 2005a). The Council's turn to human security and its move into AIDS education has been endorsed by UNAIDS, and has resulted in new partnerships and institutional relationships between UNAIDS, the DPKO, and the Security Council. It has more recently drawn the attention of academic writers, who have, since the debate, begun to generate scholarship focusing on HIV as a security issue (David 2001; Brower and Chalk 2003; Elbe 2003; Ostergard Jr. 2007).

The interesting question here is not why the Security Council chose to frame HIV as a security threat, since that is the lens through which the Council views the world; the question is why it chose to consider HIV at all, and why, having made this choice, it was able to garner state support for new programs and institutional arrangements to carry out its HIV prevention and education work. A related puzzle is why other UN actors who were not security focused should have endorsed this point of view, allowing this framing of HIV to gain currency beyond the Security Council. In addition to looking back in time to find explanations for the Security Council's debate and the subsequent uptake of its ideas, it is also essential to look to the future, asking what the implications of treating HIV as a security threat might be.

I would like to suggest that we can see the Security Council debate as one component of a larger, ongoing contest in the UN system over the meaning of human security. At the moment of the debate, other UN bodies, particularly the Economic and Social Council (ECOSOC) agencies, were discussing HIV with reference to human rights, gender equity, poverty, and other human development issues. Rooted in an expansive understanding of human security, some of these alternate frames engage with thorny questions of global inequity, discrimination, racism, patriarchy, poverty, debt, and social justice; prescribed solutions imply more far-reaching reform, including changes to drug patent laws and trade policies, debt forgiveness, and the creation of enforcement mechanisms for international human

rights laws. In some instances these reforms, if implemented, would create global governance structures that could limit the sovereignty of states and result in significant changes to global political economies.

In contrast, the Council's framing of HIV as a security threat ultimately led to education and condom distribution programs intended to promote behaviour change among peacekeepers and soldiers as individuals. These prevention models are technical, in that they depend on the dissemination of material goods (condoms) and technical knowledge (the biomedical facts of HIV prevention). The proposed solutions do not require significant political intervention, as do the more substantive economic and legal reforms noted in the previous paragraph. Neither do they require structural reform in the UN system.

Furthermore, the debate, and especially the HIV prevention strategies that followed from it, served to reinforce the "South-ness" and particularly the "African-ness" of HIV in the Security Council and DPKO's work. HIV was framed as something both threatening to, and emanating from, African states and militaries, and requiring a change in the behaviours and sexual proclivities of (largely developing-nation) countries that send peacekeepers to African conflicts. This framing precludes significant consideration of the political and economic practices of non-African states that might contribute to the spread of HIV in sub-Saharan Africa. Such practices include trade policies that aggressively promote the international expansion of developed-country industries, firms, and agricultural producers, while maintaining tariff barriers that limit developing-country access to lucrative markets in North America and Europe; the use of tied aid to promote donor countries' products and interests; and, more germane to this paper, agreements like TRIPS that limit the abilities of generic drug producers to manufacture and distribute HIV medicines.

Thus the debate, while unprecedented in its consideration of a disease as a security threat, resulted in solutions that in some ways had less to do with preventing the spread of HIV among Africans, peacekeepers, and uniformed services than with inhibiting the spread of alternate approaches to the AIDS pandemic in the UN system, approaches that would entail a higher political and economic cost for developed states. While treating a non-military issue as a threat to security was initially welcomed as a progressive move to a human security frame, in fact, the Security Council debate brought the concept of human security, at

least as applied to HIV, back in line with a traditional understanding of security in ways that advanced (or, at least, did not inhibit) the interests of developed states. At the same time, though, the DPKO-UNAIDS partnership that resulted from the debate successfully expanded the scope of the Council's work, providing it with increased authority to participate in HIV education activities that had previously been the exclusive domain of ECOSOC, and especially UNAIDS.[1]

Before proceeding to the analysis, two clarifications are needed. First, "HIV as a security threat" is not the only, or even the dominant, approach to HIV in the UN system. While the outcomes of the Council debate, particularly the UNAIDS-DPKO partnership, have probably limited the extent to which either body can employ alternate interpretations of human security in their HIV prevention work with uniformed services, this contest of meanings is ongoing.

Second, the following analysis suggests that the position of some states in the debate, and especially that of the United States, was motivated primarily by instrumental and strategic concerns. However, these state interests and motivations must be distinguished from those of the debate's initial champion, American Ambassador Richard Holbrooke. Notwithstanding the limitations of the programs resulting from the debate, Holbrooke's ability to persuade the Council to discuss HIV in the first place, and subsequently to undertake substantive action, was an impressive diplomatic accomplishment. Significant political will was required and political capital expended, both domestically and within the UN, for him to accomplish this work, and I do not mean to suggest that Holbrooke himself promoted this debate for purely instrumental reasons.

Why the Debate Occurred: Origins

Most accounts suggest that Ambassador Holbrooke was the driving force behind the Security Council's decision to debate HIV (David 2001; Faluso 2004; Prins 2004; Traub 2006). When Holbrooke assumed the rotating presidency of the Security Council in January 2000, he immediately declared that this was to be the "month of Africa," wherein the Council would debate a wide range of factors contributing to conflict and insecurity in Africa, including HIV, and consider strategies to promote peace, security, and development. This was a largely unilateral endeavour, and one that other Council members greeted with reactions ranging from mild resistance to bemused tolerance (Prins 2004, p. 941).[2]

Holbrooke appears to have been motivated at least in part by a genuine personal concern with the impact of HIV on Africa, possibly as a result of a 1999 visit to a Zambian program for AIDS orphans (Prins 2004, p. 941). His friendship with then Secretary-General Kofi Annan (Traub 2006, p. 139), who championed the continent throughout his time at the head of the UN, may also have been a contributing factor to his decision to hold a "month of Africa," though this cannot fully explain the decision to include HIV as topic during that month.

The timing of the debate vis-à-vis American domestic politics also suggests some instrumental reasons for Holbrooke's decision. The issues discussed in the "month of Africa" were reasonably consistent with the Clinton administration's foreign policy, but at odds with the ascendant Republican Party's position on involvement in the UN and on financial support for African peacekeeping missions. James Traub argues that Africa "offered the grossest evidence of US disengagement from its international responsibilities…It was hard enough as it was [for Holbrooke] to counter the view of his African colleagues that the United States and the other Western powers were willing to send 100,000 soldiers to the Balkans but barely a platoon to Africa" (2006, p. 143). The "month of Africa" can be seen in part as an attempt to appease troop-contributing states, and African states more generally, that were critical of the American failure to support African peacekeeping, and to demonstrate to the international community that the United States was still committed to the United Nations as an institution, in spite of recent Republican claims to the contrary.

Beyond domestic concerns, we must also consider the larger question of outcomes at the level of the UN: why was framing HIV and human security in this way so successful? Why and how did the debate succeed in creating new education programs for peacekeepers and new institutional arrangements between the DPKO and UNAIDS? Thomas Weiss (2004) has suggested that the post-Cold War Council has been increasingly motivated by a "humanitarian impulse," by which he means both an increased willingness to intervene in conflicts on humanitarian grounds and an embrace of an expanded understanding of human security. I contend, however, that part of the reason for the outcome of the debate, even if not intended by Holbrooke, is precisely that it *limits* the extent to which and the manner in which the Council can promote human security. While the move to AIDS education does expand the scope of the Council's work,

this expanded agenda was articulated in ways that carefully define the outer limits of human security, consequently limiting the types of HIV-related humanitarian claims that can be made by states. It is these limitations that made Resolution 1308 and subsequent work palatable to the permanent members of the Council, and facilitated the widespread institutional uptake of the notion that HIV is a threat to security in Africa and in peacekeeping.

Evidence for this claim can be at least partly discerned through a careful analysis of statements made to the Council during the debate. Specifically, two related arguments in the debate can be traced to illustrate the ways in which a new issue, HIV, became the basis upon which more traditional security and especially political economy interests were advanced, while foreclosing other possible interpretations of and responses to the AIDS pandemic. First, by following arguments about the definition of human security, and competing notions of sovereignty that underpin those definitions, we can discern how the narrow definition that was ultimately adopted protected particular state interests. Second, by considering how HIV was constructed first as a specifically African issue, and then as a threat to peacekeepers, we can see how attention to non-African states, and to global political economies, was deflected, while also further narrowing the scope of the debate.

The Debate:[3] An Overview

The January 2000 debate began with an open discussion on the impact of HIV on human security in Africa, in which all member states were invited to address the Council, and to which both the president of the World Bank and the head of UNAIDS were invited (UNSC 2000a; UNSC 2000b). The Council subsequently adopted a resolution in July 2000 that called for the development of HIV education and prevention programs for its peacekeepers, thus formally marking the Council's decision to become, at least by proxy, an AIDS educator (UNSC 2000d). Since then, DPKO and UNAIDS have signed a memorandum of understanding to collaborate on HIV prevention programs for peacekeepers, regular follow-up meetings have been held to track the DPKO's progress in implementing these programs, and the Security Council president has issued numerous statements reaffirming the Council's commitment to HIV prevention in peacekeeping.

When addressing the Council in 2005 to provide an update on the progress of their work, both the DPKO Under-Secretary-General Jean-

Marie Guéhenno and the UNAIDS Executive Director Peter Piot acknowledged that their joint activities had resulted in the increased interdependence of the two bodies. As a result of their partnership, DPKO and UNAIDS each have staff seconded to the others' organization, they jointly deliver programs including pre-deployment and in-theatre AIDS awareness training, the distribution of male and female condoms, peer education programs, the provision of voluntary counselling and testing facilities, and peacekeeper-led outreach activities with local populations. Additionally, UNAIDS is working with 53 states to develop comprehensive HIV programs for their uniformed services and has also signed partnership agreements with 15 ministries of defence (UNSC 2005b, pp. 3–7).

As a result of Resolution 1308, then, there has been a significant expansion in the sorts of activities the Council supports, both institutionally and on the frontlines of peacekeeping missions. However, it is crucial to note that from an initial focus on the impact of HIV on Africa in January 2000, there was a shift in focus to HIV prevention in peacekeepers by the July 2000 resolution—an extraordinary narrowing of the debate's scope, and one made all the more remarkable by the extent of support expressed for this narrow understanding of the ways in which HIV might pose a security threat to Africa.

The Debate: Human Security

Indicating the extent to which the Security Council's treatment of HIV as a security threat has been endorsed by other actors, UNAIDS director Peter Piot stated in a January 2001 address to the Council, "I can tell the Council that its deliberations on AIDS have been enormously helpful. I do not think that can be overestimated. The simple fact that the world's ultimate tribunal on questions of peace and security devotes its attention to AIDS sends a very powerful message...The Security Council has helped transform the way in which AIDS is viewed. Only when we understand AIDS as a fundamental issue of human security can we grasp the extent of destruction it has caused" (UNSC 2001a). But what exactly has been the nature of this transformation, and to what extent has it marked a genuine turn to a human security agenda?

Reviewing the debate, it is apparent that several different concepts of human security were articulated, contested, and, ultimately, resolved in a manner that favoured a narrow rather than an expansive

definition of security, and one that favoured developed–rather than developing–country interests. Significantly, the term "human security," while widely employed during and after the debate, does not appear in Resolution 1308 (or, indeed, in any Security Council resolution). Embedded in the debate, we also find a reaffirmation of the "South-ness" and particularly the "African-ness" of HIV; and an eventual focus on peacekeepers rather than on other HIV-affected populations, or more generally on the connections between HIV and state stability.

The overall tone of the January 2000 debate is suggested by an anecdote recounted by Holbrooke in his final address to the Security Council: "My friend to my right, Sir Jeremy Greenstock [the UK Ambassador], passed me a hand-written note in the middle of the meeting, saying, with characteristic British understatement, 'I dare say this is the first time the word "condoms" has been used in the Security Council.' Now we just throw that word around" (UNSC 2001a). This image of a mature, highly educated diplomat passing notes like a schoolboy because the subject under discussion involved the word "condoms" indicates that discussion of HIV was not only unprecedented for the Security Council as a whole, but was also, at least at first, embarrassing and uncomfortable for some of its individual members. This in turn suggests a lack of familiarity with the issue on the part of some representatives, or at best a lack of experience talking about the facts of HIV transmission, even 20 years into the epidemic.

To be sure, in many ways the debate *was* a significant departure for the Council. Up to this point, the Council had understood security threats to be military in nature, involving violent conflict or the threat of violent conflict between states (and, more recently, within them). In contrast to the state-centric, geographically specific understanding of security embedded in this definition, member states used the debate on HIV to introduce human security analyses into the Security Council's dialogue. Such analyses focus on issue areas rather than on geographically bounded conflicts, and put individuals rather than states at the centre of analysis (MacFarlane and Khong 2006). Because most human security work comes from the human development tradition (see, e.g., Commission on Global Governance 1995; United Nations Development Programme 1994), analyses of inequity and social justice can typically find greater purchase here than in traditional realist security studies approaches (e.g., Waltz 1979). However, even within the human security tradition, there are debates

about just how expansive the concept is and ought to be (MacFarlane and Khong 2006). These competing interpretations of human security are clearly evident in the debate, with endorsements of narrower or broader interpretations of security breaking down largely, though not exclusively, along North-South lines.

Al Gore, addressing the Council as the United States representative, said that "this meeting demands of us that we see security through a new and wider prism and, forever after, think about it according to a new and more expansive definition" (UNSC 2000a). Other representatives took similar positions, and speakers from both the North and the South made reference to the role that war, the breakdown of states and social relations, extreme poverty, and inadequate health care and education play in facilitating the spread of HIV, and the role that HIV plays in perpetuating poverty and underdevelopment. In this sense, all speakers seemed to endorse an expansive definition of security that paid significant attention to human development concerns. However, in their prescriptions for resolving the threat that HIV poses to security and development, there are marked differences in the solutions proposed by the North and the South.

In outlining the course of action that must follow from the recognition that HIV is a security threat, Gore stated that "we know that the first line of defence against this disease is prevention…We also must do much more to provide basic care and treatment…This requires affordable medicine, but also more than medicine. It requires that we train doctors, nurses, and home-care workers, that we develop clinics and community-based organizations to deliver care to those who need it" (UNSC 2000a). He additionally outlined the investment in biomedical research that the "Clinton-Gore" administration had proposed to Congress.

To be sure, Gore's call for strengthening human and community capacity to respond to the epidemic does have radical potential, especially if he is referring to an inclusive "we" that encompasses African states as equal partners, rather than a paternalistic developed-country "we" who must save a continent. However, it is notable that, in his address, he privileges prevention over care and treatment and, when discussing treatment, focuses on issues "beyond medicine"—the strengthening of service delivery systems in the South and research investments in the North—rather than on the provision of medication

per se. Similarly, the representatives from the United Kingdom, France,[4] Netherlands, Canada, Australia, Italy, and Japan used their addresses to highlight their financial contributions to UNAIDS, the DPKO, development assistance in Africa, and vaccine research in their own countries; they did not significantly engage with the question of treatment access. In spite of their apparent embrace of an expansive human security agenda, even in this initial debate the developed nations tended to support a prevention- rather than a treatment-based solution to the threat of HIV.

More intriguingly, even in this first debate, in which the agenda was "the impact of AIDS on peace and security in *Africa*" (emphasis mine), several delegates seemed to interpret this to mean the impact of HIV in conflict settings, and especially in peacekeeping. References to peacekeeping in the January 2000 meeting were, overall, fairly general; where it was mentioned at all, it was in a sentence or two of a much longer statement. A typical example is the Canadian delegate's observation that "all peacekeepers, international observers and relief aid workers need to fully understand, both personally and professionally, the risks associated with inappropriate sexual behaviour" (UNSC 2000b). Even this remark seems incongruous in a debate that did not even refer to peacekeeping in its title. By the six-month follow-up session to this debate, when the Council passed the resolution that committed it to becoming involved in HIV prevention, the sexual behaviour of peacekeepers was overwhelmingly the focus of discussion for developed-country delegates. Proposed solutions to the problem of HIV in Africa (rapidly being reframed as the problem of HIV in peacekeepers), were now not just prevention-focused, but also peacekeeper-focused: they included the voluntary, confidential testing of peacekeepers prior to deployment, the development of educational materials for peacekeepers, and increased institutional coordination between the Council, the DPKO, and UNAIDS.

In contrast to this myopic view of the impact of AIDS on Africa, two of the three African ministers of health (from Uganda, Namibia, and Zimbabwe) who attended the January 2000 meeting, and the majority of African delegates who addressed the Council, understood human security much differently. First, their proposed solutions were for all Africans, not just peacekeepers, and second, they were emphatic in the need for improved access to HIV medication and for recognition of the right to health.

The Namibian health minister, for instance, addressed Gore directly in the first debate: "We call on your government to take the lead in mobilizing the pharmaceutical industry to work with African Governments and the World Health Organization for more affordable access by Africa to life-saving and life-enhancing drugs for the treatment of AIDS" (UNSC 2000a). She went on to link security directly to drug access issues, noting that "although it is known that there are drugs that can prolong and improve the quality of life, African governments cannot afford them. Therefore, in our view, security needs to be visualized as part of a complex of issues affecting the manner in which we perceive and deal with socioeconomic and political problems" (UNSC 2000a). While including a reference to the importance of HIV prevention in peacekeeping, the vast majority of her address to the Council dealt with drug access. The Ugandan minister, too, called for developed-country assistance in improving drug access. He acknowledged the success of behavioural change and ABC ("abstain, be faithful, use a condom") programming in reducing HIV incidence in Uganda, but emphasized that, although Uganda was being held up as a "success story" in the debate, the prohibitive cost of medications was preventing the country from providing adequate care to its HIV+ population.

In the subsequent July 2000 meeting, African and developing-country delegates continued to emphasize treatment access and the right to health, even as developed states shifted their focus to the sexual behaviour of peacekeepers. The Namibian delegate reiterated that "prevention alone is not sufficient, given the large numbers of people already infected with the virus. Due to a lack of resources and the inaccessibility of HIV drugs, not much progress has been attained in the treatment of the disease" (UNSC 2001d). In contrast, developed countries retreated from their initial embrace of an expansive definition of security and a concurrent expansion in the scope of the Council's work. Holbrooke justified the narrowing of the Council's focus by observing that the proposed resolution "focuses appropriately on the area where the Security Council has primary responsibility and the most at stake, in particular in addressing the impact of AIDS on peacekeeping…it is a fact that without proper training, education and steps towards prevention, peacekeepers may also be spreading AIDS inadvertently" (ibid.). He also observed that "this draft resolution in no way infringes on the sovereignty or authority of countries" (ibid.), a statement that betrays some of the

underlying strategic concerns that shaped the outcomes of the debate.

Normatively, the concept of human security can be expansive and leaves ample room for attention to a range of development, governance, and social justice concerns. Once it was pared down to a functional core, however, human security for the Security Council meant a narrow focus on providing peacekeepers with condoms and HIV education, and on regulating their sexual behaviour. The only reference to treatment comes at the end of the resolution, where the Council "expresses keen interest in additional discussion among relevant United Nations bodies, Member States, industry and other relevant organizations to make progress, *inter alia*, on the question of access to treatment and care, and on prevention" (UNSC 2000d)—hardly the forceful endorsement of the right to health and improved treatment access that African representatives had called for. What is more, in spite of perpetuating the "African-ness" of the disease, a point to which I return below, the resolution fails to acknowledge high rates of HIV in African civilian populations, who are the majority of Africans affected by HIV.

Why does it matter, with reference to HIV, whether human security is understood in its narrower rather than its broader sense? When HIV is seen as a security threat because of the sexual practices of individual peacekeepers, the solution becomes individual behaviour change, education, and efforts to regulate and control individual transgressions. But when human security encompasses perspectives like the "right to health" and treatment access, the HIV pandemic is framed as a structural problem. Here, it is the socio-economic conditions under which transmission occurs that are foregrounded, and the focal point is global health care infrastructure, medical service delivery systems, and above all global inequities in access to treatment. By framing HIV transmission as a right to health issue, far-reaching economic and political reforms are indicated, ones that might intrude on state sovereignty by compelling states, against their will, to make structural and legal changes to facilitate access to health care services. In contrast, the solution to HIV as a narrowly defined security threat proposes a solution no more radical than the status quo, plus condoms.

The resolution that the Council passed effectively limited the extent to which treatment access could be linked to security concerns, both at the time and in the future, by entrenching this narrower version of

human security in policy. This in turn has placed treatment issues, including patent laws and access to medicines, beyond the scope of the human security agenda as interpreted by the Security Council. It has, furthermore, restricted the Council to a concern for the health of peacekeepers, rather than the other millions of Africans living with or affected by HIV.

It would be far too simplistic to argue that a narrow version of human security was endorsed by developed states because they were protecting the interests of corporate actors. However, I suggest that one of the reasons for the uptake of the "condoms and prevention education" solution is that it does not require structural reform, and is therefore not a very domestically costly approach for the Council's permanent members to endorse. What is fundamentally at issue here is the implication of a broad understanding of human security for state sovereignty and global governance—and, especially in the context of this debate, implications for patent laws, pharmaceutical company profits, and developed states' funding commitments to UNAIDS and other multilateral health and development actors. Conceptions of sovereignty become particularly acute when discussed at the level of the Security Council, because it has more extensive enforcement power and authority to make legal declarations (Ratner 2004, p. 602). Were they to adopt a definition of human security that acknowledged the right to health, developed states would put themselves in a position where their national laws and policies could be challenged on legal grounds, albeit on the grounds of unenforceable (but still normatively important) international law. The cost to sovereignty that this would entail is unacceptable to most states, and certainly to the permanent members of the Council.

The Debate: The "African-ness" of HIV

As noted above, a second difference between traditional and human security approaches is that in the case of the former, threats are contained in specific geographic locations: conflicts occur in specific countries, or across specific border lines. In contrast, HIV is a security threat that is located in people. It is at once everywhere and nowhere. By recognizing HIV as a security threat, the Council seemed to indicate a willingness to change the scope of its operations from geographically and temporally located conflicts to a scale at once much smaller (human bodies) and much larger (the world entire) than that at which it had previously operated.

However, the debate, by virtue of its African focus, began by implicitly locating HIV in African bodies. Then, by taking a non-geographically specific human security issue but retaining the traditional security approach that treats threats as geographically bounded (but threatening because they might spill beyond their state borders), the Council's debate and proposed solutions reaffirmed the "South-ness" and the "African-ness" of the disease in ways that may further preclude genuinely global efforts to mitigate the impact of HIV. Holbrooke, in acknowledging the global nature of the pandemic, argued that "[AIDS] is not just an African problem…we have to recognize that while interdependence provides economic opportunities, it also poses global threats. You cannot deny AIDS a visa; you cannot embargo it or quarantine it. You cannot stop it at a border. That is why we must work together" (UNSC 2001d, p. 5). In this construction, HIV is "not just an African problem" not because it occurs elsewhere, not because the socio-economic conditions in Africa that contribute to disease transmission might have their origins elsewhere, but because it cannot be contained within African borders.

In the context of a debate on HIV and Africa, this linking of the disease, the continent, and peacekeeping implies one of two things. First, it suggests that African peacekeepers are themselves likely to be HIV+, and so may be a vector through which the disease can spread beyond its African borders. Alternately, it may suggest that peacekeepers stationed in Africa are at risk of contracting HIV from a dangerous, infected civilian population. In either case, the "African-ness" of the threat is reaffirmed through these constructions. In neither case is there much supporting empirical evidence.

One might object that a focus on the sexual behaviour of peacekeepers is not exclusively African in focus, even if, by nesting the discussion of HIV and peacekeeping within a larger debate on Africa, this was strongly implied. It is true in principle that peacekeepers can come from, and be deployed to, any country. However, in practice most of the large post-Cold War peacekeeping missions have been in African states, and, increasingly, responsibility for those missions has devolved to African troop-contributing countries. In January 2001, for instance, there were 39,061 UN staff on 16 peacekeeping missions, 4 of which were in Africa. Of these staff, 13,748 were on African missions. More than nine thousand (9,287) of all staff were from Africa, of whom 6,930 were serving in Africa. Of the top 10 contributing countries, which sent 21,078 people (54% of all troops), 3 were African and the

majority were developing nations.[5] Hence, contemporary discussions of peacekeepers and peacekeeping are largely (though not entirely) *de facto* discussions of Africans. Crucially, even where they are not African, the majority of troop contributors are developing nations, which means that even when the Council's gaze does shift from Africans to all peacekeepers, it is still the sexual behaviours of developing-country troops that are under scrutiny—and, because of the linking of HIV to African civilian populations in areas where these troops might be stationed, the "African-ness" of the threat remains even here.

What are we to make of this portrayal of HIV as an African disease? In part this reflects an epidemiological truth: of the 29 million people who were estimated to be living with HIV in 2001, 20.9 million of those (of whom 60% were female) were in sub-Saharan African countries (UNAIDS and WHO 2007).[6] Having a powerful body like the Security Council recognize the devastating impact of the pandemic on African states and citizens drew attention to what many would argue was still a neglected issue. However, recognizing that the cost of HIV is borne disproportionately by Africans need not lead to the conclusion that the *solution* to this problem is also uniquely African in nature, nor that it is most logically addressed through a focus on peacekeepers.

This fact was not lost on developing-country representatives. The Indian ambassador, as a representative of a major troop-contributing country, was highly critical of suggestions that peacekeepers were vectors of HIV transmission, saying that "not one Indian peacekeeper has either arrived in theatre in Africa with AIDS or left with it" (UNSC 2001b, p. 13). He noted, correctly, the absence of any compelling epidemiological evidence to suggest that a large number of peacekeepers are HIV+, or that HIV transmission increases in conflict and/or peacekeeping missions.[7] Indeed, while there is evidence that peacekeeper-civilian sexual contact had occurred in UN missions to Cambodia, Yugoslavia, and Mozambique (Slim 1997), the logical leap made by many delegates—that HIV transmission was a significant problem in peacekeeping missions in Africa or staffed by Africans—was and continues to be unsupported by evidence.

Furthermore, as the Malawian delegate noted (UNSC 2001d, p. 23), linking HIV specifically to conflict and peacekeeping obscures the fact that the southern African states with the highest rates of HIV (of

which Malawi is one) are almost all free of violent conflict. It was clear to the Malawian delegate that an approach to HIV in Africa that focuses exclusively on peacekeepers would be of minimal assistance to his state, and would certainly not reach the majority of Africans living with HIV.

Implications and Significance

In summary, while the meaning of security was initially contested in the debate, this was followed by a move to narrowly technical and prevention-focused solutions and a program of education for peacekeepers that was endorsed by UNAIDS through its partnership with the DPKO. What we have here is a debate that is, on the one hand, novel, involving the unprecedented treatment of a disease as a threat to peace and security, and the expansion of Council responsibility as it moved from policing and enforcement into AIDS education. On the other hand, the debate also perpetuates the construction of HIV as an African problem and defines human security in a limited manner that continues to incorporate elements of traditional security approaches, preventing the consideration of far-reaching structural or political-economic reform, restricting interventions to a very limited segment of the global HIV-affected population, and focusing on prevention rather than treatment issues.

In more general terms, what are the implications of this debate? Above all, this case shows us the potential, but also the limitations of human security approaches, and serves as a cautionary example of what can happen when a health and development issue is viewed through a security lens. The human security approach to HIV, in this case, has been neither as radical as the concept's human development advocates have hoped; nor, as its critics have feared, has it led states too far astray from traditional security concerns. Rather, it appears to have affirmed some of the fears of critical and post-development writers (Deudney 1990; Whiteside, de Waal, and Gebre-Tensae 2006), who have expressed concern that applying a security lens to traditionally non-security issues promotes narrow approaches to development.

Within the United Nations, the United Nations Development Programme (UNDP) has been perhaps the strongest proponent of an expansive human security agenda. The UNDP and other state and civil society actors have contended that framing health and other issues as threats to human security constitutes a progressive move (United Nations Development Programme 1994). As we saw in the initial

debate, many observers expressed the hope that, when framed in human security terms, approaches to HIV would become more holistic, promoting more equitable global programs of care, treatment, and support, and recognizing the role that macroeconomic forces, gender inequity, poverty, migration, and other socio-economic phenomena play in facilitating the spread of the virus.[8]

In contrast to the human development advocates who welcomed the turn to human security, some security studies writers have remained deeply suspicious of the concept and of the treatment of HIV as a security threat. MacFarlane and Khong, for example, argue that "it is organized individuals with the capability and intent to injure and kill that constitutes the quintessential security threat" (2006, p. 250). In this view, disease is not an agent with intent, so it cannot be understood to threaten state security. Even where it may increase state instability and therefore vulnerability to conflict, these authors contend that HIV is too far removed from the ultimate outcome of violent conflict to assert the existence of a causal relationship between the two; at best, HIV *may* be a contributing factor to insecurity given a variety of other conditions (ibid., p. 249). They further argue that redefining HIV as a security issue provides no additional analytical leverage, and that "the indiscriminate expansion of the scope of security threats stretches the concept to the point of incoherence" (ibid., p. 264). The Security Council debate, however, seems to be an instance where human security rhetoric was ultimately used to reaffirm traditional approaches to security and to stifle more radical and more costly approaches such as the right to health. While concerns about the analytical utility of treating HIV as a security threat may have merit, the Council debate, contrary to the fears of these authors, has not resulted in a substantially revised conception of security and threats to security.

The triumph of a narrow understanding of human security in the Council suggests particular causal arrows for organizational learning. It may initially have seemed that the Security Council's position on security had evolved and become more inclusive of human security concerns, perhaps as a result of a move to human security elsewhere in the UN (Weiss 2004). In actuality, it seems that this debate, and its aftermath, is an instance where the Security Council brought the idea of human security back in line with more traditional security concerns and with the interests of powerful states in the North. Rather than being a case of institutional learning on the part of the Security

Council, the debate may be seen as the Security Council partially absorbing the language of human security, but then reconstituting it in a way that strips the term of its radical normative implications. Then, through new institutional arrangements with other UN actors, particularly the DPKO and UNAIDS, the Council redeployed this new version of human security in a way that limits the ability of those actors to employ more expansive understandings of human security in their work with peacekeepers.[9]

We might also regard this case as an illustration of institutional constraints. The UN was created at a time when security and development issues were seen as separate and distinct, whereas they are increasingly—certainly in the human security literature—seen as mutually dependent. As a result, while Security Council discourse now includes references to development and human security, its structure and mandate remain rooted in traditional military approaches to security. In other words, it is not just the agendas of states that constrain Security Council approaches to HIV, but the institutional framework within which those discussions occur.

As a health and development issue, HIV is not a Security Council responsibility; only when treated as a threat to security does it become unequivocally a matter for Council attention. Conversely, delivery of services to peacekeepers clearly falls under the mandate of the Council and DPKO, and no other UN body is in a position to insist upon health programs in the context of peacekeeping missions. This institutional division of labour places both the Security Council and UNAIDS in a quandary. In order to legitimately discuss HIV within its mandate, the Council must frame it as a security problem. In reducing HIV to a security threat, crucial issues of international political economy, state governance, and power relations are placed beyond the scope of the debate. Yet, if these and other development issues are brought into the debate, it is no longer an entirely appropriate topic for the Security Council. Similarly, while UNAIDS and other ECOSOC bodies can legitimately discuss HIV as a development issue, peacekeeping and security are beyond their mandates—or at least were beyond their mandates at the time of the debate. Institutional structures thus force a choice between an optimal response and one that conforms to institutional and bureaucratic imperatives. Even if individual Security Council members had no vested interest in promoting a particular interpretation of security, given these constraints there may be organizational limits to the extent to which

expansive human security approaches could be institutionalized, at least at the Security Council level.

Conclusions

Alex de Waal has argued that "'war' on AIDS has more to do with the supposed warriors than their proclaimed struggle. It signals an attempt to continue business as usual" (2006, p. 108). He made this observation with reference to the African state, as part of his attempt to explain why high rates of HIV in many of the sub-Saharan countries have not led to the governance crises and state instability that many observers had predicted. We might, however, equally apply this observation to the Security Council and its treatment of HIV as a security threat. Though the debate was unprecedented in its consideration of a disease as a threat to peace and security, the proposed solutions are in many ways "business as usual" in their focus on technical, behavioural, and individual interventions endorsed by the North rather than engagement with African-endorsed solutions that would require broader structural reform. In another sense, the debate is "business as usual" because it reflects the institutional constraints that limit actors in the UN system even where there is a genuine desire for innovation. However, there is also a more encouraging way to regard the debate as "business as usual": while the outcomes may have been more limited than right to health advocates had hoped, the debate is also an example of the ongoing contest of meanings in the UN system, and of the commitment of some states to continue pushing for a more comprehensive and effective response to the global HIV pandemic.

Acknowledgements

I am grateful to the Ontario Graduate Scholarship Program, the Lupina Foundation, and the Comparative Program on Health and Society for providing financial support and the opportunity to workshop this paper. I would also like to thank Antoinette Handley and Lisa Forman for their comments on this paper and the larger dissertation project of which it is a part. My thinking on this issue has evolved substantially in the two years since this paper was written, due in no small part to their insightful comments and critiques, provided over the course of many conversations since 2007.

Notes

1. For more on the tendency of bureaucracies to expand in ways that promote their authority and legitimacy, see Barnett and Finnemore (2004).

2. Some permanent members, and subsequently some of the non-members who addressed the Council during the debate, argued that HIV was well beyond the Council's mandate and best addressed by the General Assembly and the Economic and Social Council (ECOSOC). The ongoing power contest between the Security Council and ECOSOC, although not fully explored in this paper, is also an important backdrop to the debate.

3. For the sake of brevity, in the rest of the paper I use "the debate" to refer both to the initial debate that occurred in January 2000, and to the subsequent Council meetings, discussions, and resolutions on the topic of HIV and security in Africa. Thus, I am conceiving of the debate as ongoing, though I suggest that most of the decisive moments involving the negotiation of terminology and the scope of policy recommendations occurred early in the debate.

4. France, however, subsequently initiated multilateral discussions on treatment access and discussed the importance of equitable drug access in a 2005 address to the Council.

5. These and all subsequent peacekeeping statistics are from the UN Peacekeeping Monthly Summary of Contributors of Military and Civilian Police Personnel for January 2001 (available at http://www.un.org/Depts/dpko/dpko/contributors/jan.htm, last accessed 29 June 2008). Numbers include troops, police, and observers. While African missions make up a minority of the total missions at any given point in time, cumulatively they make up a large proportion of post-Cold War missions.

6. These and subsequent HIV statistics are from 2001, rather than the year the debate was held, because this is the earliest year for which there are complete data sets for both UNAIDS and DPKO statistics. However, these are the 2001 estimates that were retroactively revised downwards by UNAIDS and WHO in December 2007. While the difference between the original and the revised estimates is significant for program planners, I do not believe that the content or outcome of the debate would have been radically different if it had been known that approximately 29 million, not 32 million, people were HIV+ in 2001.

7. A recent Lancet article (Spiegel et al. 2007) has suggested that there is currently no evidence to support the claim that conflict contributes significantly to the spread of HIV. To be fair, this information was not available to the Council at the time the resolution was passed, but it further calls into question the utility of this narrow framing of HIV as a security threat in Africa.

8. For a detailed analysis of the many environmental factors contributing to the spread of HIV in the developing world, see Stillwaggon (2006).

9. The assumption here is that because the Security Council framing of HIV has gained traction and influenced the work of UNAIDS and DPKO, it has also limited the extent to which other approaches, such as the right to health, can now ascend and shape the work of these agencies.

References

Barnett, M., and M. Finnemore. 2004. *Rules for the World: International Organizations in Global Politics*. Ithaca: Cornell University Press.

Brower, J., and P. Chalk. 2003. *The Global Threat of New and Reemerging Infectious Diseases: Reconciling US National Security and Public Health Policy.* Virginia: RAND.

Commission on Global Governance. 1995. *Our Global Neighbourhood: The Report of the Commission on Global Governance*. Oxford: Oxford University Press.

David, M. 2001. Rubber Helmets: The Certain Pitfalls of Marshalling Security Council Resources to Combat AIDS in Africa. *Human Rights Quarterly* 23 (3): 560–82.

Department of Peacekeeping Operations (DPKO). 2007. UN Peacekeeping Monthly Summary of Contributors of Military and Civilian Personnel for January 2001, available at the DPKO website, available at http://www.un.org/Depts/dpko/dpko/contributors/95-05.htm (accessed 24 January 2008).

Deudney, D. 1990. The Case against Linking Environmental Degradation and National Security. *Millennium* 19 (3): 461–76.

de Waal, A. 2006. *AIDS and Power: Why There Is No Political Crisis—Yet*. New York: Zed Books.

Elbe, S. 2003. *Strategic Implications of HIV/AIDS*. Adelphi Paper 357. Oxford: Oxford University Press.

Faluso, L. 2004. *An Insider's Guide to the UN*. New Haven: Yale University Press.

MacFarlane, S.N., and Y.F. Khong. 2006. *Human Security and the UN: A Critical History*. Bloomington: Indiana University Press.

Ostergard Jr., R.L., ed. 2007. *HIV/AIDS and the Threat to National and International Security*. New York: Palgrave Macmillan.

Prins, G. 2004. AIDS and Global Security. *International Affairs* (London) 80 (5): 931–52.

Ratner, S. 2004. The Security Council and International Law. In *The UN Security Council: From the Cold War to the 21st Century,* edited by D. Malone. Boulder: Lynne Rienner.

Slim, H. 1997. The Stretcher and the Drum: Civil-Military Relations in Peace Support Operations. In *Beyond the Emergency: Development within UN Peace Missions*, edited by J. Ginifer. London: Frank Cass.

Spiegel, P.B., et al. 2007. Prevalence of HIV Infection in Conflict-affected and Displaced People in Seven Sub-Saharan African Countries: A Systematic Review. *Lancet* 369 (9580): 2187–95.

Stillwaggon, E. 2006. Determinants of the HIV Pandemic in Developing Countries. In *The HIV Pandemic: Local and Global Implications*, edited by E.J. Beck et al. Oxford: Oxford University Press.

Traub, J. 2006. *The Best Intentions: Kofi Annan and the UN in the Era of American World Power*. New York: Farrar, Straus and Giroux.

UNAIDS. 2005. *On the Front Line: A Review of Policies and Programmes to Address AIDS among Peacekeepers and Uniformed Services*. New York: Joint UN Programme on HIV/AIDS (UNAIDS).

UNAIDS and WHO. 2007. UNAIDS *Epidemic Update: December 2007*. New York: Joint UN Programme on HIV/AIDS (UNAIDS).

United Nations Department of Peacekeeping Operations. 2008. Monthly Summary of Contributors (Military observers, civilian police, troops) as of 31 January 2001, available at http://www.un.org/Depts/dpko/dpko/contributors/jan.htm.

United Nations Development Programme. 1994. *Human Development Report 1994*. New York: Oxford University Press for UNDP.

United Nations Security Council (UNSC). 2000a. Security Council 55th year, 4087th meeting, 10 January 2000. UN DOC S/PV.4087.

———. 2000b. Security Council 55th year, 4087th meeting (Resumption 1), 10 January 2000. UN DOC S/PV.4087 (Resumption 1).

———. 2000c. Security Council Holds Debate on Impact of AIDS on Peace and Security in Africa. UN DOC SC/6781.

———. 2000d. Security Council Resolution 1308. UN DOC S/RES/1308.

———. 2001a. Security Council 56th year, 4259th meeting, 19 January 2001. UN DOC S/PV.4259.

———. 2001b. Security Council 56th year, 4259th meeting (Resumption 1) 19 January 2001. UN DOC S/PV.4259 (Resumption 1).

———. 2001c. Statement by the President of the Security Council. UN DOC S/PRST/2001/16.

———. 2001d. Security Council 55th year, 4172nd meeting, 17 July 2000. UN DOC S/PV.4172.

———. 2005a. Statement by the President of the Security Council. S/PRST/2005/33.

———. 2005b. Security Council 60th year, 5228th meeting, 18 July 2005. UN DOC S/PV.5228.

Waltz, K. 1979. *Theory of International Politics*. New York: McGraw-Hill.

Weiss, T.G. 2004. The Humanitarian Impulse. In *The UN Security Council: From the Cold War to the 21st Century*, edited by D. Malone. Boulder: Lynne Rienner.

Whiteside, A., A. de Waal, and T. Gebre-Tensae. 2006. AIDS, Security and the Military: A Sobering Appraisal. *African Affairs* 105 (419): 201–18.

The Codex in Historical Perspective: Food Safety Standards and the Codex Alimentarius Commission (1962–1973)

Brigit Ramsingh

Abstract

The Codex Alimentarius, or "food code," emerged in 1962 as a joint effort of the United Nations' World Health Organization (WHO) and Food and Agriculture Organization (FAO). Inspired mainly by the work of its European predecessor, a group known as the Codex Europaeus, which existed from 1958 to 1962, these two agencies enlisted member states, scientific experts, and industry observers to assist in the drafting of food standards for specific commodities, additives, hygiene, pesticide residues, methods of analysis, and sampling. The early years of the Codex were marked by tensions, including interagency tensions, disputes between the European regional group and the new international Codex, and, finally, disagreements over the question of what counts as "correct science." This paper will describe these main sources of tension which attended the international Codex during the early stages of its formation.

Introduction

Reflecting upon the first decade of the Codex Alimentarius Commission, its Chairman, John Davies, wrote to his colleague at the Food and Agriculture Organization in Rome: "I think we can now say we have got a satisfactory, varied and significant block of standards through; this is what we simply had to do at the last session if the Codex was to mean anything. Now the standards have got to be sold to Governments. This, I conceive, will be your main task in the next few years and I don't really think there is much I can do to help. Lack of response by Governments is the second rock on which the Codex might founder—though not in my time, I imagine!"[1]

Since the early 1960s, this Codex was published as a joint effort by the World Health Organization (WHO) and Food and Agriculture Organization (FAO) and did not founder once introduced to governments—in fact, its membership in the first 10 years grew from 30 to 98 countries and continued to grow and serve as an instrument for setting international food standards (Beacham 1973, pp. 10–11). At the time of Chairman Davies' reflection in 1969, the Codex had developed a variety of standards for products ranging from honey to lard, canned fruit and vegetables, rendered pork fat, and fish sticks; it established food labelling guidelines, standards for scientific methods of analysis, as well as tolerances for pesticide residues. By 1985, a resolution passed by the UN General Assembly strongly urged that the Codex serve as the basis for informing national food policy of governments worldwide. By the mid-1990s, the World Trade Organization (WTO) identified the Codex as a key reference point for scientific food standards, and the result was a shift in its impact on international public health and trade policy, often culminating in disputes over scientific evidence that pitted the EU against the United States (Jukes 2000; WHO/FAO 2005). Long before the recent controversies over growth hormones in milk and meat which plagued the Codex in the 1990s, however, there existed disagreements in developing and approving standards in the early years, many rocks upon which it could have foundered. This paper will address what were the main influences in the creation and early years of the Codex.

Several tensions attended the formation and early years of the founding of the Codex. Using historical archival evidence, I will explore some of these tensions, centred around three broad categories: (1) interagency tensions, particularly between the WHO and the FAO; (2) tensions between the predecessor Codex Europaeus and the new international one; and, finally, (3) tension over what counts as "correct science" within a highly economic and political context. For the latter category, in this paper I will limit my discussion to the area of food hygiene, although there were several other key food safety areas (such as pesticides and additives) which were also the focus of the WHO and FAO.

International food standards from the beginning were very much negotiated and constructed, and the result of a long process that unfolded within a climate of intense political, economic, technical,

and scientific interest. Codex standards, especially those dealing with safety, captured the interests of scientific and medical experts because of their impact on human health. At the same time, however, the post-war increase in agricultural trade and new "food regime" which emerged in 1947—one which favoured national regulation and authorized both import controls and export subsidies—meant that states could exercise power through the tool of subsidized exports of surplus commodities (Friedmann 1993). As a consequence, amid all the other interests at play in developing standards, there was also the overarching tension between the explicit Codex twin goals of protecting consumer health and facilitating international trade.

Furthermore, the emergence of an international standards body also connects to a broader post-war trend recognized among sociologists as "making food count," that is, the statistical and numerical approach to world food supplies and distribution, as seen in many of the programs initiated by organizations like the FAO (Ilcan and Phillips 2003). Even the Codex standards as they were constructed and defined, and demonstrated below with the particular case study of food hygiene, began to acquire a "more statistical profile" (Hardy 1999).

Food standards have a long history, but in their modern form their origin is largely rooted in the mid-nineteenth and early twentieth centuries. Early attempts to standardize food were often in response to or culminated in various pieces of legislation—particular acts aimed at preventing food adulteration—such as those that were passed in France (1851), England (1860), Canada (1875), and the United States (1906).[2] Food safety has had different meanings and Anne Hardy has made this distinction in outlining the "historic roots" of it, organized around two types of illnesses associated with food: (1) one associated with adulteration (e.g., deliberate contamination, addition of foreign substances like chemicals); versus (2) the illness caused by foods which had undergone decomposition, putrefaction, or decay—the products of which were once thought of as chemical toxins and then later in the nineteenth century referred to in some circles as ptomaine poisoning (Hardy 1999, p. 294). Alongside these frameworks of food adulteration and putrefaction, however, there also arose a growing awareness of the role of the microbe in making food unsafe, and by the late nineteenth century the term "food poisoning" not only emerged, but also implicated bacteria as a causal agent. The causal agents of food additives and pesticides, however, have their roots in the older history of food adulteration in the former chemical sense, as, since the

middle of the nineteenth century, food chemists and analysts had been turning their microscopes to food to develop composition tables and chemical food standards (Guillem Llobat 2008; Hierholzer 2007; Charnley 2008).

The end of the Second World War marked a watershed for food safety standards. The rise of intensive farming techniques, increased focus on the health of developing populations, increased worldwide trade, the industrialization of food systems, increased reporting of food borne illnesses, and emergence of surveillance networks for problems like salmonellosis all meant that the time was ripe for coordinated international efforts in establishing food standards.

Although the Codex in its modern form has attracted attention from various academic circles, particularly among disciplines whose focus is on the contemporary political economy of institutions, international relations, trade, and international food law, few historians have sought to characterize the origins of the Codex before its alliance with the WTO. Histories of the organizations that authored the food code are abundant. The WHO, which turned 60 years old in 2008, has attracted a number of historians, many of whom are flocking to its archives to describe various programs, or, sometimes the *failure* of programs, for example, the Malaria Eradication Program (Siddiqi 1995). This work adds to the myriad of internal histories told by retired staff, and has also established network of historians through initiatives like the Global Health Histories project (WHO n.d.). There has been significant attention paid to the WHO's predecessor organization, the League of Nations Health Organization, and the history of its work with expert groups and institutes on biological standards—penicillin, insulin, and syphilis tests (Bud 2007; Murnaghan and Talalay 1992; Mazumdar 2003).

Similarly, historians have begun to examine the FAO and its various programs, particularly its role in post-war development and campaigns like the Freedom from Hunger initiative, or the trajectory of the Protein Advisory Group, devised to address the perceived post-Second World War protein gap (Bunch 2007; Ruxin 2000). Amy Staples has identified this post-war period (with the rise of the WHO, the FAO, and the World Bank) as a unique historical moment she calls the "birth of development" (2006).

Following on some of these themes and approaches, I suggest that through their construction, standards come to represent more than the

thing itself—whether the item under scrutiny be a unit of electrical resistance, a pound of platinum, an Ampoule of penicillin, a can of string beans, or a block of cheese. An examination of the tensions in the early Codex reveals how there are stories behind the standards: they are cultural artefacts of the entire process, they embody all of the tacit knowledge that goes in to their development, they are symbols of scientific expertise, authority, national, and economic interests, and they can be a result of a political, economic, epistemological, cultural, or ideological struggle or discussion when one standard is created or adopted over another. Bearing this in mind, the role of scientific standards within a highly politicized and economic environment such as the Codex is intriguing, and the first rock upon which the Codex could have foundered was simply getting the international group started.

The Early International Codex

The international Codex was conceived as a result of an offstage conversation at a London Food Law meeting in 1960, partly out of frustration at the then "chaotic" state of international food standards.[3] The first Joint Food Standards Conference was held in October 1962 in Geneva, and brought together 45 countries and 25 international organizations. *Paris-Match* magazine carried a one-line statement on the conference: "En preparation, un Codex international (O.M.S.) pour tous les produits alimentaires." As one scientist wrote wryly to his colleague: "In this single line they managed to make two mistakes: omitting FAO and speaking of 'all' food products."[4]

Certainly, the initial aims of the Codex were not to handle all food products, but rather to focus more on those which had been processed and were ready for the market, for trade, and for consumption by the consumer. Moreover, there were other groups working on food standards, such as the International Organization for Standardization (ISO), a private industry body, and the Codex was created to help simplify and integrate existing food standards and provide a mechanism for obtaining government acceptance of these standards, through their publication in its *Codex Alimentarius*.

The members of the Codex were drawn from countries that already held a membership in the FAO or WHO. The first formal meeting of the Codex occurred in Rome at FAO Headquarters in June 1963. Here, the group adopted its general principles, most of which were inspired by the European Codex, or Codex Europaeus. These general

principles emphasized the nature of Codex food standards, and how they had to be verifiable by "scientifically proven methods" in order to be acceptable for adoption.[5]

The 120 participants included representatives of 30 countries and observers from 16 international organizations. It was a mixture of scientists and civil servants, recruited from their nations' ministries of health, agriculture, fisheries, industry, and trade. Representatives from international organizations also served on these expert committees alongside national delegates, and among others, included groups like the aforementioned ISO, the United Nation's Economic Commission for Europe (ECE), the International Dairy Federation (IDF), the Association of Official Agricultural Chemists of North America (AOAC), the International Federation of Margarine Associations (IFMA), the International Vine and Wine Office, and the International Association for Cereal Chemistry.[6]

Aside from the UN's ECE, the presence of these industry and interest groups at the early Codex meetings points to a larger issue of the time period, related to the post-war industrialization of food and agriculture and the increased use of food additives and pesticides in order to preserve supply for farther distance and longer-term consumption.

The Codex subcommittees (which would often focus on a particular commodity, like fish and fish products, or a broader area, like food hygiene) drew upon the expertise of international multidisciplinary teams of science professionals: microbiologists, chemists, public health inspectors, veterinarians, and medical doctors.

During these first few annual meetings, the Commission ironed out some guidelines for its standards:

> Standards should specify the following criteria, as appropriate: (1) *Product designation, definition and composition*: These should describe and define the food (including its scientific name when necessary) and cover compositional requirements which may include quality criteria; (2) *Hygiene requirements:* These should include such factors as specific sanitary and other protective measures and safeguards to assure a sound, wholesome, and marketable product; (3) W*eight and measure requirements:* such as fill of container, weight, measure or count of units based o*n an appropriate method or criterium;* (4) Labelling requirements: These should include specific requirements for labelling and presentation; and (5) *Sampling, testing and analytical methods:* These should cover specific sampling, testing and analytical procedures.[7]

Once these standards were developed they would be distributed to the nations' ministries of health and agriculture, trade, and industries in the form of advice to governments. It is worth emphasizing how this was a multi-stage process: drafts would go back and forth filtered through the hands of many experts and member governments, eventually ratified by the Executive of Codex before they could be distributed.

Built into the Codex structure were regional coordinating committees to ensure that regional interests were protected, despite the decidedly international flavour of Codex. The first regional coordinating committee was for Europe, and it was not until the late 1970s that regional committees for Asia and Africa emerged. Developing countries were very keen to adopt Codex standards as a way of facilitating exports to worldwide markets, however, given the advisory (e.g., not mandatory) nature of Codex standards, not to mention the cost of implementing standards, their adoption did not always benefit developing countries, and as we will see below with the European Codex, developing countries were often perceived as having lower quality standards.

Interagency Tensions: Relationship between the FAO and the WHO

Tension is common within or between large agencies or organizations, and thus it is not surprising that there existed tensions between the FAO and the WHO as they designed the new Codex in the early 1960s. Although formal agreements existed, in practice the lines were not always clearly drawn between their respective jurisdictions over the standards, as will be seen later in the case study of food hygiene. The WHO explicitly placed greater emphasis on food safety and its health effects—domains such as food additives, food hygiene (pathogenic contamination of food), pesticides—whereas the FAO had a broader interest in the standards. The Assistant Director-General of the FAO, P. Terver, described these separate but related duties when he wrote to Dr. Dorolle, the Deputy Director-General of the WHO: "The lines of interest in each House in the Program are clear enough and no difficulties have arisen in the scientific and technical points concerning the Program. In essence, as I see it, all health aspects of food standardization fall to WHO and all technological and trade aspects to FAO. I would however like to make the following suggestion so as to simplify the handling of broad questions of

common interest and to ensure, especially when dealing with outside bodies, that cooperation between our two Houses not only in fact runs smoothly but also appears to do so."[8] (His suggestion was to have a single central contact point for day-to-day administration.)

By 1966, Dr. M.G. Candau (Director-General of the WHO) wrote to B.R. Sen, the Director-General of FAO: "I am happy to note that the health aspects of the joint programme are increasingly being recognized as an essential component of this activity and that the Codex Committee on General Principles has re-confirmed that the food standards should, in the first place, aim at protecting the health of the consumer."[9]

The early Codex efforts, however, were marked by real budget strains between the two agencies, and there were requests sent back and forth between the Directors-General to discuss and secure funding. The Codex was funded by a voluntary trust fund which member states made contributions to, but in the first few years, "as a result of the rapid and significant progress" made by the Codex, "acute shortages of staff and funds have arisen" and continued to exist within its first decade, and brought forth some complaints from member states.[10] The lack of resources meant there were early problems with the processing of documents, as well as translation and distribution of the standards to member states, which became all the more labour intensive with the gradual introduction of Spanish translation in the end of the 1960s.[11]

Even administratively within the houses, however, there were snags and misunderstandings about the role and behaviour of the Codex members. The same FAO Assistant Director-General, P. Terver, received an angry letter from the Chief of the Food Standards Programme, Graham Kermode, upon learning that someone had leaked information about staffing and reorganization to delegates from member governments. He wrote:

> Attempts have been made by certain persons in FAO to enlist the support of governments to their point of view…I want to say that I deplore this sort of behaviour. It is ungentlemanly and it is time that a number of persons accepted the view that we are all employed in FAO to be international civil servants, not politicians…I am extremely angry about this; I regard it as being unscrupulous and unfair and it is not a practice to which anyone in the whole of [the Program and Budgetary Service] would have stooped. It is this sort of behaviour which lowers the reputation of FAO and one which I can say I have never encountered before in ten years of experience in the British Civil Service in Whitehall even among British politicians.[12]

Like these tensions between and within the agencies, there also existed early tensions between the Joint international Codex and its European predecessor, the Codex Europaeus. The tensions were reflected in the diverse concerns voiced by its members: namely those of cost, of European autonomy, and of trade, along with concerns over safeguarding and protecting public health. The WHO addressed some of these concerns, highlighting the distinction between the two Houses as it responded to concerns from Professor Otto Högl, the president of the European regional group, who feared that the health aspects would not be covered adequately if left to FAO. He was reassured that "action will be taken to ensure that WHO is represented at those meetings of the Joint FAO/WHO Codex Alimentarius Commission and its Expert Committees which are set up for drafting standards where it is considered essential that the public health aspects shall be safeguarded."[13] But this written assurance was often times not enough for the European Codex.

Tensions with the European Codex

The European Council of the Codex Alimentarius was established following a conference in Paris in April 1958, and had formally existed for only four years. Certain members of the European regional group expressed concerns over the question of merging with the international Codex. Particularly vociferous were the representatives from France, Switzerland, and Austria, who feared that the standards set for commodities from other nations would not be on par with European requirements, and could not compare to those of Europe not only in terms of quality, but also in terms of safety. The first president of the European Council of the Codex, Dr. Hans Frenzel, at the end of his term in 1962, posed the question to his delegates: "Why do we want a connection to this large and powerful organisation?"[14] But the answer to this question proved to be less than straightforward, and the path to union with the UN agencies was not a smooth one.

By way of response, the FAO/WHO Codex attempted to strike a balance by constructing international standards that would also allow for regional differences, or a combined hybrid regional and worldwide approach. This attempt at a shift of focus from regional to worldwide standards would haunt the new Codex leadership during its first few years.

European Council (1958–1962)

The European Council of the Codex Alimentarius had its roots in a number of institutions and was a joint project of like-minded predecessors. One view was that the initiative for sponsoring this body came from the Commission Internationale des Industries Agricoles (International Commission for Agricultural and Food Industries, abridged [CIIA in French]), an international organization of private industrialists founded in 1934 which worked closely with the Bureau Internationale Permanent de Chimie Analytique, founded in 1912 and constituted in 1923. Some of the other founders did not agree with this perspective.[15] The CIIA was an international and inter-governmental organisation created by 49 states and, by 1949, it was recognised by the United Nations Economic and Social Council (ESC) not only as an intergovernmental organization but also as a vocational body. That same year, an agreement for co-operation and work relations was established with the UN's Food and Agricultural Organisation (FAO), which granted the CIIA consultative status.

Working closely with the CIIA was another predecessor to the European Codex originating from Austria. The idea of an Austrian Codex had been around since the late nineteenth and early twentieth centuries, but there was little support for it until the mid-twentieth century once under the energetic guidance of Dr. Hans Frenzel, a minister in the Austrian government. Frenzel said of his efforts: "Often I believed myself like an itinerant preacher who moves through the land Germany, Poland, France, Holland, Switzerland, and in the native country Austria. I was not lonesome on these ways, followers were found in all lands. But on all these traveling journeys Dr. Wildner accompanied me as a faithful Palladin. Thanks are to be said to him for his loyalty to the idea."[16]

It is important to note that there was another key region which contributed to the emergence of the Codex: Latin America. A chemistry congress in 1924 had proposed the drawing up of a Código Latino Americano de alimentos, launched under the Leadership of Dr. Carlos C. Grau of Argentina.[17] A portion of the Latin American Code was included and absorbed into the international Codex (general provisions, definitions, additives lists), but the establishment of a coordinating committee for this region did not formally occur until 1976.

The European Codex was meant to be only a provisional ad hoc organization with a very firm founding principle not to create a new organization, but to work within the auspices of member governments. (This would later prove to be a sticking point with the international Codex.) The experts from 19 of the member countries had dealt mainly with preservatives, dyes, additives, honey, cocoa products and chocolate, fruit and vegetable preserves, jams, and jellies.[18] The secretariat functions were covered by the CIIA and the presidential office was maintained in Vienna, where most of the meetings were hosted.

On April 8, 1962, just as the International Codex was forming, Dr. Frenzel decided to step down as president of the European group, and passed the leadership on to his Swiss colleague, Professor Högl. Addressing his Council in a farewell speech, and perhaps foreshadowing the problems to come, he admitted in his "non-binding private opinion" that he could never agree with the fact that one does all sorts of work and then starts completely from the beginning under "quite a new flag." He said, "we are ready with pleasure—this was said repeatedly—to work in agreement with existing professional circles—but can you imagine that a completely new, foreign circle can continue our efforts, without having acquired in the last four years the spirit of our unselfish work?"[19] He went on to suggest that the issue of money would not be in line with the same unselfish spirit of the Codex that he tried to promote. Finally, he pointed to "the question of the independence, that the present circle of experts continues the begun work, this question…remained unanswered."[20] It was a question his successor would have to address.

Högl's Term (1962–1964)

The year 1962 was not only the year in which Frenzel stepped down as president of the European Council, it was also the year in which the first Joint Food Standards Conference was held in Geneva between the WHO and the FAO to lay the groundwork for an international Codex. Members of the European group were in attendance, and the idea (as least in the eyes of the UN agencies) was that the Europeans would be subsumed by the larger organization and form a subcommittee—a Regional Coordinating Committee for Europe—controlled under the jurisdiction of the principles as set out by the FAO and WHO. The new Codex would absorb certain bits of work already completed by the Codex Europaeus, namely, the definitions of foodstuffs and the

descriptions of how to go about collecting of samples for analysis. The new president of the European Council was in attendance at two key meetings of the new Joint FAO/WHO program: one, in 1962, and the other, the first session of 1963. But merely being absorbed by the larger international body and acting as an "advisory group" did not sit well with the European Council.

Nevertheless, the worldwide approach, the idea that a standard could be "used as a passport for food products in international trade, valid for entry into all countries," had great appeal.[21] But how would a regional approach to standards work within a worldwide body?

The 1962 conference stressed the need for both worldwide standards and for those of primary interest to a specific region or groups of countries, stating that international food standards are "largely conditioned by similar food habits. As a result, international trade in food is often localized within regions but may also cut across regional groups. In some cases, therefore, a standard will be required for a given region but in other [cases] by groups of countries belonging to more than one region or even for worldwide use. Health aspects, being of the widest interest, will usually need to be handled on a worldwide basis."[22]

As the representative from France similarly reasoned: "The concept of region should not be understood in its strictly geographical sense and can be applied to countries with similar ecological, economic and social conditions having brought about closely related food habits. These countries with similar standards of living and whose inspection services have similar structures and comparable methods will naturally be led to enforce unified rules in trade."[23]

Implied in this reasoning, however, was the idea that European standards were inherently higher. In a hypothetical example, the same French representative argued:

> It is quite conceivable that a given region, Europe for example, will try to define the acceptable purity criteria—let us say for jams—and that another region, Africa for instance, will be working on the same problem. If the European requirements are stricter than the African ones, what will this situation lead to?…Foodstuffs complying with the strictest standards will circulate without difficulty from the standpoints of consumer health protection and of the existence of fair practices in food trade in the countries of the region which have adopted them. They should meet no limitations dictated by

> considerations of consumer health protection of fair trade practices when exported to countries of the other region, Africa in this case, which would have adopted less strict standards: He who can do more can do less.[24]
>
> Regarding foodstuffs complying with African standards, they could circulate in the African countries which will have adopted these standards. They could be admitted into Europe—in our theoretical example—only inasmuch as they comply with the European standards, going beyond the African ones.[25]

Other arguments had little to do with health at all, but more to do with adulteration and fraud in the classic sense, as the French representatives suggested: "the guarantee of honesty in commercial transactions must be insured against any fraud without incidences of the medical…the WHO for example is badly prepared to guarantee the quality and the origin of a cognac, a whisky or even a high-class wine or *foie de gras*."

Echoing all of these concerns, Professor Högl constantly wrote letters to the Directors-General of the WHO and the FAO, and attempted to schedule private meetings separately with each, a process that became cumbersome and tricky for the agencies when trying to respond as one common voice to the European President of the Council.

As a WHO representative wrote to his FAO colleague about a private meeting with the Professor: "I think this visit could be really worth while especially in view of the European Council's comments on the draft Rules of Procedure for the Codex Commission which give added emphasis to regionalisation and do not follow the October Conference's recommendations on this important point."

One month before the first joint meeting of 1962, Paul Lamartine Yates, the FAO Regional Representative for Europe was accosted with a surprise visit by a team of Europeans led by Professor Högl, wanting to inform him of the European position on the Codex. Yates wrote to his colleague, the Assistant Director-General:

> They assert that, on account of the different levels of development in different parts of the world, inter-governmental agreements on food standards must first be worked out on a regional level among governments which are more or less at the same stage of advancement. They say, for example, that a set of standards which would be extremely high and ambitious for a group of Latin American countries would be much too lax for Europe. They therefore suggest that the

> task of the World Commission should be to secure uniformity in methodology and procedures while, under its general auspices, regional committees or commissions should be entrusted with the work of developing food standards appropriate to the member countries of the region.[28]

At this meeting, the European reps also raised concerns over the financing of the initiative—they felt the WHO and FAO should cover the costs and not individual countries. Finally, they urged the importance of working closely with the EEC and ISO.

In response to this news, the Assistant Director-General could not hide his agitation at the Swiss President of the European Codex because

> Prof Högl has…given you only one side of a problem with many facets. You will see had he carefully studied the many FAO and WHO papers on these points he might have wasted less of your time. Unfortunately we have had the same experience with him before. The FAO/WHO had already agreed to the regional approach, and thought that a strictly international approach would be 'unworkable'. At the recent FAO conference this was adopted after debate, 20 out of 21 countries agreed with the exception being France. Various papers and proposals had been drafted, and the idea of a regional approach within an international Codex was not only unanimously supported, but also submitted to the Common Market (the EEC), the ISO, OECD and even to Professor Högl![29]

And, he continued, definitely all food standards work affecting Europe must be done "hand in glove with the Brussels Organization." As for finance:

> Within three months after publication of the report, the first contributions began to come in to the Special Trust Fund by which the program is to be financed, and which are still coming in. To date we have received some $36,000 for 1962 alone from Denmark, Netherlands, United Kingdom, Poland and Switzerland. $1,000 has also been received from an overseas country: New Zealand. Eight other countries are in contact with the FAO as to their contributions and details from these are expected any day. In light of these facts Professor Högl's remarks to you that "European governments might be unwilling to contribute significant sums" seems very quaint indeed." …The present European Council of the Codex has used funds from various national sources (Minster Frenzel, the president until this year was auditor general of Austria). Professor Högl has informed us that he and his office are costing the Swiss government something in the

> region of $30,000 a year. To do nothing should cost nothing. To do a lot always costs quite a bit.[30]

And so began the rough few years during Professor Högl's term between 1962 and 1965, when the European resistance to joining Codex loomed below the surface. At official meetings of the international Codex, there would be heated debate followed by a tentative agreement to collaborate, but at the European Codex meetings, back in Vienna or Berne, the terms of union were brought into question once more.

The report from the 1962 meeting went before the 16th Session of the World Health Assembly in early 1963, with recommendations that the Codex be officially brought into being under the auspices of the two UN agencies. At this meeting in Geneva, the FAO representative sang the praises of the new Codex as an "excellent example of interagency co-operation, stress[ing] that it had no financial implications for WHO."[31]

Although no debate had been anticipated, an unforeseen "crisis" emerged as the Austrian, Dutch, and Swiss delegates "immediately expressed reservations, asking that the programme be financed by the regular budget, and not by a trust fund; that it be financed by government contributions alone; that more emphasis be given to the existing regional set-up and that the initiative rest with the regions; and finally, that the health aspects of the programme be given priority."[32]

The meeting had to adjourn this discussion and call in the Food Standards Officer-in-Charge—Frank Townshend—from Rome to address these questions, having already intervened in the European regional meetings at The Hague to bring about a deal with the regional group.

The officer in charge of the FAO/WHO food standards program, Frank Townshend, wrote to his colleague Dr. John Harvey at the US FDA for advice on how to deal with Professor Högl. After Högl began to call his own meetings of European-only Codex representatives, and fearing that this might result in a "European" camp forming against the US and non-European committees, Townshend wrote to Harvey: "His approach is strictly, even narrowly European and his experience has been exclusively in the laboratory and lecture hall, his work therefore needing careful and tactful guidance if in fact his role is to be carried out in the sense the Commission had in mind."[33]

When considering whether to reign in Högl's mini-Europe meetings, he admitted: "in any case it would be extremely difficult to oppose it at this stage since this might upset the Europe/Worldwide balance we hope the Commission achieved…we have however a coordinator and we must in all fairness give him a chance to coordinate, even if this work might be done more effectively in a different way."[34]

Högl would also write to the Director-General of the WHO, strongly urging that there be greater involvement from that agency out of fear that the FAO, with its agricultural interests, made it "almost impossible to prevent that the economic aspects appear rather strongly."[35] He explained that in their European Codex, "in the majority of the States, these laws are enacted by the ministry responsible for the public health. The economic aspects would have as much as possible remain[ed] apart from the discussion."[36]

Townshend sought Harvey's counsel once more and in a letter attached

> a copy of a curious "Circular No. 1" sent out by a body calling itself the "Public Health Workshop, Foodstuffs Legislation Section" in Vienna, dated December 20, 1963. The list of addresses to which it was sent by the originators is also attached. You will see that it is sent out "at the suggestion" of Doctor Hans Frenzel and reads very like a carefully edited "anti-Codex" manifesto (although the dateline is indeed Vienna and not Avignon!). There is clearly much more in this than can be read between its closely typed lines. It is, however, quite new to me and I have not yet been able to find out what exactly lies behind it. If I may mix some good metaphors, it looks a bit like a kite sent up by…leaving what they hope is a sinking ship. If it is so they will surely be steam-rollered! I want to try to analyse thoroughly the various reasons behind European separatism in Codex matters; there are at least a half dozen, some of which are not emotional and could I believe and must indeed be cured.[37]

Regional to Worldwide: "This sounds like a holy mess"

It is unclear whether these reasons were "cured" as Townshend had wished. After the first session of the Commission was held in Rome in June of 1963, everything seemed set, although debate lingered over the powers of the regional group. The European Council needed to approve the terms of its entry into the international Codex, to be decided in May of 1964 at a meeting in Berne. But months after the first session of the Codex, and just months prior to the Berne meeting,

a "holy mess" erupted over the report from the Rome 1963 session. The English version did not correspond with the French translation, or, as the European Council charged, changes were made to the English translation after the meeting, which would ultimately strip the European region of its desired autonomy. The blame for the changes (according to the Swiss and French representatives) was attributed to Townshend, the officer-in-charge, who had just coincidentally resigned his post, and they felt that it was a deliberate attempt to embarrass the Europeans at the upcoming meeting in Berne. Högl indicated that the European Codex Council might have to back out of the deal altogether: "We must at all costs find a compromise solution that will win the approval of both sides. To side with one group of countries would mean losing the collaboration of the other group and make the 'European' ideal an illusion."[38]

Perhaps in less subtle terms, a French representative wrote to the FAO Director-General B.R. Sen, reminding him that: "These changes refer to an article which was the object of very heated debate…I hope, Mr. Director-General, that without waiting to receive a more formal protest from the French Government, you will kindly give this matter the attention it deserves."[39]

This sparked a flurry of correspondence and memos within and between the FAO and WHO, with one handwritten note scrawled in the margins, "this sounds like a holy mess."[40] In actual fact, the holy mess appears to have been a "set-up" by Högl; the mistake was caught by Townshend a week after the Rome meeting and he communicated it to Högl, who never raised the issue to the European Council, presumably knowing that the French version would be to the liking of some of the members (e.g., France), and the English version might be labelled as the "mistake." As an FAO Liaison Director wrote: "It is obvious [Högl] is attempting to exploit the unfortunate oversight in translation to get over an important decision of the Codex Commission which is not to the liking of a few members of his Council (France, Germany, Switzerland and Austria)."[41]

At the May Meeting in Berne, the European Council decided the terms of its entry into the worldwide Codex, and insisted on its autonomy in establishing its own standards for problems of a particularly European nature. After Berne, and on the advice of the its Executive Board and Working Party on Rules of Procedure, the international Codex decided to change the Rules of Procedure for Advisory Bodies,

admitting that it was in error and that it "seemed advisable…not to define the functions of these bodies too restrictively…since it might become necessary to assign other functions to them in the future which might be markedly different from those contemplated at present".[42]

They added that the term "Co-ordinating Committees" would more properly express the proposed functions of these subsidiary bodies rather than the expression "Advisory Groups." Högl was immediately appointed the Chair of this Coordinating Committee for Europe; however, he stepped down after less than one year in the post. Dr. Frenzel came out of retirement to chair the committee in 1965, but passed away the following year, and another Austrian representative, his loyal paladin, Dr. Richard Wildner took his place. On their own terms, they had finally established the connection to the large and mighty organization.

These tensions between the European and the international Codex in the early years show how many complex political, economic, and health ("scientific") concerns were at play in crafting the Codex standards. Concerns over health were certainly voiced, although the larger concern appears to have been the question of autonomy and control in the decision making for the European region.

Another key area affecting consumer health, particularly in the eyes of the WHO, and source of tension in the early years of the Codex was that of food hygiene standards. The Assistant Director-General of the WHO described how "the importance of food hygiene in relation to human health is obvious, and WHO and FAO have collaborated on this work for many years."[43] The collaboration was uneasy at times, however, mainly due to the WHO's reliance upon an external group of microbiologists and biometricians, who espoused a statistical approach to food safety. As alluded to earlier, this case study not only further illustrates interagency tension over constructing standards, but also points to the larger issue of what counts as "correct science" when standards are being crafted by large international organizations.

Tensions over "Scientific Correctness": The Case of Food Hygiene

In May of 1969, an international team of microbiologists met in Dubrovnik to discuss microbiological specifications for food and, for the first time, their invited guests included biometricians. The

International Committee for the Microbial Specifications for Food (ICMSF) had convened before, but on this occasion they required the assistance of experts trained in statistical methods. The Chairman, Fred Thatcher, announced to his colleagues how their "biometrican friends" deserved thanks for developing and explaining statistical sampling plans to the microbiologists.[44] With a biometrical approach to microbes, the ICMSF began to exert great influence on the science of food hygiene, and received the attention and financial support of the WHO, which had already been investing in the Codex standards program. The Codex Food Hygiene Committee approached food hygiene standards from a more qualitative perspective, providing descriptions in lieu of numerical values such as microbial counts for food products. With the emergence of the ICMSF, however, and the advent of new statistical sampling techniques for microogranisms like salmonella, the Codex increasingly—at first reluctantly—had to incorporate the expertise of biometricians into its work. The WHO worked much more closely with the international microbiologists than the FAO and insisted, along with the ICMSF, on the input of biometricians and highlighted their role in changing the shape of food safety.

This reliance upon another discipline of experts for microbiological food issues marked an apparent shift in approach to food hygiene standards, one which called for a greater emphasis on quantification and statistical methods in order to be deemed "scientifically correct." Appealing to statistical knowledge was not novel; it has been well documented that statistical styles of reasoning had been occurring within pockets of scientific and medical research communities starting in the early nineteenth century and continuing throughout.[45] Initially, however, the Codex standards focused on whether microbes were present in a food product without any concern over methods of enumeration or statistical significance.[46] The statistical approach to food safety involves, for example, establishing numerical limits to describe "safe" amount of microbes or maximum tolerance levels for pesticides or additives or other contaminants in a given sample of food. As I argue, this was part of a broader post-war shift in approach to food standards, and the Food Hygiene Committee serves as one case study of how tensions emerged over the meaning of "correct science" when constructing standards.

Codex Food Hygiene Standards

The first meeting of the Codex Expert Committee on Food Hygiene was held in Washington, DC, in 1964 at the Pan American Health Organization/WHO Building, where it remained for the next several years under the chairmanship of the Government of the United States. The Chairman was Lavega Robert (Bob) Shelton, a bacteriologist working in the Division of Microbiology at the United States Food and Drug Administration (FDA). After completing his master's, he spent a year working in the laboratory in his home state with the Missouri Board of Health and then went on to join the FDA as a Seafood Supervisor.

Other members of this committee included delegates from Australia, Canada, Cuba, Denmark, Ireland, Italy, the Netherlands, Poland, Portugal, Sweden, Switzerland, Turkey, and the UK. The members were scientists ranging from professions such as veterinary inspectors, medical inspectors, bacteriologists, directors of laboratories for food zoonoses, and professors of food hygiene and environmental sanitation, and many were also representatives from member states' ministries of hygiene, health, or agriculture—essentially a range of experts in the field of food hygiene, all charged with the task of drafting standards.

Its secretariat was particularly active in arranging tours of processing plants and meat packaging facilities, along with distributing reports, information, and the draft hygiene standards to countries, industries, or members who requested them. The focus of the hygiene standards was not solely on the final *product*; rather, they described the entire *process* of food production, including methods, practices, and behaviours. These standards, once fully elaborated, were thick documents giving the definition, storage requirements, plant and operational facilities, and end product description.

For example, the final product (i.e., ready for human consumption) description would be qualitative. Taking a canned fruit and vegetable example, we learn that

> To the extent possible in good manufacturing practice the products should be free from objectionable matter including insects and insect parts, insect webbing, soil, sand, or stone fragments, faecal matter of any kind, human or animal hair, and free from fungal filaments (mold) to an extent indicative of decayed ingredients; The product should be free from any pathogen infectious to man and from any toxic

substance originating from bacteria or fungi; Products with an equilibrium pH above 4.5 should have received a processing treatment sufficient to destroy all spores of *Clostridium botulinium*, unless growth of surviving spores would be permanently prevented by product characteristics other than pH.[47]

The ICMSF and the Biometrical Approach to Hygiene Standards

Meanwhile in Geneva, as mentioned at the beginning of this section, the WHO had decided to provide funding to the ICMSF,[48] which had emerged in 1962—the same year as the Codex Alimentarius—and was a standing committee of the International Association of Microbiological Societies.[49] This group was formerly known as the International Society for Microbiology, founded in 1927, and by the late 1960s became part of the International Union of Biological Sciences (IUBS).

The Co-Chairs of the ICMSF were Canadian microbiologists, Drs. Fred Thatcher and David Clarke, based out of the Department of Health and Welfare's Food and Drug Directorate in Ottawa. Other members were from the Netherlands, such as Dr. David Mossel, as well as Dr. Betty Hobbs, a British bacteriologist and an internationally recognized authority on food poisoning and food hygiene. Dr. Hobbs was one of the first members of the Public Health Laboratory Service in London and became director of its Food Hygiene Laboratory. She produced an influential series of books aimed not only at microbiologists but also at those working in the food service industry and for those preparing food in the home. These included *Food Poisoning and Food Hygiene* (1953) and *Hygienic Food Handling* (1962).

Collectively, the ICMSF produced a monograph and series of volumes which became very influential worldwide and adopted in universities for teaching purposes.[50] The first edition of *Microorganisms in Foods: Their Significance and Methods of Enumeration* (1968) identified the main culprits of food poisoning (salmonella, Escherichia coli, clostridium, staphylococcus) and the methods and reagents needed to detect, isolate, and analyze them. The second edition, however, was to be revised based on decisions made in Dubrovnik.

The meeting in Dubrovnik was attended by a WHO representative, the food hygienist Dr. Znedek Matyas, and the agenda included a day-and-a-half discussion led by the biometricians.[51] The three biometricians

present were Dr. D.F. Bray from Canada, Irving W. Burr from the United States, and Dr. E.F. Drion of the Netherlands.[52] Dr. Bray, an employee of the Department of Health and Welfare, was a member of the Statistical Society of Canada and had a background in poultry science.[53] Irving W. Burr was a Purdue University faculty member and leading expert in quality control and industrial statistics. Dr. Drion had collaborated previously with Dr. Mossel on the importance of establishing microbial limits in food.

The sampling plans themselves were very detailed and relied upon concepts of sample size, probability, population estimates, the use of lots and batches and coding systems, and the value "c," the maximum allowable number of microbes found in a product. These numerical values were specific to the microbe and specific to the food product (e.g., for salmonella, "c" had to be relatively lower, if not "0," depending on the food it was found in) (ICMSF 1974).

In preparation for this meeting, the ICMSF decided that "assuming some agreement on 'standards' the committee will again meet with the biometricians. If sampling plans have been recommended, these will be referred to the biometricians for comment; if not, then appropriate data and proposals for 'standards' will be offered to the biometricians and their advice sought."[54] And, "In due course, the overall objective is to publish a sequel to our book…under a title such as 'Sampling Plans for the Microbiological Analysis of Foods.' Our biometrician colleagues have kindly consented to assist in such a venture."[55]

The increased reliance upon biometrical approaches to microbial food issues clashed with the standards set by the Codex Food Hygiene Committee, mainly over the question of what was "scientifically correct." Despite the Codex being a joint FAO/WHO initiative, the WHO went so far as to create its own expert food hygiene group independent of the Codex, and worked more closely with the ICMSF, and thus with the biometricians, which brought an onslaught of confusion. As one member of the Codex wrote: "The argument over the functions of the Codex Committee on Food Hygiene and the [WHO] Expert Committee on Food Hygiene respectively seems to be going round in circles." The WHO had greater interest in pathogens and in methods of sampling, whereas the Codex seemed to focus more broadly on the products themselves and approached their standards with a broader meaning of hygiene, not limited strictly to microbes nor relying on numerical definitions.

"Not Scientifically Correct"

For example, one sticking point between the two food hygiene approaches was with regard to standards defining the microbial counts in the final product, items which would go on the market ready for consumption. The crux of the issue was a phrase that appeared in most of the Codex Group's standards for the final product: "that products should be free from any pathogen infectious to man and from any toxic substance originating from micro-organisms."[57] The ICMSF strongly urged that simply describing the final product as "free from pathogens" was not enough; there had to be more rigorous numerical standards and methods for sampling. The WHO's position was that simple descriptive qualitative hygiene standards were "not scientifically correct." In a report describing the role of the Food Hygiene Laboratory, the WHO explained:

> The food hygiene laboratory can only give the microbial status of the food in relation to the sensitivity of the test used and with the consideration that tests can be used only for a limited range of pathogenic organisms. Further, the finding of no pathogens does not mean that such pathogens are absent from the food. They may not be found by the methods used. This caution is especially important since any other interpretation of microbiological data would lead to the impression that foods certified by a laboratory as pathogen-free did not contain pathogens, or might be so safe that subsequent care in handling could be disregarded. In order to maintain the scientific integrity of the food hygiene laboratory, the Committee recommends that no microbiological results be issued without a qualifying statement which indicates the exact number of samples examined in relation to the total lot in question, the quantity of sample, and the methods used. (WHO 1967)

Dr. Znedek Matyas admitted that the sampling plans were "a very controversial issue" at the time of development. And, in terms of uniting the two food hygiene groups, Matyas reported that when he "spoke privately at the [Codex] sessions in Washington with some of the recognized microbiologists present, they were against such closer relationships, saying that there are many other international committees or commissions working in the same field as ICMSF and which have also excellent results."[58]

The Chairman of the Codex Food Hygiene group, Bob Shelton, wrote to his liaison officer in Rome:

> The question of Codex/ICMSF relations has always been intriguing to me. As a food microbiologist by training and experience, I am very sympathetic to the aims of the ICMSF and appreciative of the problems with which the group is wrestling. The method testing program and the studies on statistical design for sampling are splendid—no other group is in a position to take on such work. The individual members are experts in the field and I have high regard for each of them. Consequently, in my opinion, it is entirely logical to depend heavily on their expertise in food microbiology. For some reason which I never been able to fathom, there seems to be a deep antipathy toward the ICMSF on the part of Australia (Mr. Smith) and the U.K. (Dr. Ross). It is my understanding that these gentlemen were the principal antagonists to the liaison proposed at the second session of the Food Hygiene Committee. Believe we have our work cut out for us to achieve a good working relationship![59]

Indeed, as the Australian Mr. Smith aptly complained, "I find it somewhat surprising that one of the Governing Bodies should be engaged in the elaboration of microbiological specifications for food with I.C.M.S.F. while the Codex subsidiary whose terms of reference encompass all aspects of food hygiene, is uninformed of the nature of that work."[60]

Shelton tried his best, but as one FDA observer to the Codex food hygiene meetings noted: "I have dropped by the Hygiene Committee meeting this week. These hygienists are difficult to please. Both McNally and Shelton have to work hard in order to move them past any controversial point."[61]

By 1973, methods for salmonella isolation for eggs and egg products were of particular concern and a growing problem, and at this point another international group, the International Organization for Standardization (ISO), wrote to the Codex to notify them of its interest in developing salmonella standards. This made for a total of three groups of international experts attempting to draft microbiological food hygiene standards. Within days of this news, Bob Shelton submitted his resignation as the Codex hygiene chairman, citing how it was evidently a time for new ideas in food hygiene: "While I am a staunch advocate that continuity of membership is desirable for the work of Codex Committees, I believe length of service by the Chairman can be carried to extremes. There is a risk of becoming fixed in ideas and this could be damaging to the work of the Committee."[62]

His replacement, J.C. Olson, worked with the US FDA and was also a member of the ICMSF and, with this connection, he was well prepared to tackle the question of salmonella standards for egg and egg products. At this stage, all groups were prepared to embark upon interlaboratory testing of methods of detection of salmonellae in egg and egg products and attempt to settle on salmonella testing standards. Olson wrote to a colleague about the ISO, Codex, and ICMSF attempts to collaborate: "We must recognize that microbiological methodology is in a dynamic state, and probably always will be. Unanimity of agreement of what is the best method is perhaps too much to ask. Undoubtedly some compromise will be necessary to reach a position that the Committee can be comfortable with. Personally, I feel that such can take place without consequence of adverse practical significance."[63]

Indeed, salmonella became a sort of microbial "poster child" for studies in standardization; articles and studies on how to control and test for this pathogen appeared in many public health and medical journals during this time.[64] Approaches and methods may have differed among laboratories, and even among farms, but the message became clear: standardization using statistics was the "best" route for addressing the problem.

Conclusion

The historical evidence of these early tensions serves as an important basis for understanding the trajectory and evolution of the Codex and the development of international food safety standards since the post-war period. In fact, the constructed nature of the standards and the tensions of the early years seem to have been echoed in more recent debates. The Codex was thrust into the limelight in the mid-1990s, when high-profile health concerns, import bans, and court cases began to surface as a result of the use of growth promoting hormones in beef, as well as the rise of production aids such as the milk hormones Bovine Somato-tropin (BST). Debates over maximum residue limits (MRLs) for these substances often pitted countries like the United States against the European Union, and forced the Codex to consider whether to base its decisions strictly on sound science or consider "other limiting factors" such as consumer concerns, animal welfare, fraudulent or unfair trading practices, labelling, and other ethical and cultural considerations (Jukes 2000). The Sanitary and Phytosanitary (SPS) agreement came into force in January 1995, and

emphasized how any disputes between member states over the health risks of a particular commodity must be resolved through the use of sound scientific principles. And since many members of the WTO were also members of the Codex, it naturally followed that the Codex was to become the source of these scientific principles for food standards (ibid.). But, as I have documented, long before recent controversies over growth hormones in milk and meat which plagued the Codex in the 1990s, there existed disagreements and tensions in the development of standards and the structures that would attempt to govern them.

Acknowledgements

This research was funded by an Ontario Graduate Scholarship, for which I am appreciative. I would like to thank the Lupina Foundation and the Comparative Program on Health and Society at the University of Toronto's Munk Centre for International Studies for the opportunity to present this work, as well as for a fellowship in support of this research. I would especially like to thank Harriet Friedmann for her comments on a draft of this paper, and to Pauline M.H. Mazumdar for her thoughts and support in conducting the archival research for this project.

Notes

1. Davies, J.H.V. 23 April 1969. Letter to G.O. Kermode. CX2/20. Box 12 C x 6. FAO Archives, Rome.

2. See Stanziani (2006); Davidson (1949); Hilts (2004).

3. Nehemiah, J.V.A. 18 March 1963. Origin—Joint FAO/WHO Program on Food Standards (Codex Alimentarius Program.) Note to B.R. Sen. SP10/1 Box 12 C x 74. FAO Archives, Rome.

4. Townshend, F.H. 6 November 1962. Letter to Dr. C. Fedele. SP10/2 Box 12C x 74. FAO Archives, Rome.

5. FAO/WHO. 1963. Report of the First Session of the Joint FAO/WHO Codex Alimentarius Commission. Rome. ALINORM 63/12. FAO Archives, Rome.

6. Ibid.

7. FAO/WHO. 1965. Joint FAO/WHO Program on Food Standards, Codex Alimentarius Commission. Appendix III: General Principles of the Codex. Report of the First Meeting of the Codex Committee on General Principles. Paris. ALINORM 65/9. FAO Archives, Rome.

8. Terver, P. 31 January 1964. Letter to Dr. P.M. Dorolle. SP10/1. Box 12 C x 74. FAO Archives, Rome.

9. Candau, M.G. 5 August 1966. Letter to B.R. Sen. BU2/1. Box 12 C x 1. FAO Archives, Rome.

10. FAO. 1966. Draft Note for the Eleventh Session of the Program Committee. CX2/20. Box 12 C x 6. FAO Archives, Rome.

11. Davies, J.H.V. 23 April 1969. Letter to G.O. Kermode. CX2/20. Box 12 C x 6. FAO Archives, Rome.

12. Kermode, G. 18 November 1966. Letter to Mr. P. Terver. SP 10/1. Box 12C x 74. FAO Archives, Rome.

13. Verhoestraete, L. 6 February 1964. Letter to Professor Högl. BU 2/1. Box 12 C x 1. FAO Archives, Rome.

14. Frenzel, H. 1962. Speech to Codex Europeaus. [Warum wollen wir einen Anschulss an eine grosse mächtige Organisation?] SP10/1 Box 12 C x 74. FAO Archives, Rome.

15. World Health Organization (WHO). 1962. Draft Document of WHO Exec Board Jan 62 Meeting of "Joint FAO/WHO Programme on Food Standards (Codex Alimentarius)." FAO Archives, Rome.

16. Frenzel, H. 1962. Speech to Codex Europeaus. [Warum wollen wir einen Anschulss an eine grosse mächtige Organisation?] SP10/1 Box 12 C x 74. FAO Archives, Rome.

17. FAO/WHO. 1962. Report of the Joint FAO/WHO Conference on Food Standards. Geneva. ALINORM 62/8, p. 3. FAO Archives, Rome.

18. Cépède, 1964. Letter to Nehmiah. SP 10/1 Box 12 C x 74. FAO Archives, Rome.

19. Frenzel, Hans, 1962, Speech to Codex Europeaus. SP10/1 Box 12 C x 74. FAO Archives, Rome.

20. Ibid.

21. Cépède, 1964. Letter to Nehmiah. SP 10/1 Box 12 C x 74. FAO Archives, Rome.

22. FAO/WHO. 1962. Report of the Joint FAO/WHO Conference on Food Standards. Geneva. ALINORM 62/8, p. 9. FAO Archives, Rome.

23. Cépède, 1964. Letter to Nehmiah. SP 10/1 Box 12 C x 74 FAO Archives, Rome.

24. Ibid.

25. Ibid.

26. Ibid.

27. Townshend, F.H. 4 April 1963. Letter to Claus Agthe. SP 10/2 Box 12 C x 74. FAO Archives, Rome.

28. Yates, P.L. 7 September 1962. Letter to Mr. Oris V. Wells. SP 10/1 Box 12C x 74. FAO Archives, Rome.

29. Orbaneja. 10 September 1962. Letter to P. Yates. SP 10/1 Box 12C x 74. FAO Archives, Rome.

30. Ibid.

31. Nehemiah. 17 March 1963. Letter to P. Terver, SP 10/2 Box 12 C x 74. Archives, Rome.

32. Ibid.

33. Townshend, F.H. 10 September 1963. Letter to Dr. John C. Harvey. SP 10/1 Box 12C x 74. FAO Archives, Rome.

34. Ibid.

35. Högl, O. 2 April 1964. Letter to B.R. Sen. SP10/2 Box 12 C x 74 FAO Archives, Rome.

36. Högl, O. 16 January 1964. Letter to M.G. Candau. SP 10/2 Box 12 C x 74. FAO Archives, Rome.

37. Townshend, F.H. 23 January 1964. Letter to Dr. John C. Harvey. SP 10/3, 10/8, 10/4 (Austria) SP 10/2/WHO. Box 12 C x 74. FAO Archives, Rome.

38. Högl, O. 2 April 1964. Letter to B.R. Sen. SP10/2 Box 12 C x 74. FAO Archives, Rome.

39. Cépède, M. 3 April 1964. Letter to B.R. Sen. SP10/2 Box 12 C x 74. FAO Archives, Rome.

40. Ibid.

41. Nehemiah, J. 9 April 1964. Letter to Mr. O.V. Wells. SP 10/2 Box 12 C x 74. FAO Archives, Rome.

42. FAO/WHO. 1964. Report of the Working Party on the Rules of Procedure and Related Matters in Report of the Second Session of the Joint FAO/WHO Codex Alimentarius Commission. FAO Archives, Rome.

43. Dorolle, P. 1964. Opening Speech of the Assistant Director-General at the Second Session of the Joint FAO/WHO Codex Alimentarius Commission. Geneva. CX2/20. Box 12 C x 6. FAO Archives, Rome.

44. Thatcher, F.S. 13 January 1969. Letter to the International Committee on Microbiological Specifications for Foods. CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

45. See, for example, Hacking (1990), who describes the end of determinism in the nineteenth century and the onslaught what he calls an "avalanche of numbers," and Porter (1986). See, also, Matthews (1995, p. 100), who describes how bacteriologists and biometricians have met before in the early twentieth century. Matthews describes how debates over who should be "the final arbiter of medical knowledge" (clinicians, biometricians, or bacteriologists) influenced the disputes between the British biometrical school (Karl Pearson) and the bacteriologist Sir Almoth Wright over the effectiveness of Wright's antityphoid inoculation.

46. See Ilcan and Phillips (2003). Phillips and Ilcan approach this topic by employing sociologist Nikolas Rose's concept of "technologies of government" to describe the role of FAO and WHO in "making food count." Technologies of government are "an assemblage of forms of practical knowledge" and include modes of perception, vocabularies, types of authority, or practices of calculation, employed not only to govern conduct, but also as they suggest, to make objects, in this case, food, more "knowable."

47. Food and Agriculture Organization of the United Nations, Report of the Fourth Session of the Codex Committee on Food Hygiene, Washington D.C., USA, 1967. Appendix III.

48. Matyas, Z. 28 January 1969. letter to F.S. Thatcher , CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

49. Sand, Dr. P. 28 January 1972. Letter to W.L. de Haas, CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

50. Thatcher, F.S. 13 January 1969. Letter to the International Committee on Microbiological Specifications for Foods. CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

51. Matyas, Z. 28 January 1969. letter to F.S. Thatcher, CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

52. Thatcher, F.S. 13 January 1969. Letter to the International Committee on Microbiological Specifications for Foods. CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

53. See Bray (1999) and http://www.landfood.ubc.ca/alumni_history/poultry_science.htm (accessed 26 May 2008).

54. Thatcher, F.S. 13 January 1969. Letter to the International Committee on Microbiological Specifications for Foods. CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

55. Ibid.

56. Götzsche, N.O. 10 April 1968. Letter to Dr. F. Winkelmann. CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

57. Food and Agriculture Organization of the United Nations. Report of the Fifth Session of the Codex Committee on Food Hygiene, Washington, D.C., p. 4.

58. Matyas, Z. 27 March 1972. Letter to Graham Kermode, CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

59. Shelton, R. 9 February 1972. Letter to de Haas. CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

60. Smith, I.H. 6 January 1972. Letter to Graham Kermode. CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

61. Kermode, G. 28 July 1970. Letter to H. McNally. CX4/20 Vol. I : Box 12 C x 12. FAO Archives, Rome.

62. Shelton, B. 6 September 1973. Letter to William de Haas. CX 4/20 Vol III: Box 12C x 12. FAO Archives, Rome.

63. Olson. 12 November 1973. Letter to K. Buchli. CX 4/20 Vol III: Box 12C x 12. FAO Archives, Rome.

64. See, for example, Edel and Kampelmacher (1968, 1969); Roberts et al. (1975); Williams and Newell (1970).

References

Archival references were obtained from the Food and Agriculture Organization (FAO) Archives in Rome.

Beacham, L.M. 1973. Ten Years of Codex Alimentarius—A Progress Report. *Foreign Agriculture*, May 28, 10–11.

Bray, D.F. 1999. A History of the Statistical Society of Canada: The Formative Years. *Statistical Science* 14 (1): 123–24.

Bud, Robert. 2007. *Penicillin: Triumph and Tragedy.* Oxford: Oxford University Press.

Bunch, Matthew. 2007. Canada and the Freedom from Hunger Campaign. PhD diss. University of Waterloo.

Candau, M.G. 5 August 1966. Letter to B.R. Sen. BU2/1. Box 12 C x 1. FAO Archives, Rome.

Cépède, M. 3 April 1964. Letter to B.R. Sen. SP10/2 Box 12 C x 74. FAO Archives, Rome.

———. 1964. Letter to Nehmiah. SP 10/1 Box 12 C x 74. FAO Archives, Rome.

Charnley, Berris, Arguing Over Adulteration: The Success of the Analytical Sanitary Commission. *Endeavour* 32 (4): 129–33.

Davidson, A. 1949. *The Genesis and Growth of Food and Drug Administration in Canada.* Ottawa: Food and Drug Directorate.

Davies, J.H.V. 23 April 1969. Letter to G.O. Kermode. CX2/20. Box 12 C x 6. FAO Archives, Rome.

Dorolle, P. 1964. Opening Speech of the Assistant Director-General at the Second Session of the Joint FAO/WHO Codex Alimentarius Commission. Geneva. CX2/20. Box 12 C x 6. FAO Archives, Rome.

Edel, W., and E.H. Kampelmacher. 1968. Comparative Studies on Salmonella-isolation in Eight European Laboratories. *Bulletin of the World Health Organization* 39 (3): 487–91.

———. 1969. Salmonella Isolation in Nine European Laboratories Using a Standardized Technnique. *Bulletin of the World Health Organization* 41 (2): 297–308.

FAO. 1966. Draft Note for the Eleventh Session of the Program Committee. CX2/20. Box 12 C x 6. FAO Archives, Rome.

Food and Agriculture Organization of the United Nations. 1967. Report of the Fourth Session of the Codex Committee on Food Hygiene, Washington D.C., USA, Appendix III.

———. 1967. Report of the Fifth Session of the Codex Committee on Food Hygiene, Washington, D.C.

FAO/WHO. 1962. Report of the Joint FAO/WHO Conference on Food Standards. Geneva. ALINORM 62/8. FAO Archives, Rome.

———. 1963. Report of the First Session of the Joint FAO/WHO Codex Alimentarius Commission. Rome. ALINORM 63/12. FAO Archives, Rome.

———. 1964. Report of the Working Party on the Rules of Procedure and Related Matters in Report of the Second Session of the Joint FAO/WHO Codex Alimentarius Commission. FAO Archives, Rome.

———. 1965. Joint FAO/WHO Program on Food Standards, Codex Alimentarius Commission. Appendix III: General Principles of the Codex. Report of the First Meeting of the Codex Committee on General Principles. Paris. ALINORM 65/9. FAO Archives, Rome.

Frenzel, H. 1962. Speech to Codex Europaeus. [Warum wollen wir einen Anschulss an eine grosse mächtige Organisation?] SP10/1 Box 12 C x 74. FAO Archives, Rome.

Friedmann, H. 1993. The Political Economy of Food: A Global Crisis. *New Left Review,* no. 197: 29–57.

Götzsche, N.O. 10 April 1968. Letter to Dr. F. Winkelmann. CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

Guillem Llobat, Ximo. 2008. El control de la qualitat dels aliments. El cas Valencià en el context internacional, (1878–1936). PhD diss., Universitat de Valencià.

Hacking, Ian. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.

Hardy, A. 1999. Food, Hygiene and the Laboratory. A Short History of Food Poisoning in Britain circa 1850–1950. *The Society for the Social History of Medicine* 12: 293–311.

Hierholzer, Vera. 2007. Searching for the Best Standard: Different Strategies of Food Regulation during German Industrialization, *Food & History* 5 (2): 295–318.

Hilts, P.J. 2004. *Protecting America's Health: The FDA, Business, and One Hundred Years of Regulation.* Chapel Hill: University of North Carolina Press.

Högl, O. 16 January 1964. Letter to M.G. Candau. SP 10/2 Box 12 C x 74. FAO Archives, Rome.

———. 2 April 1964. Letter to B.R. Sen. SP10/2 Box 12 C x 74. FAO Archives, Rome.

Ilcan, S., and L. Phillips. 2003. Making Food Count: Expert Knowledge and Global Technologies of Government. *The Canadian Review of Sociology and Anthropology* 40 (4): 444–62.

International Commission on Microbiological Specifications for Foods (ICMSF). 1974. *Microorganisms in Foods: Sampling for Microbial Analysis: Principles and Specific Applications*, vol. 2. Toronto: University of Toronto Press.

Jukes, D. 2000. The Role of Science in International Food Standards. *Food Control* 11: 181–94.

Kermode, G. 18 November 1966. Letter to Mr. P. Terver. SP 10/1. Box 12C x 74. FAO Archives, Rome.

———. 28 July 1970. Letter to H. McNally. CX4/20 Vol. I : Box 12 C x 12. FAO Archives, Rome.

Matthews, Rosser J. 1995. *Quantification and the Quest for Medical Certainty*. Princeton: Princeton University Press.

Matyas, Z. 28 January 1969. letter to F.S. Thatcher, CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

———. 27 March 1972. Letter to Graham Kermode, CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

Mazumdar, P.M.H. 2003. "In the Silence of the Laboratory": The League of Nations Standardizes Syphilis Tests. *Social History of Medicine* 16: 437–59.

Murnaghan, J.H., and P. Talalay. 1992. H.H. Dale's Account of the Standardization of Insulin. *Bulletin of the History of Medicine* 66: 440–50.

Nehemiah, J.V.A. 17 March 1963. Letter to P. Terver, SP 10/2 Box 12 C x 74. FAO Archives, Rome.

———. 18 March 1963. Origin — Joint FAO/WHO Program on Food Standards (Codex Alimentarius Program.) Note to B.R. Sen. SP10/1 Box 12 C x 74. FAO Archives, Rome.

———. 9 April 1964. Letter to Mr. O.V. Wells. SP 10/2 Box 12 C x 74. FAO Archives, Rome.

Olson. 12 November 1973. Letter to K. Buchli. CX 4/20 Vol III: Box 12C x 12. FAO Archives, Rome.

Orbaneja. 10 September 1962. Letter to P. Yates. SP 10/1 Box 12C x 74. FAO Archives, Rome.

Pieters, T. 1998. Managing Differences in Biomedical Research: The Case of Standardizing Interferons. *Studies in History and Philosophy of Biological and Biomedical Sciences* 29: 31–79.

Porter, Ted. 1986. *The Rise of Statistical Thinking, 1820–1900*. Princeton: Princeton University Press.

Roberts, Diane, et al. 1975. The Isolation of Salmonellas from British Pork Sausages and Sausage Meat. *The Journal of Hygiene* 75 (2): 173–84.

Ruxin, J.N. 2000. The United Nations Protein Advisory Group. In *Food, Science, Policy, and Regulation in the Twentieth Century: International and Comparative Perspectives*, edited by J. Phillips and D.F. Smith, 151–66. London: Routledge.

Sand, Dr. P. 28 January 1972. Letter to W.L. de Haas, CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

Shelton, B. 6 September 1973. Letter to William de Haas. CX 4/20 Vol III: Box 12C x 12. FAO Archives, Rome.

Shelton, R. 9 February 1972. Letter to de Haas. CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

Siddiqi, J. 1995. *World Health and World Politics. The World Health Organization and the UN System*. London: Hurst & Company.

Smith, I.H. 6 January 1972. Letter to Graham Kermode. CX 4/20 Vol II: Box 12C x 12. FAO Archives, Rome.

Stanziani, A. 2006. Quality Rules: What For? Consumers vs. Lobbies Protection, The Case of France, 19th–early 20th C. Paper presented at the XIV International Economic History Congress, Helsinki.

Staples, A.L.S. 2006. *The Birth of Development: How the World Bank, Food and Agriculture Organization, and World Health Organization Changed the World, 1945–1965*. Kent, Ohio: Kent State University Press.

Terver, P. 31 January 1964. Letter to Dr. P.M. Dorolle. SP10/1. Box 12 C x 74. FAO Archives, Rome.

Thatcher, F.S. 13 January 1969. Letter to the International Committee on Microbiological Specifications for Foods. CX 4/20 Vol I: Box 12C x 12. FAO Archives, Rome.

Townshend, F.H. 6 November 1962. Letter to Dr. C. Fedele. SP10/2 Box 12C x 74. FAO Archives, Rome.

———. 4 April 1963. Letter to Claus Agthe. SP 10/2 Box 12 C x 74. FAO Archives, Rome.

———. 10 September 1963. Letter to Dr. John C. Harvey. SP 10/1 Box 12C x 74. FAO Archives, Rome.

———. 23 January 1964. Letter to Dr. John C. Harvey. SP 10/3, 10/8, 10/4 (Austria) SP 10/2/WHO. Box 12 C x 74. FAO Archives, Rome.

Verhoestraete, L. 6 February 1964. Letter to Professor Högl. BU 2/1. Box 12 C x 1. FAO Archives, Rome.

Williams, L., and K.W. Newell. 1970. "Salmonella Excretion in Joy-riding Pigs," *American Journal of Public Health* 60 (5): 926–29.

World Health Organization (WHO). 1962. Draft Document of WHO Exec Board Jan 62 Meeting of "Joint FAO/WHO Programme on Food Standards (Codex Alimentarius)." FAO Archives, Rome.

———. 1967. The Role of the Food Hygiene Laboratory in Food Hygiene Programs. Report of the WHO Expert Committee on Microbiological Aspects of Food Hygiene. Geneva.

———. n.d. Global Health Histories, available at http://www.who.int/global_health_histories/en/.

WHO/FAO. 2005. *Understanding the Codex*, 3rd ed. Rome: World Health Organization, Food and Agriculture Organization of the United Nations.

Yates, P.L. 7 September 1962. Letter to Mr. Oris V. Wells. SP 10/1 Box 12C x 74. FAO Archives, Rome.

The Healing of Psychosocial Trauma in the Midst of Truth Commissions. The Case of *Gacaca* in Post-Genocide Rwanda

Regine King

Abstract

This paper discusses the challenges of healing psychosocial trauma through truth commissions in post-conflict societies and, in particular, the psychosocial role of *gacaca* in post-genocide Rwanda. *Gacaca* was considered a homegrown model with the dual role to address the past wrongs and heal psychosocial trauma. However, like many other truth commissions it became more retributive and less reconciliatory and, as a result, failed to address psychosocial trauma and repair social relations. In this paper, I examine the psychosocial role of truth commissions with an emphasis on *gacaca* and in reference to the Truth and Reconciliation Commission (TRC) from South Africa, a model that has been highly regarded as exemplary in regard to other forms of truth commissions. The paper concludes with some practical suggestions to fill the conceptual and implementation gaps observed in different forms of truth commissions.

Introduction

The objective of this paper is to discuss the challenges truth commissions face in their dual retributive and restorative role, and suggest alternatives of healing psychosocial trauma in post-conflict situations and post-genocide Rwanda in particular. In 2002, Rwanda introduced *gacaca* as a form of truth commission to address legal and psychosocial consequences of the 1994 genocide in the country. *Gacaca*, which literally means "grass," was utilized as a traditional Rwandan justice mechanism for dispute resolution to address issues related to property matters, inheritance, and family law (Schabas 2005, p. 891). In the aftermath of the Rwandan genocide, this

traditional justice model was transformed into a form of truth commission with a mandate to address legal and psychosocial issues pertaining to the Tutsi genocide. Although *gacaca* has been considered a homegrown model with the dual role to address the past wrongs and heal psychosocial trauma, it has been argued that in practice it became more retributive and less reconciliatory. In this paper, I examine the psychosocial role of truth commissions with an emphasis on *gacaca* and in reference to the Truth and Reconciliation Commission (TRC) from South Africa, a model that has been highly regarded as exemplary in regard to other forms of truth commissions.

I begin by providing a short overview of truth commissions and the general concerns they raise. Truth commissions are the most commonly used form of transitional justice. They were advocated by post-conflict governments and multilateral organizations at the end of the Cold War (Humphrey 2003, p. 172; Teitel 2003, p. 78). Their mandate was to combine the rule of law with psychosocial goals to break systemic cycles of violence and facilitate reconciliation (Amstutz 2005, p. 18). The combination of retributive and restorative justice has raised concerns about the potential negative impact of psychologizing intervention models that are driven by political and economic restructuring measures and through "the individualization discourse of law or the polarizing logic of trials which establish the guilty and the innocent" (Humphrey 2003, p. 171).

Over the last two decades, truth commissions have become a well-established fixture on the global terrain of human rights (Nagy 2008, p. 276), suggesting that countries should disengage from the past wrongs as a form of accountability and an acknowledgment of the pain caused by the experiences of brutality and violence (Stanley 2002, p. 1). Truth commissions advance the dimensions of truth-telling, public acknowledgement, apology, and forgiveness as the best approach toward national reconciliation in post-conflict societies. As they stand, it is unlikely the work of truth commissions can resolve legal and psychosocial issues at the same time. Collecting facts about past wrongs and providing public testimonies does not present a clear goal and investment in understanding the extent to which massive violence affects individuals and collectives, and may do more psychosocial damage than good. In the following section, I demonstrate the limitations of *gacaca* and, thereby, other models of truth commissions in the healing of psychosocial trauma.

Truth commissions are known for their focus on documenting past wrongs and their involvement with public testimonies largely about abuses committed or sponsored by the state. The collection of facts is, however, very selective and leaves many people unsatisfied. The informants and witnesses are often called to tell the parts of their stories that are of interest to the commissioners and not to those who lived the realities of violent conflicts. The psychological catharsis produced during the hearings, often considered as healing, may actually cause further psychological and social difficulties in communities in which the victim-witnesses continue to live beside the perpetrators of violence.

Truth commissions tend to function like other humanitarian interventions that import and impose models conceived without consultation with people in post-conflict societies. They are designed to collect evidence, hear public testimonies, and provide reports in a time-limited mandate. Their priority to meet the funders' needs and not the needs of affected individuals and communities is critical. There is a very narrow consideration of the magnitude of psychosocial issues at hand. Furthermore, requiring the affected individuals to publicly relive and recount their traumatic experience may be more traumatizing than healing. It has been argued that truth commissions may revive animosity between groups in the conflict through their emphasis on the categories of victims and perpetrators. To examine the above challenges, I use a particular case of *gacaca* in the context of post-genocide Rwanda.

Experience of *Gacaca* in Rwanda

The details about the Rwandan genocide go beyond the scope of this paper. Moreover, when discussing psychosocial issues related to *gacaca*, it is important to remember that, in a period of only one hundred days, an estimated eight hundred thousand Tutsis were murdered by their Hutu neighbours. People were subjected to acts of physical and emotional cruelty, rape, body mutilation, coerced participation in the murder of loved ones, and forced mass displacement from their communities. Today, more than a decade later, Rwandans continue to face multiple serious social issues including poverty, HIV/AIDS, collective trauma, injustices, and interethnic tensions. *Gacaca* was established in order to address legal community issues pertaining to the 1994 genocide that could not be tried through other legal systems. *Gacaca* was viewed as being more

restorative and a supplement of the work of the International Court Tribunal for Rwanda (ICTR) in Arusha and the national courts (Tiemessen 2004, p. 67).

During my visit in Rwanda in the summer of 2007, I decided to attend one *gacaca* session because many people I met over the course of that trip greeted me with statements like: "Did you know that this person has been released?" "Today is a *gacaca* day, you cannot do anything today, you go to the hearings or do nothing else." Once a week in every village, all work stopped for all citizens to concentrate on the local *gacaca* hearings. There was such a sense of urgency and worry that I wanted to understand what was happening. I attended one session of *gacaca* proceedings in my home village. The community members were gathered in various cells, in other words, different neighbourhood-based groups. After a long period of paper work preparation and registration to ascertain who was present at the session, the villagers were instructed to sit in a circle in front of four judges. Community judges were people who had been elected by other community members as righteous individuals who could be trusted to address crimes related to the 1994 Tutsi genocide during the *gacaca* courts. A group of suspected perpetrators stood nearby. Some of them were prisoners dressed in pink uniforms. Others were community members who had been listed as suspects but were still living in their respective communities.

The proceedings consisted of one judge reading the charges and asking the accused to make a statement. The hearings on that day were about four men who had been accused of attacking the home of a woman of about 75 years of age at the time of the hearings. Her entire family had been killed, except a son she had from a previous relationship with a Hutu man. Her property had been stolen and her house destroyed. Each member of the accused group was asked to give a personal account about his role in what happened to that woman as a form of self-defence. On the other side of the gathering, a few individual witnesses were asked to either confirm or disconfirm the alleged facts. The woman who had survived the attack was among the witnesses. Her surviving son was among the prisoners waiting for trial. The process allowed the members of the accused group to also act as witnesses. Although nobody seemed to deny what had happened to the widow and her deceased family, there were problems in explaining how it had happened and who should be held

responsible. One young man who had been seen riding the bike of the deceased husband denied stealing it but agreed to pay compensation.

The elderly woman seemed disoriented. She was speaking to herself and hit the ground with a long stick. Her responses to questions were short and limited to factual information. During this process, which took more than four hours, the community members remained silent, moving their eyes back and forth between the accused and the witnesses. They looked emotionless. I would have liked to protest against the deadly silence that muted everyone with the exception of the witnesses and the defendants, but there was no space for that. As I later reflected on the session, I asked myself many different questions about the whole process of truth commissions and *gacaca* in particular: Can *gacaca* bring peace to the Rwandan communities? What did community members really want to say during that process? What actions do they take after participating in similar sessions week after week? Does this approach accomplish anything substantial in redressing past wrongs and healing psychosocial trauma? Is there any hope for reconciliation after going through *gacaca*?

In the session I attended, I did not see a community member or a trained professional handling the psychosocial aspects during the hearings. The judges had the power to establish order and command the proceedings according to the basic legal training they had received prior to *gacaca* hearings. They could impose sentences of 25 years of jail time. The community members confronted with the new *gacaca* were silent witnesses rather than active contributors to the reconciliatory processes promised in this form of justice. It is in the context of these contradictions between the legal and the psychosocial that the healing of psychosocial trauma is questioned.

Psychosocial trauma in post-conflict situations is understood from a structural violence framework as a combination of individual emotional wounds and the social suffering of communities. It links the feelings about experienced violence of the past to the barriers of victimhood, guilt, and fear that characterize post-conflict situations (Martín-Baró 1994, pp. 119–21). It encompasses the struggles of individuals and collectives for systemic social transformation (Chaudhry and Bertram 2009, pp. 299–300). When these struggles are nationwide and affect different levels of social structures, a more holistic approach is advised (Wessells and Monteiro 2004, p. 330). Truth commissions with their dual role to address both legal and

psychosocial trauma seemed ideal in many post-conflict situations, especially through their promise to encourage truth-telling, apology, and forgiveness. However, there are controversial opinions about the meanings of these dimensions and how they are integrated in the overall conceptualization and implementation of truth commissions.

Truth commissions combine legal, psychological, and religious understanding of the concepts of truth-telling, accountability, public acknowledgement, apology, forgiveness, and reconciliation. Legal and religious meanings are beyond the scope of this paper. However, as I draw my analysis from a psychological perspective, the legal and religious aspects serve as references. Concepts of accountability, public acknowledgement, and apology are all forms of recognizing the wrong done and the promise of not repeating past mistakes. In this paper, I use apology as the term most commonly used in truth commissions. Reconciliation will not be examined as a separate concept as the other dimensions are part of the reconciliation process.

Truth-telling

Truth commissions consider truth-telling an important dimension of addressing legal issues and psychosocial trauma that result from massive violence. These commissions have adopted a psychoanalytical language which proposes that revealing is healing and that unresolved memory of past violence can be overcome by remembering, telling, and forgetting (Humphrey 2000, p. 9). With a concentration on truth-telling, Alex Boraine (2006, p. 21) postulates that the telling of stories about dehumanizing acts can be publicly received with dignity when relayed in a poignant manner. He also asserts that truth commissions challenge people who deliberately ignore the suffering inflicted on others to stop saying that they did not know. Instead, they are offered the opportunity to cooperate with survivors who are seeking the truth about what happened to their loved ones. Cathartic reactions that result from the provided testimonies are supposed to facilitate transition from a wounded to a healed individual; the effects of the testimonies presumably impact those who do not have a chance to testify.

Truth-telling consists of individual narratives or expressions of a collectively shared understanding of the past gained from different levels of witnessing (Gobodo-Madikizela 2008, p. 176). The role of truth-telling in healing psychosocial trauma is ambiguous. As Mendeloff (2004, pp. 374–76) argues, truth-telling through

testimonies is very hard to prove empirically because it is subjective: human beings always remember and relate stories from a spatially and temporally limited perspective. From a political point of view, truth-telling is considered a tool to break the cycle of silence surrounding massive violence so that people will not say that they did not know (Boraine 2006, p. 22), but its psychosocial effect remains unclear. Gobodo-Madikizela (2008, p. 176) argues that reflection on one's own role in the past and the capacity to confront and acknowledge the wrong done should lead to reaching out to others in attempts to repair broken relationships. Using the example of South Africa, she explains that the TRC created a space for people to come together to forge a peaceful society by sharing emotions of pain, grief, anger, and resentment associated with a history of violence in a reflective dialogue. She adds that the public spaces open to the TRC hearings were sufficiently intimate to allow some acts of recognition, apology, and forgiveness. According to my observation during *gacaca*, emotions were obvious through the telling of and listening to testimonies. However, there was no space for dialogue. People did not display or express their opinions on the proceedings. It seemed to me that the place was actually unsafe for those giving testimonies and the defendants who had to prove their innocence in regard to the accusations made against them. Each of them addressed the judges, who had the ultimate power to decide the outcome of the proceedings, which included pressing charges or releasing the accused. Other community members were silent witnesses.

Apology

Apology has been greatly emphasized in truth commissions as the moral "right thing to do." It encompasses acknowledgement of the injustice, expression of regret, and acceptance of responsibility, including material or financial compensation. Sincere apology is a critical factor in restoring broken relationships (Gobodo-Madikizela 2008, p. 178). According to Allan and colleagues (2006, p. 99), genuine apology should go beyond general verbal apologies to incorporate apologetic behaviour that reflects the level of sincerity of the wrongdoers as perceived by the victims. For full-fledged completion, a genuine apology must elicit acceptable signs of empathy on the part of the offended party (Williams, Jr. 2003, p. 277).

Apology is often discouraged by the legal system, which focuses on the manipulation or denial of facts in order for the offence to avoid

any form of accountability. Accountability ranges from sincere acknowledgement of the wrong done and expression of regret to material or monetary compensation or other forms of punishment (Amstutz 2005, pp. 8–10), such as imprisonment. Although truth commissions put great emphasis on the process of apology, they tend to concentrate on factual truth and other simple public acknowledgment (Wolf 2006, pp. 28–30) and are reluctant to incorporate apologetic measures. In places like Peru, where the unrest is rooted in old animosity, compensation has been considered unreasonable (Garcia-Godos 2008, p. 79). In other countries where truth commissions recognize the importance of material compensation, financial compensation is seldom provided as a form of recognition of the injustice and pain inflicted on the victims, with a common excuse of insufficient funds. In places where massive violence was sponsored by the state, financial apology is seen as responsibility of the government (e.g., South Africa).

In the *gacaca* courts, however, compensation has been the burden of individual offenders rather than the state. Because of the poverty level of many of the front-line killers, individuals who cannot afford monetary compensation do community work, such as repair of roads or other tasks related to the public interest and not to the victims, many of whom also live in extreme poverty. This approach leaves both the victims and the perpetrator dissatisfied by the outcomes of *gacaca*, thereby creating the perception of another form of injustice. While the community work is good for all, survivors resent the fact that the government has done very little to help them rebuild their lives and the houses that were destroyed during the genocide. At the same time, the accused resent and blame the victims for what they view as an exploitative punishment. In this context, the process of apology becomes overwhelmed by emotions of fear, anger, and resentment, and has little social space to offer sincere apology and forgiveness.

The process of apology is emotional work that involves dealing with guilt, shame, anger, and pity for the different parties in conflict (Brendel 2006, pp. 15–16). The manner in which apology is expressed influences the nature of response that is offered by the injured party. There is a lack of empirical studies on the process of apology in the context of truth commissions. However, studies done on the concept of apology in social contexts other than truth commissions show that when offenders deny their offence and try to justify their wrongdoing, or ignore, avoid, or exclude the offended, they develop instances of

active or passive dissociation (Leary, Springer, Negel, et al. 1998, p. 1233). These attitudes may indicate lack of accountability on the part of the perpetrators, which in turn, can evoke more negative reactions from the offended and augment threats to attempts at re-establishing peace.

Although many truth commissions are theoretically supposed to enhance apologetic statements, they do not offer space to let emotions be expressed and processed so that genuine apology can take place. Rather, the focus is often on the testimonies, especially the perpetrators' accounts, to provide facts of the massive violence that occurred. Granting the perpetrators amnesty when information in testimonies matches the evidence sought by the commissioners also poses problems of whose justice is being sought. The amnesties are detached from remorse for the violation of victims' rights. The victims of violence are required to go along with the amnesty decision, hence creating another obstacle in the journey toward forgiveness by hindering opportunities to obtain the desired psychosocial relief and benefits.

Research shows that people who go through formalistic processes of apology such as amnesty or insincere apology may later regret having apologized or experience anger and protest their innocence by blaming the victims for the violence that occurred (Exline, Deshea, and Holeman 2007, p. 499). Genuine apology constitutes an important foundation for forgiveness (Girard and Mullet 1997, p. 219) and a two-way emotional process of giving and receiving: the presumed offenders feel relieved when they are able to express their regrets and shame about the wrong done. Similarly, the offended experience relief from the pain endured when their offenders genuinely admit the wrong done. It is this mutual communion of pain and emotion that transforms the affected individuals into renewed human beings who can in turn share the gift of forgiveness.

Forgiveness

The topic of forgiveness is considered one of the dimensions of truth commissions. As mentioned above, forgiveness and reconciliation, like truth-telling and apology, have different interpretations from different disciplines such as theology and some branches of psychology, including social psychology and, more recently, peace psychology. In this section, I examine the concept of forgiveness from a psychological perspective.

Although a dimension of truth commissions, forgiveness is rarely mentioned during the hearings. While forgiveness may be part of the discussions employed by other models of intervention in post-conflict situations, there are no organized activities dedicated to this particular dimension in a truth commission process. There is also a lack of empirical literature on this concept in the context of truth commissions. One study that was conducted in South Africa with the TRC shows that the theme of forgiveness was not mentioned during testimonies (Chapman 2007, p. 56). Participants reported that they were motivated to participate in the truth commission because of their desire to discover the truth about the human rights violations and the perpetrators. Some of the participants in Chapman's study stated that they gave testimonies either because they wanted to tell their stories and gain public acknowledgement, or have their names cleared in order to restore their dignity.

Why is the dimension of forgiveness not part of truth commission practice? In the next section I offer a theoretical and empirical analysis of forgiveness and argue that the current state of truth commissions does not provide space for this dimension. Generally, social relationships and psychosocial well-being are built on constant acts of forgiveness and reconciliatory processes. From a psychological perspective, people have the power to choose how they handle their emotions; therefore, they can consciously decide to forgive without necessarily eliminating all negative emotions (Worthington, Jr., Witvliet, Lerner, et al. 2005, p. 170). Emotional forgiveness suggests juxtaposing positive emotions such as empathy, compassion, sympathy, and love against the negative begrudging and bitter emotions associated with the inability to forgive. Releasing negative emotions such as anger, resentment, and the desire for revenge allows the forgivers to abandon not only their negative emotions but also their indifferent behaviour toward the offender (Staub 2006, pp. 886–87). This process may help the offended party to move past their negative emotions and even create peaceful coexistence with their offenders. However, this is a fragile forgiveness in which the root problem of disagreement remains untouched. Feelings of animosity are easily revived towards self or other human beings when there is no space to hear how the damage done affects the lives of people at conflict. This kind of forgiveness does not provide a good foundation on which to rebuild strong relationships. The concept of authentic forgiveness is embedded in inner and outer factors. By inner factors, I

mean internal human processes that may facilitate or hinder forgiveness. By outer factors, I refer to external factors from the social world of the person. In the next section, I offer an example of each category.

Research on individual differences identifies narcissism as a major intrinsic factor that hinders forgiveness and psychosocial well-being (Exline, Baumeister, Bushman, et al. 2004, pp. 908–9). Accordingly, narcissistic individuals tend to concentrate their efforts on the self and self-interests. They are easily offended and are often preoccupied with defending their rights and demanding legal justice to be rendered after harmful events. For the narcissists, transgression is a debt requiring payment and forgiveness is costly and morally unacceptable. Their unforgiving nature can lead to anger, anxiety, and other negative emotions, especially when society suggests acts of forgiveness after painful experiences. Interpersonal and psychological anxiety observed in narcissistic individuals tends to be negatively correlated with social connectedness (Lee and Robbins 1998, pp. 343–44) and prone to mental health issues. Opting for forgiveness through a truth commission without considering the different needs of the community members may create further psychological problems, especially for those who focus on their needs for legal justice.

External factors such as socio-cultural and political variables that encourage people to forgive their offenders are often rooted in cultural and religious belief systems. However, in some cases of religious forgiveness, people are motivated to "forgive and forget," particularly when violence is committed by members of the same religious community. This is also the case in some interdependent cultures. In these contexts, forgiveness becomes a requirement for maintaining peaceful coexistence.

Both internal and external factors play an important role in in-group and out-group relations. One of the conditions of belonging to a certain group is to accept group behaviours and attitudes that separate the in-group from the out-group in a conflict, show in-group favouritism, and denigrate members of the out-group (Reed II and Aquino 2003, p. 1271). In such a context, forgiving someone from an opposing group can threaten in-group cohesion or an individual right to choice. Forgiveness often goes beyond the exoneration of the out-group from past injuries with the expectation that direct engagement

with the out-group and reconciliatory processes will follow (Noor, Brown, and Prentice 2008, pp. 106–7).

Studies of in-group and out-group relations find the process of forgiveness to be influenced by the level of regard in-group members have for out-group members. Stangor and colleagues (2001, p. 493) showed that belonging to a group with high negative attitudes toward members of the out-group hinders the willingness to forgive. Empirical literature on the concept of forgiveness in post-conflict situations outside the work of truth commissions shows that when massive violence is understood as a general human tragedy, people are more forgiving and less inclined to assign collective guilt to the opposing group (Wohl and Branscombe 2005, pp. 300–301). A more recent example is the study Cehajic and colleagues (2008, pp. 361–64) conducted with high school and university students in Bosnia and Herzegovina. Their study revealed that, when these young people identified themselves as Bosnian, they showed reduced social distance from the out-group, greater tendency towards forgiveness, and increased trust in the other group in conflict. In contrast, other studies found that competitive victimhood and a high level of in-group identification discourage forgiveness of members of the out-group, especially when the in-group membership had links with political membership (Noor, Brown, and Prentice 2008, p. 106).

Although most of the empirical literature is mainly experimental and not conducted in post-conflict situations, it can be argued that truth commissions lack a careful analysis of factors that can enhance or hinder their implementation. There are assumptions that the ultimate goal of truth commissions is to achieve forgiveness and thus lead to national reconciliation. Research shows that social categorizations, especially the competing discourses of victims versus perpetrators, are obstacles to forgiveness. It is well known that in many divided societies people form new social categories based on shared experiences and other in-group identification such as ethnicity, race, or gender. The question to ask is whether truth commissions, in their mandate to facilitate forgiveness, have the capacity to minimize this divide between the in-group and out-group membership. I would argue that focusing on a particular event and collecting facts that are shared as testimonies at hearings can divide communities rather than enhance a unified process of forgiveness. In many countries where truth commissions are promoted, there is a great emphasis on victims and perpetrators as opposing groups in the conflict. However, this

black and white separation in many post-conflict situations is blurred by other social contributory factors which may discourage restoration of social connectedness and psychosocial healing.

In Rwanda for instance, *gacaca* hearings involved people who might be related through intermarriage. In the example provided above, I was shocked to learn that one of the accused men in pink (jail uniforms) was the son of the elderly woman. In some other anecdotal cases from *gacaca*, genocide suspects were accused of killing the loved ones of the people they hid and protected. These realities make it difficult to forgo negative emotions, thoughts, and actions in various post-conflict contexts. Imposing truth commissions without paying attention to the complex social contexts of post-conflict situations and the complex emotions at play may complicate forgiveness processes.

Studies conducted on the emotional state of people who participated in *gacaca* found that they experienced feelings of greater insecurity and fear after testifying (Brounéus 2008, pp. 66–71). They also reported physical and psychological difficulties immediately before, during, and after their testimonies, including shaking uncontrollably, fainting, and intense feelings of isolation. The assessment by Kanyangara and colleagues of the emotional climate and intergroup perceptions involving a group of survivors and prisoners accused of genocide crimes showed that *gacaca* hearings heightened the negative perceptions, which then prevailed for an extended period (2007, pp. 398–400). Emotions of sadness, fear, disgust, insecurity, and shame increased during the course of the hearings, especially for survivors. Prisoners who accepted their role in the genocide during *gacaca* hearings reported feeling an intensified sense of guilt.

Studies on emotions such as shame and guilt have shown high patterns of unique relationships between emotions and motivations when participants were asked to recall "other-caused events," whereas feelings of anger, sadness, and anxiety were highly interrelated with the "self-caused event" (Schmader and Lickel 2006, pp. 53–54). When these emotions are not given space to be processed, they can lead to the deterioration of individual and community psychosocial well-being. The *gacaca* hearings did not offer space to recognize the emotions or opinions of community members who attended the hearings week after week. Although the long-term impact of *gacaca* on the psychosocial well-being of individuals and communities is still unknown, the few studies and anecdotal evidence about the process

raise great concerns. There is a need for systematic longitudinal studies to assess the impact of the dual role of *gacaca* in healing psychosocial trauma and reconciling Rwandans. The lack of strategies to address emotions that come up before, during, and after the hearings can be evaluated as a general indication of conceptual weakness of many truth commissions.

Conceptual and Practical Issues Regarding Truth Commissions

Some of the conceptual issues about truth commissions are theoretical while others are methodological in nature. In the analysis of the different dimensions above, I presented some of the challenges that arise when the complex realities embedded in the social issues in many post-conflict societies are not considered. It may also be argued that the existing empirical literature on the different dimensions has not been utilized to provide the necessary background information on the dynamics of intergroup relations and internal and external factors that influence the different dimensions. Another central conceptual issue is the combination of retributive and restorative models of justice, which does not clarify the place of psychosocial trauma and what is done to address it. The combination of two opposite forms of justice is further explored in the following section.

Retributive and restorative justice models are two approaches rooted in dichotomous philosophical frameworks. Retributive justice, rooted in legal justice and supported by the liberal human rights tradition, is not an equalizing discourse. Although this approach seeks to punish the offender and vindicate the victim (Philport 2007, pp. 95–96), it does not always offer opportunities to explore problems that caused violence or the resulting complex issues such as psychosocial trauma. Restorative justice, on the other hand, originates from moral and religious discourse (Philport, 2007, p. 96; Teitel 2003, pp. 81–82), which encourages restoration of relationships and involves apology, forgiveness, and reconciliation between individuals and groups.

In practical terms, balancing retributive justice with restorative and reconciliation processes is a struggle for truth commissions. The combination of opposing perspectives in one intervention model has been considered a political strategy to manage post-conflict societies, but it fails, in my view, by misunderstanding the psychosocial issues in post-conflict situations and thus misusing the psychological terms. Truth commissions are often state-driven and imposed from above

(Ingeleaere 2008, p. 44; Nagy 2008, p. 276). They are conceived outside the context in which they are implemented and this abstracts them from the local realities of post-conflict societies. The lack of clarity and frustration resulting from various commissions is due, in part, to the lack of participation of local people in conceptualization and implementation. Their voices are silenced not only by violence and the inflicted pain, but also by models that apply a top-down perspective, do not reflect the local realities, and fail to identify and build on local coping mechanisms and resources. Nevertheless, truth commissions assume that giving testimonies, which is central to their functioning (Allan and Allan 2000, p. 469), contributes to psychosocial healing.

Asking people to relive traumatic experiences without adequate and appropriate support in place and trying to achieve on the social level what the psychological models attempt to accomplish on an individual level may actually be more traumatizing than healing (Allan 2000, p. 195; Allan and Allan 2000, p. 472; Ingeleaere 2008, pp. 50–51; Mendeloff 2004, p. 365). The TRC, which is upheld as a role model for other truth commissions, has itself been criticized as underestimating the extent of the psychological impact of public testimonies about acts of violence of the apartheid regime in South Africa (Allan 2000, p. 193). According to Allan, the initial draft of the TRC did not have a plan to support witnesses who had to relive their traumatic past through testimonies. Even when the testimonies were recognized as potentially retraumatizing, and psychologists were invited to provide help, the support was restricted to individuals who had given testimonies and not those who heard them. Testimonies of massive violence can be traumatic for both the witnesses and the public who follow these unfolding stories in silence.

Conceptual and Practical Issues Related to *Gacaca*

At the conceptual level, *gacaca* was initially praised for involving the local population in the election of judges (Schabas 2005, p. 893), known as *inyangamugayo*, which means the trusted or "righteous" members of the community. Both Hutu and Tutsi were chosen to implement this form of community justice. However, at the implementation level, *gacaca* created contradictions with the initial plan to hear the truth and reunite Rwandans. Ingeleaere's analysis (2008, pp. 57–58) shows that the collection of facts consisted mostly of accusations and the defendants' statements of self-defence; the

accused frequently gave evasive testimonies to cover up their own deeds or those of family members by admitting to minor crimes while attributing more serious violent acts to those who had died or disappeared. In addition, the collection of facts about past wrongs contradicted the information that had been used to motivate people to participate in *gacaca*. The written objectives and goals of this form of truth commission put emphasis on prosecution,[1] whereas people were told that it was reconciliation that was being sought.

During the proceedings, *gacaca* became more retributive than restorative, and the collected facts were utilized to prosecute the wrongdoers rather than restore relationships through processes of apology and forgiveness. Anecdotal reports revealed that the complicated cases that took more than one session were often cases that presented complex legal challenges rather than reconciliatory processes.

Another serious issue involved the judges training background. *Gacaca* was commonly praised as a homegrown approach in which the judges were elected by community members. However, the basic training these judges received was about legal prosecution and not processes of community reconciliation. Furthermore, the training was inadequate given the importance of the decisions they had to take. As a result, *gacaca* failed to meet some basic requirements of formal trials such as cross-examination and legal representation (Venter 2007, pp. 589–90) and it failed to heal and reconcile the people involved. Anecdotal evaluations of this process by older Rwandans who understood the difference between the old and new *gacaca* found the new one to be an instrument of the state influenced by international donors. *Gacaca*, like other many truth commissions, became what Humphrey calls "a bureaucratic response to bureaucratic murder" (2003, p. 14) that is driven by political and economic restructuring and governance rather a psychosocial grassroots intervention.

In old *gacaca*, the neutral and respected members of the community who were invited to resolve a particular issue met at the scene where the wrong had been done. The process included hearing the explanations of the parties in the conflict, collecting factual evidence from observation, and hearing testimonies of witnesses. Rather than imposing punishment, the two parties were given an opportunity to express their feelings and needs, and out of a negotiated process they agreed on the compensation for the wrong done. Other community

members were not silent witnesses. They participated in the discussions and even helped negotiate the compensation. If the offender did not have the means to compensate the victim for the wrong done, other members offered some assistance. Assessment of the wrong done was always followed by recognition of responsibility and an apology. The process was accompanied by rituals of shaking hands and sharing a drink which was purchased by both parties and other community members. The restoration of the broken relationship was an important outcome of this old grassroots form of *gacaca*. The involvement of the community in the process was a sign of solidarity. The old *gacaca* model aligns with some other grassroots approaches applied in post-conflict countries such as Peru[2] and East Timor's village-based Community Reconciliation Procedures (CRPs) (Burgess 2004, pp. 146–53).

Although these grassroots approaches tend to have little support and legitimacy, the applied methodology of bringing local people together to challenge individual actions and utilizing socio-cultural resources to repair the damage done seems to address psychosocial issues and reaches some desired outcomes by the local people.

Suggestions for Healing Psychosocial Trauma

The action of remembering and telling the stories of violence is necessary in order to stop the cycles of violence that threaten livelihood in post-conflict societies. Breaking the silence surrounding massive violence through public testimonies is important both politically and socially. Truth commissions can still have their role to act at the societal level to collect stories of violence and to use some case presentations as a form of recognizing the pain inflicted and the failure of the state to protect its citizens. However, it should be made clear that truth commissions are symbolic alternatives to legal justice and deal inadequately with psychosocial trauma. Thus, other models of intervention are called for to handle the psychosocial trauma that results from massive violence.

For psychosocial purposes, testimonies can be utilized not as just facts but as stories of lived experiences of the witnesses with the intent of motivating the population to seek psychosocial support and initiate creative ways to heal one another. An educational approach that uses case studies of testimonies can be a way to normalize mental health issues resulting from violent conflicts and their consequences. Grassroots emotional work to facilitate contact and exchange between

offenders and the offended about what had happened to them both as individuals and as a community must be encouraged and resources made available. This process can enable people who must live together to create new individual and collective narratives.

Local governments and external supporters have the responsibility to ensure security and the rule of law and to acknowledge the impact of massive violence on the psychosocial well-being of individuals and their communities. Local people should be identified as key players in the healing of their own trauma and the rebuilding of their communities. Affected individuals need to be aware of the suffering of other community members on all sides of a conflict. The depth and extent of healing will depend to a large degree on the willingness of individuals to mobilize other community members and share resources through social and psychological group processes.

On a concluding note, the healing of psychosocial trauma cannot start and finish with truth commissions or other imposed models (Stanley 2002, p. 3). Truth commissions should build on the local resources to support the healing of psychosocial trauma by enhancing processes of apology and forgiveness. To achieve the objective of healing psychosocial trauma in post-conflict situations, such as post-genocide Rwanda, requires active involvement of the members of parties in conflict and the creation of a contained space to facilitate the exploration of emotions that result from violence and its resulting psychosocial consequences.

Notes

1. *Gacaca* had the following objectives: (1) Find out the truth about what happened since residents shall be called upon as eyewitnesses to the acts of committed in their cells, and they shall compile a list of victims and perpetrators; (2) accelerate the prosecution of genocide since those who know what happened shall testify in the presence of their neighbours on the hills; (3) continue the eradication of the culture of impunity by using any method that makes it possible to identify a person who took part in the tragedy, since once the truth is known, none of those who were complicit shall escape punishment, and the people will understand that an offense results in the conviction of the criminal without any exception whatever; (4) punish those who played a part in the tragedy, reconcile the Rwandans, and strengthen their unity since the *gacaca* jurisdictions' system shall induce the residents of the same cell, sector, commune, and prefecture to collaborate in judging those who participated in the genocide, to discover the victims, and to restore their rights to innocent people; and (5) prove the capacity of the Rwandan custom, since, although the cases that the *gacaca* jurisdictions will have to hear are different from those that are normally resolved within the *gacaca*

framework, these jurisdictions fit well into the custom of settling differences by arbitration, even amicable arbitration (Transitional National Assembly of Rwanda Adopted Organic Law (TNARAOL) no.40/2000 of 16 January 2001).

2. The villagers from the Ayacucho region decided to render justice and reintegrate former Senderistas in a rehabilitative manner instead of handing them over to the military (Theidon 2006, pp. 434–56). Senderistas is the name of the guerrilla movement involved in ongoing fighting with the Peruvian government. The return to the community was facilitated through sharing accounts of the suffering the community experienced at the hands of the Senderistas and their own experiences in the mountains. A public commitment to never go back to guerrilla activities was obtained from the members of Senderistas who returned to the community. Although they received some form of physical punishment through a number of whippings, they were offered a piece of land on which to build a house and grow crops, and settle among other community members.

3. The CRPs were initiated in East Timor after the war between East Timor and Indonesia. During the war Indonesia hired East Timorese to carry out violence against their own fellow citizens. They participated in the killing and looting of their own villages. The CRPs' objective was to bring victims, perpetrators, and other community members together to undertake acts of reconciliation. The majority of perpetrators tried by the CRPs came from the same villages as their victims. On their release, they were allowed to request to be reaccepted into their communities. Hearings were facilitated by a committee formed by a panel of local church and spiritual leaders and elders, and chaired by a regional commissioner appointed by the East Timor government. Perpetrators were required to listen to their victims' stories, be confronted face to face with the pain and anger they caused, and given the opportunity to apologize. Punishment to restore some of the damage to the individual and community (e.g., building a school or paying back the stolen property) was decided by the community, and perpetrators had to comply. The hearings were completed by traditional ceremonial practices that included communal chewing of betel nut on a large mat before sending the final agreement to the district court for registration.

References

Allan, Alfred. 2000. Truth and Reconciliation: A Psycholegal Perspective. *Ethnicity and Health* 5 (3–4): 191–204.

Allan, Alfred, and Marietjie M. Allan. 2000. The South African Truth and Reconciliation Commission as a Therapeutic Tool. *Behavioural Sciences and the Law* 18 (4): 459–77.

Allan, Alfred, Maria M. Allan, Debra Kaminer, et al. 2006. Exploration of the Association between Apology and Forgiveness amongst Victims of Human Rights Violations. *Behavioural Sciences and the Law* 24 (1): 87–102.

Amstutz, Mark R. 2005. *The Healing of Nations: The Promise and Limits of Political Forgiveness*. Lanham, Md.: Rowman & Littlefield Publishers.

Boraine, Alex. 2006. Transitional Justice: A Holistic Interpretation. *Journal of International Affairs* 60 (1): 17–28.

Brendel, David H. 2006. Psychotherapy and the Truth and Reconciliation Commission: The Dialectic of Individual and Collective Healing. In *Trauma, Truth*

and Reconciliation. Healing Damaged Relationships, edited by Nancy Nyquist Potter, 15–27. New York: Oxford University Press.

Brounéus, Karen. 2008. Truth-telling as Talking Cure? Insecurity and Retraumatization in the Rwandan Gacaca Courts. *Security Dialogue* 39 (1): 55–76.

Burgess, Patrick. 2004. Justice and Reconciliation in East-Timor. The Relationship between the Commission for Reception, Truth and Reconciliation and the Courts. *Criminal Law Forum*, 15 (1–2): 135–58.

Cehajic, Sabina, Rupert Brown, and Emanuele Castano. 2008. Forgive and Forget? Antecedents and Consequences of Intergroup Forgiveness in Bosnia and Herzegovina. *Political Psychology* 29 (3): 351–67.

Chapman, Audrey A. 2007. Truth Commissions and Intergroup Forgiveness. The Case of South African Truth and Reconciliation Commission. *Peace and Conflict* 13 (1): 51–69.

Chaudhry, Lubna Nazir, and Corrine Bertram. 2009. Narrating Trauma and Reconstruction in Post-conflict Karachi: Feminist Liberal Psychology and the Contours of Agency in the Margins. *Feminism & Psychology* 19 (3): 298–312.

Exline, Julie Juola, Roy F. Baumeister, Brad J. Bushman, et al. 2004. Too Proud to Let Go: Narcissistic Entitlement as a Barrier to Forgiveness. *Journal of Personality and Social Psychology* 87 (6): 894–912.

Exline, Julie J., Lise Deshea, and Virgina Todd Holeman. 2007. Is Apology Worth the Risk? Predictors, Outcomes, and Ways to Avoid Regret. *Journal of Social and Clinical Psychology* 26 (4): 479–504.

Garcia-Godos, Jemima. 2008. Victims Reparations in the Peruvian Truth Commissions and the Challenges of Historical Interpretation. *International Journal of Transition Justice* 2 (1): 63–82.

Girard, Michèle, and Etienne Mullet. 1997. Propensity to Forgive in Adolescents, Young Adults, Older Adults, and Elderly People. *Journal of Adult Development* 4 (4): 209–20.

Gobodo-Madikizela, Pumla. 2008. Trauma, Forgiveness and the Witnessing Dance: Making Public Spaces Intimate. *Journal of Analytical Psychology* 53 (2): 169–88.

Haleri, Shlomo, and Zvi Eisikovits. 2006. The Role of Communicating Social Emotions Accompanying Apologies in Forgiveness. *Motivation and Emotion* 30 (3): 189–97.

Humphrey, Michael. 2000. From Terror to Trauma: Commissioning Truth for National Reconciliation. *Social Identities* 6 (1): 7–27.

———. 2003. From Victimhood: Truth Commissions and Trials as Rituals of Political Transition and Individual Healing. *The Australian Journal of Anthropology* 14 (2): 171–87.

Ingeleaere, B. 2008. The Gacaca Courts in Rwanda. In *Traditional Justice and Reconciliation after Violent Conflict. Learning from African Experiences*, edited by Luc Huyse and Mark Salter, 25-59. Stockholm: International Institute for Democracy and Electoral Assistance.

Kanyangara, Patrick, Bernard Rimé, Pierre Philippot, and Vincent Yzelbyt. 2007. Collective Rituals, Emotional Climate and Intergroup Perceptions. Participation in

Gacaca Tribunals and Assimilation of the Rwandan Genocide. *Journal of Social Issues* 63 (2): 387–403.

Leary, Mark R., Carrie Springer, Laura Negel, et al. 1998. The Causes, Phenomenology, and Consequences of Hurt Feelings. *Journal of Personality and Social Psychology* 74 (5): 1225–37.

Lee, Richard M., and Steven B. Robbins. 1998. The Relationship between Social Connectedness and Anxiety, Self-esteem, and Social Identity. *Journal of Counselling Psychology* 45 (3): 338–45.

Martín-Baró, Ignacio. 1994. *Writings for a Liberation Psychology*. Ed. Adrianne Aron and Shawn Corne. Cambridge, Mass.: Harvard University Press.

Mendeloff, David. 2004. Truth-seeking, Truth-telling, and Postconflict Peacebuilding: Curb the Enthusiasm? *International Studies Review* 6 (3): 355–80.

Nagy, Rosemary. 2008. Transitional Justice as Global Project: Critical Reflections. *Third World Quarterly* 29 (2): 275–89.

Noor, Masi, Rupert Brown, and Garry Prentice. 2008. Prospects for Intergroup Reconciliation: Social-psychological Predictors of Intergroup Forgiveness and Reparation in Northern Ireland and Chile. In *Social Psychology and Intergroup Reconciliation*, edited by Arie Nadler, Thomas Malloy, and Jeffrey Fisher, 97–115. Oxford: Oxford University Press.

Philport, Daniel. 2007. What Religion Brings to the Politics of Transitional Justice. *Journal of International Affairs* 61 (1): 93–111.

Reed II, Americus, and Karl F. Aquino. 2003. Moral Identity and the Circle of Moral Regard towards Out-groups. *Journal of Personality and Social Psychology* 84 (6): 1270–86.

Schabas, William. A. 2005. Genocide Trials and Gacaca Courts. *Journal of International Criminal Justice* 3 (4): 879–95.

Schmader, T., and B. Lickel. 2006. The Approach and Avoidance Function of Guilt and Shame Emotions: Comparing Reactions to Self-caused Wrongdoing. *Motivation and Emotion* 30 (1): 43–56.

Stangor, Charles, Gretchen B. Sechrist, and John T. Jost. 2001. Changing Racial Beliefs by Providing Consensus Information. *Personality and Social Psychology Bulletin* 27 (4): 486–96.

Stanley, Elizabeth. 2002. What Next? The Aftermath of Organized Truth Telling. *Race & Class* 44 (1): 1–15.

Staub, Ervin. 2006. Reconciliation after Genocide, Mass Killing, or Intractable Conflict: Understanding the Roots of Violence, Psychological Recovery, and Steps toward a General Theory. *Political Psychology* 27 (6): 867–94.

Teitel, Ruti G. 2003. Transitional Justice Genealogy. *Harvard Human Rights Journal* 16: 69–94.

Theidon, Kimberly. 2006. Justice in Transition: The Micropolitics of Reconciliation in Postwar Peru. *Journal of Conflict Resolution* 50 (3): 433–57.

Tiemessen, Alana Erin. 2004. After Arusha: Gacaca Justice in Post-genocide Rwanda. *African Studies Quarterly* 8 (1): 57–75.

Venter, Christine. M. 2007. Eliminating Fear through Recreating Community in Rwanda: The Role of Gacaca Courts. *Texas Wesleyan Law Review* 13 (2): 577–97.

Wessells, Michael G., and Carlinda Monteiro. 2004. Healing the Wounds following Protracted Conflicts in Angola: A Community-based Approach to Assisting War-affected Children. In *Handbook of Culture, Therapy, and Healing*, edited by Uwe P. Gielen, Jefferson M. Fish, and Juris G. Draguns, 321–41. Mahwah, New Jersey: Laurence Erlbaum Associates Publishers.

Williams, Jr., Robin. M. 2003. *The Wars Within. People and States in Conflict*. New York: Cornell University Press.

Wohl, Michael J.A., and Nyla R. Branscombe. 2005. Forgiveness and Collective Guilt Assignment to Historical Perpetrator Groups Depend on Level of Social Category Inclusive. *Journal of Personality and Social Psychology* 88 (2): 288–303.

Wolf, Miroslav. 2006. *The End of Memory. Remembering Rightly in a Violent World*. Grand Rapids, Mich.: Eerdmans Publishing.

Worthington, Jr., Everett L., Charlotte vanOyen Witvliet, Andrea J. Lerner, et al. 2005. Forgiveness in Healing Research and Medical Practice. *Explore* 1 (3): 169–76.

The Effect of Contextual Differences on Health Resource Allocation Decision Making

Roger Chafe, Doreen Neville, and Thomas Rathwell

Abstract

Little consideration has been given to the influence contextual factors can have on health resource allocation processes. This article uses comparisons among nine case studies of resource allocation decision making to determine the influence that service area, the health care system, and health care organization have on the processes by which these decisions are made and on the factors considered in making specific allocation decisions. The study found that contextual issues clearly affect the structure of allocation processes, the type of allocation questions an organization faces, the level of decision maker who make allocation decisions, and the factors considered in making these decisions. One of the main implications of the findings is that proposals for reforming resource allocation will likely need to be more attuned to the institutional and program-level environments in which they are being employed.

Introduction

There is a growing recognition of the importance context plays in determining both health care decision-making processes and decision outcomes (Abelson 2001; Abelson et al. 2007; Cummings et al. 2007; Culyer and Lomas 2006; Dobrow et al. 2006; Lomas et al. 2005). Although they are crucial to how decisions get made, contextual factors are usually factors which only concern the organization making the decision, e.g., considerations about whether the organization has available resources or sufficient staffing to offer a new service, whether the proposal fits with other organizational initiatives, etc. These factors are usually contrasted with scientific research evidence, which is seen to be applicable across decision-making settings, i.e., it is "context independent" (Lomas et al. 2005).

Given the range of possibly relevant contextual factors, researchers need to specify which contextual factors they plan to focus on in their analysis. In this article, we examine three contextual factors that seem particularly important to health resource allocation: (1) the type of service for which resources are being allocated, (2) the health care organization, and (3) the province where the decision is being made. In their studies comparing hospital care rationing in the United Kingdom and the United States, Aaron and colleagues (Aaron and Schwartz 1984; Aaron, Schwartz, and Cox 2005) conclude that the mechanisms of health care rationing, and the public's reaction to it, are quite different between the two countries due, in part, to organizational and cultural differences between their health care systems. Others, too, have found contextual differences in the allocation of resources across different organizational and political settings to be important (Kapiriri, Norheim, and Martin 2007; Martin, Pater, and Singer 2001; Menon, Stafinski, and Martin 2007). Given that provinces have ultimate constitutional authority within the Canadian health care system, it is important to consider the extent to which the organizational and cultural differences between provincial health care systems influence the way health care resources are allocated in different parts of the country.

Most provinces in Canada employ a regional health structure in which health care is provided by individual health care organizations, each with their own geographically defined area. Not only are there different organizational cultures in health care organizations, particularly when these health care organizations are in different provinces, there may also be structural differences in the processes for allocating resources from one regional health authority (RHA) to another. Comparing decision making in different provinces also allows, then, for the opportunity to compare allocation processes in different health care organizations.

Another important contextual factor is the area of care or service for which resources are being allocated. Few studies have examined the effect that the subject matter of resource allocation decisions has on how resource allocations are made, even though the dynamics of how resources are allocated may vary dramatically in different areas. Key characteristics of the decision (e.g., level of scientific evidence underlying a decision, the structures by which decisions are made, and who ultimately makes these decisions) may vary depending on the content area.

This article explores these issues by examining how resources are allocated in three areas of care (acute care, diagnostics, and rehabilitation) in three Canadian provinces (Alberta, Newfoundland and Labrador, and Saskatchewan). The discussion focuses on one service in each area, specifically, how resources are allocated for endovascular coiling (acute care), MRI (diagnostic), and powered upper limb prostheses (rehabilitation). The purpose is to identify how the service and/or organizational context influences: (1) the decision-making structure, (2) who is involved in making allocation decisions concerning the service, and (3) the factors considered in making specific allocation decisions.

Methods

The project uses a multiple case comparison model based on nine case studies. The cases are distinguished in terms of their province and type of service (Table 1). The three provinces—Alberta, Newfoundland and Labrador, and Saskatchewan—were chosen based on variations in their population size, geographic distribution, structure of their health care systems, and financial strength. In order to provide a reasonable scope to the project, decision making in the provincial department of health and one RHA were focused on in each province. The three services selected encompassed a diverse set of cases in three different areas of care. Endovascular neurocoiling is a treatment option for cerebral aneurysms, in which the aneurysm is filled with platinum coils, blocking blood flow to the aneurysm and decreasing the risk of hemorrhaging. At the time of data collection in 2006, endovascular coiling was available only in Alberta, although both Newfoundland and Saskatchewan were considering establishing programs. Magnetic Resonance Imaging (MRI) is a well-known, non-invasive diagnostic test, used for a wide range of disease conditions. A key issue for resource allocation for this service is that MRI is still a developing technology, expanding both in its capabilities (Fujita 2007; Strijkers et al. 2007) and in the range of cases for which it is being recommended (Bagarinao, Nakai, and Tanaka 2006; Lima and Desai 2004; Richardson et al. 2005; Zur et al. 2004). Wait times for MRIs also hold a prominent place in debates in Canada about the accessibility of publicly funded health care. Powered upper limb prostheses attempt to compensate for the loss of a hand or arm, both in terms of functionality and appearance. Powered prostheses have battery-power motors, which allow for the opening, closing, and

turning of the prosthesis, thereby increasing their functionality for some patients, but also their costs when compared with more traditional, mechanical prosthesis. All three RHAs provide powered upper limb prostheses. There are, however, variations in the levels of public coverage for powered upper limb prostheses across the three provinces. The nine cases aim to provide sufficient variation in terms of the province, the health care organization, and the service area to allow contextual differences to be clearly identified.

Table 1: Identification of Nine Case Studies

	Alberta	Newfoundland	Saskatchewan
Endovascular Coiling	1	4	7
MRI	2	5	8
Powered Upper Limb Prosthesis	3	6	9

All the cases were developed using both document reviews and key informant interviews. The documents reviewed included websites, published reports, news releases, RHA business plans, and government legislation. Key informants were all directly involved in the allocation of resources related to the nine cases and included program-level decision makers (including frontline providers), regional executive team members, and officials within the provincial departments of health. Interviews were semi-structured, using a responsive interviewing approach (Rubin and Rubin 2004). Interview tapes and field notes were all transcribed, coded, and analyzed using N6 qualitative software.

Ethical approvals for the project were granted by Memorial University of Newfoundland's Human Investigation Committee, the University of Alberta's Health Research Ethics Board, and the University of Saskatchewan's Behavioural Research Ethics Board. Operational approvals were granted by the three RHAs prior to interviews with staff members.

Findings

By focusing on endovascular coiling, MRI, and powered upper limb prostheses, a good level of variation was established across the services in terms of the size of the estimated patient populations, the

potential impact of the intervention on patients, the strength of evidence for their effectiveness, the capital costs involved, their cost per case treated, and in the scope of public coverage (Table 2).

Table 2: Overview of Services Examined (Based on Key Informant Estimates)

	Endovascular Coiling	MRI	Powered Upper Limb Prostheses
Annual Patient Population	Between 1 to 20 procedures per 100,000	Between 850 to 3600 scans per 100,000	Less than 1 prosthesis per 100,000
Impact	Potential life saving	Varies depending on application	Improves functionality and self-esteem
Evidence for Effectiveness	Little clinical research; Accepted practice by most neurologists and radiologists	Well-established technology; Level of evidence for effectiveness varies depending on application	Well-established technology; Effectiveness dependent on compliance and suitability for individual patients
Cost (Est.)	$3 million initial investment; $10,000 per case	$1.5 to $4 million initial investment; $500 per scan	Between $6000 and $35,000 per case
Public Coverage	Full universal public coverage	Full universal public coverage	AB—Partial universal public coverage NL—No universal coverage SK—Full universal public coverage

Of the 62 people invited to participate, 43 key informants gave interviews, for a participation rate of 69%. Table 3 provides a breakdown of key informants by their decision-making position and province. Searches were made of a number of relevant websites and news archives. Over 40 documents relating to their organizations and resource allocation in the selected areas were also given to researcher directly by interviewees.

Although the focus is on the allocation of resources for different health care services in different provinces, the cases share the same basic

Table 3: Breakdown of Interviewees by Position and Province

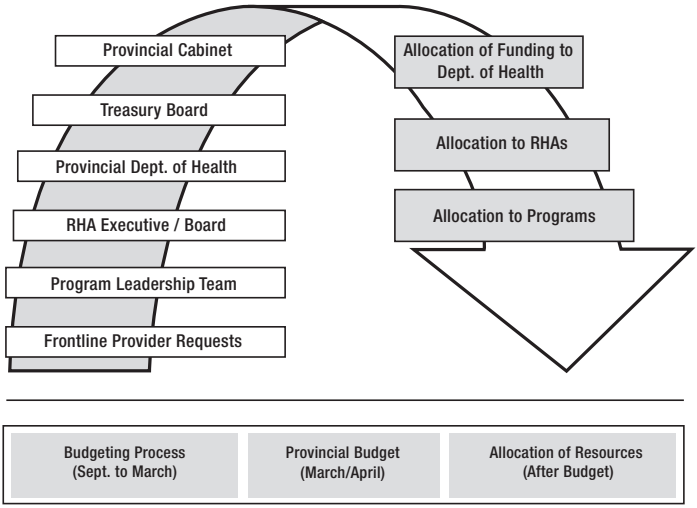
	Alberta	Newfoundland	Saskatchewan	TOTAL
Provincial Officials	3	2	4	9
Regional Executive Team Members	6	6	4	16
Program-Level Decision Maker: Endovascular Coiling/ MRI	3	4	2	9
Program-Level Decision Maker: Powered Upper Limb Prostheses	2	3	4	9
TOTAL	14	15	14	43

structure in terms of how most requests for resources are considered. This is due to the fact that almost all increases in health care expenditures ultimately must be approved as part of the provincial department of health's or the RHA's annual budgets. Where differences do occur across the services, as illustrated below, it is usually in terms of what level in the decision-making structure specific allocation decisions usually get made.

For all nine cases, requests for increased resources and/or for the adoption of a new technology usually come from providers or the program's leadership team, consisting of both a clinical and an administrative program lead. One notable exception has been with MRI, where there has been targeted federal funding provided through the 2003 First Ministers' Accord on Health Care Renewal. Typically, requests for additional resources are submitted as part of the program's annual budget submission to the RHA's executive, although some requests are made directly to the RHA's executive or the department of health outside of the normal budget cycle. The provincial department of health then incorporates the RHA's submission in its budget request to the provincial treasury board, which prepares the provincial budget for cabinet approval. Following the approval of the provincial budget, usually in March or April, the provincial department of health

allocates funding to each RHA. The RHA's executive team then allocates resources to the programs and approves particular program requests. In some cases, the provincial department of health specifically targets increases in RHA funding to specific programs or purchases. An overview of the standard RHA/provincial budget cycle is presented in Figure 1. Participants stated that the vast majority of resource allocation decisions, particularly for increases in resources, are approved through this standard budgeting structure.

Figure 1: Standard RHA/Provincial Government Budgeting Cycle



While all health resource allocations are embedded within the wider government budget structure, there were still clear contextual differences shown across services, RHAs, and provinces. One key difference among the three services is who makes most of the significant allocation decisions in the area. In Newfoundland and Saskatchewan, which were both considering starting endovascular coiling programs, discussions occurred at the executive level and, in Newfoundland, also with the provincial department of health, about whether to fund a new program. MRI capacity and wait times are closely watched by both provincial and regional executive-level decision makers across the three provinces. On the other hand, there was very little consideration given to funding powered upper limb prostheses at either the regional executive or the provincial government level. In fact, only participants who were directly

involved in the management of rehabilitation programs report giving any consideration to powered upper limb prostheses. Not only were powered upper limb prostheses not on any of the regions' or provinces' current agendas, but most participants felt that coverage and program structures likely reflected decisions made well in the past about how to handle prosthetics generally. This lack of senior decision maker attention is likely due to the service's extremely small patient population and to the fact that it is not a life-saving service, compared, e.g., with endovascular coiling. Not being on the provincial or RHA's agenda above the program level does, however, make it difficult to enact major program changes, extend public coverage, or substantially expand funding.

The three services examined require different levels of resources. Both endovascular coiling and MRI require millions of dollars of initial capital investment. This is not the case for powered upper limb prosthesis, the main cost of which relates to the prosthesis itself. Diagnostics, and MRI in particular, is an area which has received significant infusions of funds in recent years. The same cannot be said for the area of rehabilitation. These differences in the level of funding and in the likelihood of increased funding seem to affect providers' attitudes towards resource usage. While all providers reported the need for increased resources to improve patient care, providers and decision makers in the area of rehabilitation were more accepting of resource constraints and of the need to make do with existing resources. In rehabilitation, providers were much more likely to report being cognizant of cost as a consideration in their recommendation of a treatment option, i.e., whether they recommended a mechanical or powered prosthesis.

Health care organizations across the country do not face the same resource allocation questions within the same service area at the same time. The decision to establish a program requires the consideration of numerous factors and often a substantial commitment of resources. Once a program is established, the resource allocation decisions faced by a health care organization change. For example, the RHAs differ with respect to where they are in the adoption of endovascular coiling. Alberta has performed the procedure for about eight years. The executives of the RHAS in Newfoundland and Saskatchewan were, at the time of this study, considering starting programs. In Alberta, once the decision to invest in performing the procedure was made, the executive was generally no longer involved. Rather, the main resource

allocation questions were made by the health care providers and the program management team.

Similarly, while all participants with management responsibilities for MRI recognized the developing nature of the technology, no participant expressed the need to review its effectiveness for these new applications. Once a technology is adopted and providers and decision makers become familiar with it, there is much less scrutiny of the technology, even if its application is extended to other groups of patients not considered when the technology was first introduced. Decision makers did recognize this type of usage creep and its impact on demand for MRI as a significant issue when allocating resources for this service area.

Variations in existing capacities and infrastructure put some organizations in a better position to establish new services. These variations in starting points can cause RHAs to face different resource allocation decisions, even when two organizations are at the same point in the adoption of a service. Although Newfoundland and Saskatchewan are considering establishing endovascular coiling programs, they face very different start-up costs. Saskatchewan, which already has decided to purchase a bi-plane angiography suite, is considering whether to purchase an initial inventory of platinum coils and cover ongoing operational costs. Newfoundland is considering whether to purchase a bi-plane angiography suite, purchase an inventory of new coils, and cover operational costs. For Saskatchewan, the costs associated with starting an endovascular coiling program are in the hundreds of thousands of dollars; for Newfoundland, in the millions.

These cost differences are reflected in the main factors and issues considered in whether to establish a program in each RHA. In Saskatchewan, discussions have centred on the question of whether there is sufficient demand for the service for staff to be able to maintain their skill level with the procedure. In Newfoundland, the focus primarily is on whether the service is justified given its high capital costs. Furthermore, the greater the financial cost of starting a program, the more likely it is that participants will be concerned with the available research evidence supporting the service's effectiveness and raise ethical concerns. Newfoundland is alone in having conducted any review of the effectiveness of endovascular neurocoiling or considered the ethical implications of spending so many financial resources for a small group of critically ill patients.

In all three provinces, RHAs are funded almost exclusively through the provincial budget. However, there are differences in the methods used across the three provinces for determining the size of each RHA's global budget. Alberta has the most structured approach, with a clear population-based formula for determining most global funding to its RHAs. Saskatchewan Health reported using a population-based formula to examine the proportion of funding going to different RHAs, but they have in practice relied on historical funding allocations with increases targeted to specific programs. Newfoundland has also used primarily historical funding patterns and targeted funding to determine each RHA's budget.

Participants in Alberta said that Alberta Health and Wellness explicitly tries not to "micromanage" their health care system. This philosophy is reflected in their decreased use of targeted funding. One participant from Saskatchewan countered that the need to micromanage health care resources comes from the historical scarcity of resources and that Alberta's choice not to do so simply reflects its stronger financial situation. Another possible reason for the adoption of different funding methods, as reported by a participant in Newfoundland, is that the department of health in a smaller province, especially in the case of a small number of RHAs, may feel that it is familiar enough with each RHA's activities to negate the need for more structured allocation processes.

Regardless of the reason why there are differences in funding arrangements, the increased use of targeted funding results in provincial departments of health being much more involved in program-level decision making. In Newfoundland and Saskatchewan, participants also reported that the provincial cabinet is more regularly involved in making spending decisions for specific program requests. For example, the decision whether to purchase a new MRI scanner would be a cabinet level decision in the two smaller provinces. In contrast, participants in Alberta, within both the provincial government and within the RHA, reported that the decision to purchase an MRI would be made primarily at the RHA level. While all three health care systems are ultimately accountable to the provincial legislature, the regular involvement of the provincial cabinet in making specific allocation decisions brings in another level of decision making at which program requests can be fulfilled. It also increases the risk that political calculations will determine which program requests are funded. As an example of this cabinet-level political

influence, participants reported that political calculations played a large part in the Newfoundland cabinet's decision about where to locate the province's second MRI scanner.

Discussion

Contextual factors related to service type, the RHA, and provinces do affect the structure of allocation processes, the type of allocation questions an organization faces, who makes relevant allocation decisions, and the factors which are considered in making these decisions. Other contextual factors that were also identified as affecting resource allocation include where a health care organization is in the adoption of a technology or service; the financial strength of the organization and the size of its client population; an organization's existing institutional capacity; the size of the required resource expenditure, particularly in relation to the organization's overall budget; and the financial strength of the province. In short, resource allocation is clearly affected by the institutional setting in which it is carried out.

Across the three services, resource allocation questions vary in terms of decision makers' ability to get them on the agenda above the program level, providers' consideration of the use of resources in providing care, and who, in the end, decides on how resources are allocated in the area. As a result, services differ in their ability to attract greater resources or allow for major program-level changes. MRI is one of the more high-profile health care services. Powered upper limb prosthetics obviously has a lower profile. Part of the reason for the higher profile of MRI is understandable and reasonable, given its much greater client population and its role in the diagnosis of a wide range of disease conditions. Still, the lesson for lower-profile services, particularly ones with very small client bases or which can be seen as elective, is that extra effort is required by patients and providers to ensure that these programs get the attention and the resources they require. If not, clients of lower-profile services will likely face inappropriate shortfalls in resources out of proportion to their population size and level of health burden.

Because the allocation of health care resources is firmly embedded within wider government budgeting processes, there is always going to be some level of political influence on how health care resources are allocated. Contextual factors, particularly related to the population size and the financial strength of the province, do, however, affect the

level of political influence and involvement in program-level decisions. Within the smaller provinces, politicians are more likely to be involved in program-level decision making and there is a greater use of targeted funding to the RHAs. Participants report that some allocation decisions are based almost exclusively on political considerations, as opposed to considering what employment of resources is most clinically appropriate, best addresses patients' needs, or maximizes efficiencies. It is unclear if the involvement of politicians necessarily makes for poorer allocation decisions. As public representatives, they clearly have a role to play in directing how public funds are spent. Yet for more contentious coverage and allocation decisions, it is perhaps more useful to directly and explicitly involve the public rather than risk such decisions appearing to be based solely on political calculations (Abelson et al. 2007; Chafe et al. 2008).

One of the main implications for researchers is that proposals for reforming allocation processes will likely need to be more attuned to the institutional and program-level environments in which they are being employed. Although there are difficulties in affecting closer relationships with health care organizations (Mitton and Bate 2007), some researchers have already begun to recognize the need for research methods in this area which are sensitive to contextual differences, including the use of participatory action research (Mitton et al. 2003; Patten, Mitton, and Donaldson 2006).

Beyond identifying the influence of context on resource allocation, clear variations were found in access to care across the three provinces for each of the three services examined. The fact that all provincial health care systems act in accordance with the conditions set out in the Canada Health Act (1984) and receive fiscal support from the federal government to help cover health care costs may lead Canadians to think that people in different provinces and different regions receive relatively the same level of care. The decision makers interviewed recognize that there are clear variations in the level of resources different provinces and regions have at their disposal to provide care. While none of the key informants went as far as to say that there are geographical variations in patient outcomes, the clear variations in the resources available, level of access to care that is available, and in the extent of public coverage as identified by the key informants present a significant challenge to the ideal that all Canadians have equal access to the same level of medical care.

Conclusion

The article illustrates the influence context can have on resource allocation processes and decisions. These differences depend on existing staff capabilities, operational capacities, and capital infrastructure, as well as where an organization is in the establishment of the service. Improving how health care resources are allocated will not allow public coverage to meet all the existing and prospective demands. Difficult and complicated choices will still have to be made by decision makers at all levels. It is hoped that better resource allocation decisions will lead to improved health outcomes, allow for a more efficient running of the health care system, and make the system more responsive to the needs of the public. By improving our understanding of the extent to which key contextual factors can impact on these processes, we will hopefully be able to develop more viable reforms to achieve these goals.

References

Aaron, H., and W. Schwartz. 1984. *The Painful Prescription: Rationing Hospital Care*. Washington, D.C.: The Brookings Institution.

Aaron, H., W. Schwartz, and M. Cox. 2005. *Can We Say No?* Washington: The Brookings Institution.

Abelson, J. 2001. Understanding the Role of Contextual Influences on Local Health-care Decision Making: Case Study Results from Ontario, Canada. *Social Science and Medicine* 53: 777–93.

Abelson, J., P. Forest, J. Eyles, A. Casebeer, E. Martin, and G. Mackean. 2007. Examining the Role of Context in the Implementation of a Deliberative Public Participation Experiment: Results from a Canadian Comparative Study. *Social Science and Medicine* 64: 2115–28.

Bagarinao, E., T. Nakai, and Y. Tanaka. 2006. Real-time Functional MRI: Development and Emerging Applications. *Magnetic Resonance in Medical Science* 5: 157–65.

Chafe, R., D. Neville, T. Rathwell, R. Deber, N. Kenny, L. Nestman, et al. 2008. *A Framework for Involving the Public in the Medicare Coverage Decisions*. http://schoolofhealthservicesadministration.dal.ca/Files/Basket_FrameworkPublicHealth carefinal.pdf (accessed February 11, 2008).

Culyer, A., and J. Lomas. 2006. Deliberative Processes and Evidence-informed Decision making in Health Care: Do They Work and How Might We Know? *Evidence and Policy* 2: 357–71.

Cummings, G., C. Estabrooks, W. Midodzi, L. Wallin, and L. Hayduk. 2007. Influence of Organizational Characteristics and Context on Research Utilization. *Nursing Research* 56: S24–S39.

Dobrow, M., V. Goel, L. Lemieux-Charles, and N. Black. 2006. The Impact of Context of Evidence Utilization: A Framework for Expert Groups Developing Health Policy Recommendations. *Social Science and Medicine* 63: 1811–24.

Fujita, H. 2007. New Horizons in MR Technology: RF Coil Designs and Trends. *Magnetic Resonance in Medical Science* 6: 29–42.

Kapiriri, L., O. Norheim, and D. Martin. 2007. Priority Setting at the Micro-, Meso- and Macro-levels in Canada, Norway and Uganda. *Health Policy* 82: 78–94.

Lima, J., and M. Desai. 2004. Cardiovascular Magnetic Resonance Imaging: Current and Emerging Applications. *Journal of the American College of Cardiology* 44: 1164–71.

Lomas, J., A. Culyer, C. McCutcheon, L. McAuley, and S. Law. 2005. *Conceptualizing and Combining Evidence for Health System Guidance*. Ottawa: Canadian Health Services Research Foundation, available at http://www.chsrf.ca/other_documents/pdf/evidence_e.pdf (accessed 12 February 2008).

Martin, D., J. Pater, and P. Singer. 2001. Priority-setting Decisions for New Cancer Drugs: A Qualitative Case Study. *Lancet* 358: 1676–81.

Menon, D., T. Stafinski, and D. Martin. 2007. Priority-setting for Health Care: Who, How, and Is It Fair? *Health Policy* 84: 220–33.

Mitton, C., and A. Bate. 2007. Speaking at Cross-purposes or across Boundaries? *Health care Policy* 3: 32–37.

Mitton, C., S. Patten, H. Waldner, and C. Donaldson. 2003. Priority Setting in Health Authorities: A Novel Approach to a Historical Activity. *Social Science and Medicine* 5: 1653–63.

Patten, S., C. Mitton, and C. Donaldson. 2006. Using Participatory Action Research to Build a Priority Setting Process in a Canadian Regional Health Authority. *Social Science and Medicine* 63: 1121–34.

Richardson, J., R. Bowtell, K. Mader, and C. Melia. 2005. Pharmaceutical Applications of Magnetic Resonance Imaging (MRI). *Advanced Drug Delivery Review* 57: 1191–209.

Rubin, H., and I. Rubin. 2004. *Qualitative Interviewing: The Art of Hearing Data,* 2nd ed. Thousand Oaks: Sage Publications.

Strijkers, G., W. Mulder, G. van Tiborg, and K. Nicolay. 2007. MRI Contrast Agents: Current Status and Future Perspectives. *Anticancer Agents Mediclinal Chemistry* 7: 291–305.

Zur, K., S. Holland, W. Yuan, and D. Choo. 2004. Functional Magnetic Resonance Imaging: Contemporary and Future Use. *Current Opinion in Otolaryngology and Head and Neck Surgery* 12: 374–77.

An International Comparison of Priority Setting for Orphan Drugs

Zahava R.S. Rosenberg-Yunger, Abdallah S. Daar, Halla Thorsteinsdóttir, and Douglas K. Martin

Abstract

Objectives: To describe the process of priority setting (PS) for two orphan drugs (OD)—Cerezyme and Fabrazyme—in three countries, and identify best practices in making coverage decisions for orphan drugs.

Methods: We conducted qualitative case studies of how three independent drug advisory committees made decisions relating to the funding of Cerezyme and Fabrazyme. PS was evaluated against a leading conceptual framework for legitimate and fair PS. Interviews were conducted with 22 informants, including committee members, patient groups, and industry representatives.

Results: (1) Description: OD reimbursement recommendations by expert panels were based on clinical evidence, cost, and cost-effectiveness analysis; and (2) Evaluation: Committee members expressed preference for the current process used by their own committee, but were concerned with the fairness of the process for ODs. Other informants suggested the inclusion of other relevant values (e.g., drug's life saving ability) in order to improve the PS process.

Conclusion: PS for drugs is not solely a technical process (involving cost-effective analysis, evidence-based medicine, etc.). PS of expensive orphan drugs illustrates how consideration of a wider range of relevant values in drug reimbursement will enhance the fairness of the PS process.

1. Introduction

Drug expenditures are rapidly increasing in every health system and account for a large proportion of health spending. Part of the reason for this increase is that the per patient costs of some new drugs are

extremely expensive, particularly for a number of orphan drugs, that is, drugs used to treat an illness affecting less than 1 person out of 1,000 (Lavandeira 2002). Currently, over 6,000 orphan or rare disorders have been identified (Clarke 2006).

Priority setting for orphan drugs involves complex value-laden choices that are often ethically controversial, in part because they directly involve a number of stakeholders, such as government officials, pharmaceutical companies, patients, and the public (who are ultimately paying for the drugs). Expensive orphan drugs present a challenge to many drug recommendation committees because these drugs seldom meet the cost-effectiveness and clinical evidence criteria commonly used to evaluate drugs under review for reimbursement. In particular, orphan drugs cannot undergo large clinical trials due to the small number of patients affected by the disease. This problem has a potentially enormous scope because, as the science of genomics advances, medical treatments are becoming more personalized, thereby gaining quasi-orphan status. Today's policy-making for a few orphan drugs may set the precedent for thousands of future products.

Cerezyme and Fabrazyme are two examples of expensive orphan drugs. These drugs were chosen as the cases because they are both innovative and extremely costly orphan drugs. The purpose of this study was to identify the values used by three international drug reimbursement recommendation committees regarding Cerezyme and Fabrazyme. To date, there have been few studies describing priority setting in the context of orphan drug reimbursement decisions. Describing and identifying the values involved in the process of drug reimbursement decisions within an international context is an essential initial step towards understanding and improving the process.

2. Methods

2.1 Design

We conducted qualitative case studies of priority setting of two drugs, Cerezyme and Fabrazyme, in three committees across three countries. Tables 1 and 2 provide more specific detail about each of the aforementioned drugs. Data collection involved semi-structured interviews with 22 committee members, patients, and manufacturers, plus the review of several relevant documents. The case study approach was appropriate because it allowed for an in-depth and

richer understanding of the drug reimbursement process for expensive biopharmaceuticals.

Table 1. Cerezyme (imiglucerase)

Manufacturer	Genzyme, approved by FDA (US) in 1994
Use	Reduces and in some cases reverses the chronic and debilitating symptoms of Type 1 Gaucher's disease
Cost	$350,0000 US per patient per year
Reimbursement Recommendation	Prior to establishment of CEDAC and IBC; PBAC recommended funding through the Life Saving Drugs Program (LSDP)
Research Studies	1.Replacement therapy for inherited enzyme deficiency—macrophage-targeted glucocerebrosidase for Gaucher's disease Clinical trial lasting 9 months of 12 patients with type 1 Gaucher's disease Safety and efficacy regarding improving haemoglobin levels and platelet counts and in reducing splenic and hepatic enlargement were demonstrated within 5 years (Barton, Brady, et al. 1991) 2. Enzyme therapy in type 1 Gaucher disease: comparative efficacy of mannose-terminated glucocerebrosidase from natural and recombinant sources Clinical trial comparing and demonstrating the safety and efficacy of imiglucerase with alglucerase (Grabowski, Barton, et al. 1995) 3. Replacement therapy with imiglucerase for type 1 Gaucher's disease Clinical trial comparing the frequency of administration of imiglucerase (Zimran, Elstein, et al. 1995)

2.2 Setting

Because the drugs Cerezyme and Fabrazyme were the cases, this research was conducted within multiple settings, that is, both reimbursement recommendation committees and drug manufacturers. The government committees included in this study were as follows: the Canadian Expert Drug Advisory Committee (CEDAC); Australia's Pharmaceutical Benefits Advisory Committee (PBAC); and the Israeli Basket Committee (IBC). These committees were selected because they are bodies that make decisions about public funds, and they

Table 2. Fabrazyme (agalsidase beta)

Manufacturer	Genzyme, approved by FDA in April 2003
Use	Treats Fabry disease, a potentially fatal lysosomal storage disorder
Cost	$300,000 USD per patient per year
Reimbursement Recommendation	CEDAC recommended against funding; IBC recommended funding; and PBAC recommended funding through the LSDP
Research Studies	1.Safety and efficacy of recombinant human α- galactosidase A replacement therapy in Fabry's disease Randomized, placebo controlled, double-blind study of 58 patients who were treated every 2 weeks (Eng et al. 2001) 2.Agalsidase-Beta therapy for advanced Fabry disease Randomized, double-blind, placebo controlled trial across 41 centres in 9 countries (Maryam et al. 2007) 3.Long-term therapy with agalsidase alfa for Fabry disease: safety and effects on renal function in a home infusion setting Single centre, prospective open label treatment trial in 25 adult male Fabry patients (Schiffmann, Ries, et al. 2006) 4.Long-term safety and efficacy of enzyme replacement therapy for Fabry disease 58 patients were enrolled in a Phase 3 double-blind, randomized, and placebo controlled trial (Wilcox and Banikazemi 2004) 5.Enzyme replacement therapy in Fabry disease: a randomized controlled trial Randomized control trial (double-blinded) (Raphael, Jeffrey, et al. 2001)

provide guidance on drug funding to governments and other funders. Tables 3, 4, and 5 below provide an overview of each of these committees. The company that manufactured both of the study drugs, Genzyme, was included in this research because the pharmaceutical industry is a stakeholder in drug reimbursement decisions.

2.3 Sampling

2.3.1 The Cases

The two drugs were selected because they are expensive, orphan

drugs. Cerezyme (imiglucerase) was developed using recombinant DNA from the enzyme glucosylceramidase, and is used to reduce and, in some cases, reverse the chronic and debilitating symptoms of Type 1 Gaucher's disease. Today, Cerezyme is used in more than 70 countries to treat Gaucher's disease (Genzyme website). The cost of Cerezyme has been estimated at $350,000 US per year per individual (Clarke, Amato, et al. 2001). Fabrazyme (agalsidase beta) was developed using recombinant human DNA from the enzyme α-galactosidase A to treat Fabry disease (Genzyme website). Fabry disease is a potentially fatal lysosomal storage disorder (Mignani and Cagnoli 2004) which affects an estimated 1 in 40,000 males. Today, Fabrazyme is approved for use in treating Fabry's disease in a number of countries, such as the 15 European Union countries, as well as Iceland, Norway, New Zealand, Australia, and Israel (Genzyme website). The cost of Fabrazyme has been estimated at $350,000 US per patient per year (Bengtsson, Johansson, et al. 2003).

Table 3. General Information about the Canadian Expert Drug Advisory Committee

Decision-making Criteria	Forms of Publicity	Appeals Process	Committee Composition	Mandate	Documents Reviewed
Safety Efficacy Therapeutic advantage (relative to current treatments) Cost-effectiveness (related to other treatments)	Internet	Available for manufacturers	Experts Lay members	To make drug listing recommendations to Drug Plans (based on submissions)	Recommendation for agalsidase beta. 2005

In Canada, Cerezyme was marketed prior to CEDAC. Drug funding decisions of Cerezyme (and all drugs not administered in the hospital setting) were and continue to be made provincially (see Appendix 1). Obtaining reimbursement required much negotiation between healthcare professionals, government, and public advocate appeals. For example, in Ontario, the Minister of Health rejected funding of Cerezyme because of the drug's inability to meet cost-effectiveness criteria (Clarke, Amato, et al. 2001). This decision was publicly

criticized by the National Gaucher Foundation of Canada in 1993. The Minister of Health subsequently applied the Rule of Rescue and approved a provincial program for reimbursement of enzyme replacement therapy (ERT) for Gaucher's disease (ibid.).

Table 4. General Information about the Australian Pharmaceutical Benefits Advisory Committee

Decision-making Criteria	Forms of Publicity	Appeals Process	Committee Composition	Mandate	Documents Reviewed
Effectiveness (compared to alternative therapies) Cost (compared to alternative therapies)	Internet	Available for manufacturers	Experts Lay members	To make drug listing recommendations and give advice to the Minister	Guidelines for the Treatment of Gaucher Disease Guidelines for Eligibility to Receive Treatment with Agalsidase through the Life Saving Drugs Program

Table 5. General Information about the Israeli Basket Committee

Decision-making Criteria	Forms of Publicity	Appeals Process	Committee Composition	Mandate	Documents Reviewed
Clinical evidence Economic evidence Social implications Ethical implications Legal implications	Internet Radio Newspaper	Resubmit in the new year	Experts, lay people, members of the ministry of finance, members of the health insurance	To make drug listing recommendations to the Cabinet (Shani, Siebzhner, et al. 2000)	Fabrazyme recommendation Guidelines for the submission of a request to include a pharmaceutical product in the national list of health services

In Australia, drug funding decisions are made nationally through PBAC. Cerezyme was made available through the Life Saving Drugs Program. The Therapeutics Goods Administration (TGA) has published "Guidelines for the Treatment of Gaucher Disease."

In Israel, the Ministry of Health rejected funding Cerezyme on the basis of its inability to meet the cost-effectiveness criteria. In 1995, the

ministry included Cerezyme externally from the basket in a New Health Bill which gave special funding to chronic diseases, including Gaucher's disease. This decision to fund Cerezyme was based on a reduction in the cost of the drug. Israeli researchers determined that a lowered dose (without negative effects) would reduce the cost to 25% compared to the cost of the manufacturer's recommended dose (Gross 2002).

With regards to Fabrazyme, in Canada it was reviewed by CEDAC, which recommended against the funding of this drug. However, each province must make its own individual formulary decision.

In Australia, drug funding decisions are made nationally through PBAC. Fabrazyme was available through the Life Saving Drugs Program. In Israel, Fabrazyme was recommended by the IBC for the inclusion in the basket and has been part of the basket since 2002 (Kesselman, Elstein, et al. 2006).

2.3.2 The Participants

Interview participants for this study were key informants who were selected based on their experience with the drug decisions in question. This method is appropriate for in-depth studies of issues within their natural settings rather than in artificial isolation (Maxwell 1996).

Twenty-two interviews were conducted with members of advisory committees (CEDAC [4], PBAC [3], and IBC [4]), representatives of drug companies (4), and patient groups (7 respondents from Canada). Initial contact was made with individuals (i.e., advisory board members, patient groups, and industry) either in person or by email or phone. If a response was not obtained, two more attempts were made. Snowball sampling was also used, that is, participants were asked to suggest other potential interviewees. Sampling continued until the analysis reached saturation, that is, there was reiteration of the same ideas (Strauss and Corbin 1998). There was no formal calculation of sample size.

2.3.3 The Documents

Five documents and three websites related to orphan drug reimbursement decisions were also analyzed in order to explore issues of fairness related to reimbursement decisions. Documents were obtained mainly in electric format from committee and patient group websites. However, a number of documents were not publicly

accessible (particularly in Israel) and were obtained through formal letters of request to the agency.

2.4 Data Collection

Data collection involved in-depth qualitative interviews, plus the collection of relevant documents. We conducted face-to-face interviews or one-on-one telephone interviews. Interviews were 30–60 minutes in length. All interviews were audiotaped and transcribed. Interviews explored decision making in drug reimbursement of the two selected drugs (see Appendix 2: Example of Interview Guide for Committee Members).

Relevant documents related to reimbursement decisions were sampled and analyzed in order to explore issues of fairness related to reimbursement decisions surrounding both the study drugs.

2.5 Data Analysis

The interviews and documents were analyzed using a modified thematic analysis. First, the data were read to achieve a good working knowledge of the content—sometimes called "immersion" (Denzin and Lincoln 2005). Second, portions of data that related to similar concepts or ideas were identified and labeled, that is, open coding (Maxwell 1996). For example, the ideas that related to accessibility, such as the ability of the public to review recommendations, were labeled as "access." Third, concepts were compared between and within transcripts and documents to ensure consistency and comprehensiveness. Inconsistencies were corrected through either re-coding data portions into more appropriate codes or being identified as areas of further analysis. Fourth, axial coding was used to identify and organize overarching themes. Fifth, primary themes were established and related to the other themes.

During each step, analytic memos were written on observations (Creswell 1998). For example, there were comments on the location, the manner of the respondent, and the way the interview progressed. Memos are an important part of research and allow the researchers to reflect on and analyze the research methods and findings (Maxwell 1996).

The issue of validity was addressed in three ways. First, different data sources were used, including literature, documents, and interviews, which allowed for a triangulation of sources in developing emerging concepts (Denzin and Lincoln 2005). For example, the issue of insufficient data for making a reimbursement decision arose in the

interviews. This notion was further supported in a recent study by Gallego et al. (2007). Second, codes and themes were developed with other team members as a check on bias. We frequently discussed themes and codes with the research team. Third, findings were introduced to an interdisciplinary group of scholars for feedback to help ensure reasonableness of findings. Specifically, three interim analysis meetings were held with a large interdisciplinary group of scholars, including faculty members, research fellows, and doctoral students. These meeting provided an opportunity to discuss and explain the rationale behind the codes. While consensus was achieved for most of the codes, some concepts were coded under different themes as a result of the discussions during these sessions.

2.6 Research Ethics

This project was approved by the University of Toronto's Human Subject Review Committee. The consent form, along with a description of the research, was sent via email to participants prior to the interview. The consent form was reviewed with each participant at the onset of the interview and all questions and concerns were addressed. Before the start of in-person interviews, consent forms were signed and a copy was given to the participant. When interviews were conducted by telephone, the signed consent form was either faxed or sent electronically. All respondents agreed to participate and written informed consent was obtained prior to the interview. All data are confidential and anonymity of participants is protected. Additionally, all raw data are protected and available only to the research team.

3. Results

Our main finding was that participants from three reimbursement committees, across the three different health systems, reported using essentially the same values when making reimbursement recommendations for the orphan drugs Cerezyme and Fabrazyme. Those values were evidence (as assessed through cost-effectiveness and effectiveness), life saving ability (i.e., application of the Rule of Rescue), and equity. Tables 6 and 7 below compare the values used by each committee and their evaluation of whether the drug passed or failed the particular value. Please note that in Table 6, the committee names of CEDAC and IBC are omitted as these committees had not been established. Moreover, despite the drugs' inability to meet all of the values used by the three committees, all three jurisdictions have

funded the drugs either through the general funding mechanism (e.g., Israel's inclusion of Fabrazyme in the Health Basket) or through alternative mechanisms (e.g., Australia's inclusion of both drugs in the Life Saving Drugs Program). Table 8 below indicates the mechanism by which each drug was funded in each country.

Table 6. Values Used in Cerezyme Recommendations

Values Used	Canada's Ontario Ministry of Health	Israel's Federal Ministry of Health	Australia's PBAC
Evidence • Cost-effectiveness • Effectiveness	 Failed N/A	 Initially failed Passed	 Failed Passed
Rule of Rescue	Passed	Passed	Passed
Equity	N/A	N/A	Passed
Final Funding Outcome	Funded	Funded	Funded

Table 7. Values Used in Fabrazyme Recommendations

Values Used	CEDAC	IBC	PBAC
Evidence • Cost-effectiveness • Effectiveness	 Failed Failed	 Passed Passed	 Failed Passed
Rule of Rescue	N/A	Passed	Passed
Equity	Failed	N/A	Passed
Final Funding Outcome	Post-market study	Funded	Funded

Table 8. Availability of Drug by Country

Country	Cerezyme	Fabrazyme
Australia	Life Saving Drugs Program	Life Saving Drugs Program
Canada	Varies by province (See Appendix 1)	Post-market study
Israel	Available since 1995 through New Health Bill for funding of chronic diseases	Included in the Health Basket since 2004

4. Evidence

4.1 Cost-effectiveness

All the recommendation committees in this study placed a high value on clinical evidence. One concern a number of committee members

raised, particularly as it related to orphan drugs, was the lack of good clinical evidence of the drug's cost-effectiveness. For example, one member commented: "A major issue, I think, internationally [is] not only the high cost of some of these agents but, the lack of data upon which to make a proper judgment of their cost effectiveness."

In Canada, CEDAC's recommendation in 2004 against the funding of Fabrazyme was based on the lack of evidence regarding effectiveness: "This trial failed to show a clinical benefit of aglasidase beta on a range of tests" (Canadian Expert Drug Advisory Committee 2005). One CEDAC committee member explained the recommendation against the funding of Fabrazyme as related to effectiveness: "I mean, people just continued to progress on the medication, the disease progresses and this isn't a cure and it was hard to justify spending $300,000 dollars on a medication that is relatively effective at some end points but not effective at others."

Similarly, Israel's Ministry of Health initially rejected Cerezyme for funding because of the drug's inability to meet cost-effectiveness criteria. However, the Israeli government was willing to provide the treatment if costs could be reduced. Ultimately, Cerezyme was funded in 1995 because Israeli researchers determined that lower doses without negative effects (< one quarter of the manufacturer's recommended dose) would reduce the cost significantly, saving $80,000,000 (Gross 2002).

The application of cost-effectiveness criteria to orphan drugs was recognized as problematic by many committee members, as one PBAC member explained: "I don't regard those as being expensive drugs. I just regard those as being ridiculously expensive drugs. So they would never be cost-effective…in the paradigm."

Interestingly, a CEDAC committee member articulated that it was unclear that orphan drugs should be prioritized differently from other types of drugs: "We didn't have a separate process for reviewing rare drugs and, you know, no one had told us that we needed to prioritize drugs for rare conditions differently than we prioritize all drugs."

Patients believed that the lack of clinical evidence should not be an insurmountable barrier in a committee's decision: "Some of these questions are typically asked or typically answered in large phase four studies…Things like clinical significance and statistical

significance…these are very important factors, but not the only questions to ask. In particular, when trying to resolve issues around treatments for patients with rare disorders."

Industry recognized the high cost of the drugs, but thought that governments are reframing the issues in terms of cost, as one representative noted: "I think they're good products…But if you're spending 200 million dollars to treat heartburn, you can spend a couple of million dollars to provide a drug that potentially could save someone's life or prolong their life."

4.2 Effectiveness

A drug's ability to provide therapeutic benefit was considered by many respondents to be an important decision-making value. But, from one patient's perspective, increased recognition of Fabrazyme as a preventative therapy would aid in the decision-making process: "What is really frustrating for Fabry's patients…is that the world basically is now administering enzyme replacement therapy for Fabry's disease as a preventative therapy prior to organ failure…But yet, in Canada the protocol that we have to deal with is the most stringent in the world."

4.3 Rule of Rescue (ROR)

The Rule of Rescue (ROR) is a principle which values rescuing an endangered life. The categorization of drugs as life saving is an application of the Rule of Rescue. A drug's ability to be a life saving treatment was a value considered by all of the committees, some more formally than others. Patients believed that a drug's ability to save a life should be a criterion in decision making.

Saving a life was a value used formally by the IBC. As one member of the IBC explained: "You have to implement other ethical values, legal, and decide—what are the priorities?…It's our culture, Judaism…we are very concerned about life, about health."

Furthermore, the IBC prioritized life saving treatments. As one Israeli respondent commented: "The life saving drugs…will get a higher rank…[and therefore] will be provided in the basket." Life saving ability or the application of the Rule of Rescue was not clearly formulated as part of the CEDAC process. One CEDAC committee member explained: "I guess there's a distinction there that the life saving drugs could get a priority review and that would mean that they would be reviewed a little more quickly and brought to the

committee a little more quickly. The actual type of information that is sought for each medication is similar…there's other considerations that would go in as well, whether it's a specific drug for a condition that just improves quality of life or only improves life expectancy, those types of things are considered but not in a formulaic approach or anything."

In Australia, PBAC recognized the inability for both Cerezyme and Fabrazyme to meet their cost-effective criteria. Consequently, these drugs are available through the Life Saving Drugs Program, which provides financial assistance for drugs that treat rare, inherited enzyme deficiencies. As one committee member explained: "We have a rule of rescue…so it's a condition for which there's no other therapy available and yet there's still demonstrable suffering from the disease and we may list the condition…I mean we still have…we need to inform the Minister of the consequences of that, the financial consequences, but we may not be able to apply as rigorously a cost effectiveness analysis to a small…to a group of patients with a rare disease."

Patients believe that a drug's life saving ability should be a consideration in reimbursement decisions. One patient-respondent discussed the approach they would like to see used in such decisions,

> It's sort of like the hospital…One doesn't come into a neonatal ward and say we're going to put the child on life support, but you know what, when his bill begins to go over a certain amount than we have to pull the plug…we leave the child on, until it becomes clear that the child's either going to survive or [there's] not going to be a benefit and if the child's not going to benefit then that's fine…I think that's the kind of approach we're trying to do here. In many cases these are life saving treatments we're talking about diseases for which no other treatment available, not even other types of interventions that one would make. So it is the case that we either have the drug or we have nothing.

4.4 Equity (of Access)

Equity of access was a value used by some committees and discussed by a number of patient and industry respondents. Patients discussed their experience in accessing their particular drug and their use of advocacy to gain access to drugs. As well, patients discussed variations in access across and within countries. Patients also

discussed access in terms of their ability or inability to access the reimbursement decision-making process.

Access was not a criterion typically used by committees when making recommendations. Equity was cited by CEDAC in their rationale for their decision against reimbursement of Fabrazyme (Canadian Expert Drug Advisory Committee 2005). Additionally, a CEDAC committee member explained their conception of equity: "So, it was difficult to justify how we could say yes to that and no to, you know, medications for a more common condition. I mean, that has some equity issues as well, that you fund an expensive medication for a person with a rare disease who might get the same benefit as a less expensive for a common condition but you haven't funded that because it has much bigger budget implications."

One industry respondent, when asked about the Fabrazyme federal-provincial-territorial joint "research" protocol with industry in Canada, discussed the issue of equity in access: "I think, part of the chassis for the agreement had to do with recognition that the distribution of patients was not equal across the populations of the various provinces. So, that led to the idea that there needs to be some kind of national solution, because there was no way realistically to expect a small province like, Nova Scotia, to really be able to support the very high number of patients with that rare disease in relation to their population."

Patients also discussed variation in access of drugs across countries. For example: "It [Fabrazyme] was already made available to patients in 40 other countries, many of which are, you know, considered not developed countries…countries like Argentina and Turkey and Bulgaria. So we didn't think it would be a big issue but we found there were a number of obstacles to getting access."

5. Synthesis

There are a number of common values which emerged from the discussions with stakeholders, most notably evidence, the Rule of Rescue and equity. The participants' views regarding values were supported by the committees' assessment mandates. That is, all three values identified—evidence, Rule of Rescue, and equity—were values stated on the committees' websites and/or related documents as decision-making criteria. However, CEDAC's recommendation regarding Fabrazyme indicates (in addition to some other values)

equity reasons as part of their rationale. Equity is not mentioned on CEDAC's website as a decision-making criterion. Additionally, the IBC listed a number of decision-making criteria which were not discussed by many respondents, particularly, legal considerations.

The inability of orphan drugs to meet the cost-effectiveness criterion was problematic for both the Canadian and Australian systems, which clearly weight this criterion heavily. Canada's drug priority setting system is predisposed against funding drugs which do not meet this criterion. In Australia, drugs which the PBAC considers clinically effective but fail to meet cost-effectiveness standards are made available through a different route: the Life Saving Drugs Program. Israel's Basket Committee weighted the value of life more heavily, and it occasionally makes positive funding recommendations for drugs with an undesirable cost-effectiveness ratio.

6. Discussion

In this paper, we have described the values used by drug reimbursement recommendation committees in three countries pertaining to two expensive orphan drugs, Cerezyme and Fabrazyme. Our main finding was that participants from three different priority setting committees, working in three different health systems, and from three very different cultures reported using essentially the same values when making reimbursement recommendations for the orphan drugs Cerezyme and Fabrazyme. Those values were evidence, rule of rescue, and equity.

The similarity in values used across the systems provides evidence about the global dominance of a particular approach—economic—to drug priority setting. During the 1990s, there was increasing interest in and use of economic assessments of new therapies and explicit rationing in decision making (Dean 1991; Klein 1993; Coast 1997; Sabin 1998). Even though countries continue to use economic assessments, problems remain. Emphasis on meeting economic criteria such as cost-effectiveness places the value of efficacy above other values which are also important in decision making. Evidence-based medicine (EBM) is another popularly used tool which is used to understand effectiveness. However, it does not weigh effectiveness against other values (i.e., benefits, costs, etc.). Limitations of the economic approach include the inability to place a numeric value on

a health outcome. In addition, benefits and costs are subjective and therefore dependent on the person conducting the evaluation. Limitations of EBM include the frequent lack of sufficient evidence to make decisions (Martin and Singer 2000). Thus, it is not surprising that committees who use EBM and are adept at assessing a drug's ability to meet cost-effectiveness are challenged when a drug does not lend itself to such an assessment.

Nonetheless, drug priority setting is not solely a technical process. At its core, it involves adjudicating between and among a wide range of relevant values (Martin, Singer, et al. 2001). As Gallego et al. recently indicated, priority setting for high cost drugs is often based on other factors in addition to effectiveness and cost (2007). Thus, the differences among committees with regards to their application of the Rule of Rescue (ROR), that is, valuing a life saving treatment, are fascinating.

The ROR was used and applied by Canada, Australia, and Israel. Both the Australian and Israeli committees considered the ROR as part of their process. Australia considered three factors when considering whether to reimburse a drug through the LSDP: (1) whether an alternative exists; (2) whether the medical condition is severe, progressive, and expected to lead to premature death; and (3) whether the medical condition affects only a very small number of patients. In Israel, the IBC, as part of their formulary listing process, considered the following: (1) life saving technology with full improvement; (2) potential of technology to prevent mortality/morbidity; and (3) new technology for serious disease with no alternative treatment.

The tension between the ROR and cost-effectiveness is best demonstrated through the Canadian example. Ontario's Ministry of Health applied the ROR, recognizing that saving a life takes precedence over cost considerations after Cerezyme failed the latter criterion. Alternatively, CEDAC did not consider the ROR for Fabrazyme, and it was subsequently not recommended for funding as it did not meet cost-effectiveness criteria. CEDAC clearly considered issues of efficiency over those of saving a life.

Despite the inability of orphan drugs to meet the cost-effectiveness criteria, and differences in the application of the ROR, a number of countries are publicly funding these drugs through special drug access programs or by considering a fuller range of values (e.g.,

social and ethical impacts, etc.). This may be because committees recognize the value of saving a life over that of cost to the system.

Priority setting for orphan drugs involves deliberation about values, many of which conflict or are not quantifiable. Priority setting committees are very proficient at identifying quantifiable criteria, but struggle with other non-quantifiable values—such as the ROR. Israel has tried to develop a more inclusive strategy for making decisions. In addition to cost, Israel considers life saving ability and prevention of mortality (for a more detailed account, see Shani, Siebzhner, et al. 2000). It is necessary to create a drug listing system which is able to formally assess these non-quantifiable values in order to establish consistency among drug reimbursement decisions within health systems.

7. Conclusion

Drug funding decisions which provide some benefit to only some patients is highly contentious and morally controversial. It is clear that priority setting decisions will need to be made about which orphan drugs to reimburse, how to regulate them, and who will have access to them. Describing and evaluating decision making in specific contexts, such as in Canada, Australia, and Israel, and for two orphan drugs, Fabrazyme and Cerezyme, is the first step toward improving drug priority setting. This study has demonstrated that in order to create a fair and legitimate drug reimbursement process we need to ensure the incorporation of a wide range of values within the process.

Acknowledgements

Thank you to all of the participants.

This research is supported by the Canadian Priority Setting Research Network from an Interdisciplinary Capacity Enhancement grant from the Canadian Institutes of Health Research. Zahava R.S. Rosenberg-Yunger is supported by a Canadian Priority Setting Research Network doctoral fellowship. Abdallah S. Daar is with the Program on Life Sciences, Ethics and Policy (PLEP) of the McLaughlin-Rotman Centre for Global Health, University Health Network and University of Toronto. PLEP receives most of its funding from Genome Canada through the Ontario Genomics Institute, the Ontario Research Fund, and the Bill and Melinda Gates Foundation. A full list of funders is available online at

http://www.mrcglobal.org/about/funding. He is also supported by the McLaughlin Centre for Molecular Medicine at the University of Toronto. Halla Thorsteindóttir is supported by the Canadian Institutes of Health Research, New Principal Investigator Award. Douglas K. Martin is supported by a New Investigator Award from the Canadian Institutes for Health Research.

References

Barton, N.W., R.O. Brady, et al. 1991. Replacement Therapy for Inherited Enzyme Deficiency—Macrophage-targeted Glucocerebrosidase for Gaucher's Disease. *New England Journal of Medicine* 324 (21): 1464–70.

Bengtsson, B.-Å., J.-O. Johansson, et al. 2003. Enzyme Replacement in Anderson-Fabry Disease. *Lancet* 361 (9354): 352.

Canadian Expert Drug Advisory Committee. 2005. CDR Drug Database Listing on Agalsidase Beta. Available at http://www.cadth.ca/index.php/en/cdr/search/?&status=all&keywords=agalsidase%20beta (accessed 17 November 2007).

———. Available at http://www.cadth.ca/index.php/en/cdr/committees/cedac (accessed 12 March 2008).

Clarke, J.T.R. 2006. Is the Current Approach to Reviewing New Drugs Condemning the Victims of Rare Diseases to Death? A Call for a National Orphan Drug Review Policy. *Canadian Medical Association Journal* 174 (2): 189–90.

Clarke, J.T.R., D. Amato, et al. 2001. Managing Public Payment for High Cost, High-benefit Therapy for Gaucher's Disease in Ontario. *Canadian Medical Association Journal* 165 (5): 595–96.

Coast, J. 1997. The Case Against. *British Medical Journal* 314 (7087): 1118–23.

Creswell, J.W. 1998. *Qualitative Inquiry and Research Design: Choosing among Five Traditions*. London: Sage Publications.

Dean, M. 1991. Is Your Treatment Economic, Effective, Efficient? *Lancet* 337 (8739): 480–81.

Denzin, N.K., and Y.S. Lincoln, eds. 2005. *The Sage Handbook of Qualitative Research*, 3rd ed. London: Sage Publications.

Eng, C.M., et al. 2001. Safety and Efficacy of Recombinant Human α- Galactosidase A Replacement Therapy in Fabry's Disease. *New England Journal of Medicine* 345 (1): 9–17.

Gallego, G., S.J. Taylor, et al. 2007. Priority Setting for High Cost Medications (HCMs) in Public Hospitals in Australia: A Case Study. *Health Policy* 84 (1): 58–66.

Genzyme. Available at http://www.genzyme.ca/thera/fz/ca_en_p_tp_thera-fz.asp (accessed 17 February 2005).

Government of Australia. Alternative Arrangements for Medicines: Life Saving Drugs. Available at http://www.health.gov.au/internet/main/publishing.nsf/Content/health-pbs-general-supply-othersupply (accessed 12 March 2008).

———. Guidelines for the Treatment of Gaucher Disease. Available at http://www.health.gov.au/internet/wcms/publishing.nsf/Content/71D15BD1EE4A4A8DCA256F85007A7F66/$File/gguidelines.pdf (accessed 14 August 2007).

———. Pharmaceutical Benefits Advisory Committee. Available at http://www.health.gov.au/internet/main/publishing.nsf/Content/PBAC-Meeting-Agenda-and-Consumer-Comments-lp (accessed 31 December 2008).

Grabowski, G., N. Barton, et al. 1995. Enzyme Therapy in Type 1 Gaucher Disease: Comparative Efficacy of Mannose-terminated Glucocerebrosidase from Natural and Recombinant Sources. *Annals of Internal Medicine* 122 (1): 33–39.

Gross, M. 2002. Ethics, Policy, and Rare Genetic Disorders: The Case of Gaucher Disease in Israel. *Theoretical Medicine and Bioethics* 23 (2): 151–70.

Kesselman, I., D. Elstein, et al. 2006. National Health Budgets for Expensive Orphan Drugs: Gaucher Disease in Israel as a Model. *Blood Cells, Molecules, and Diseases* 37 (1): 46–49.

Klein, R. 1993. Dimensions of Rationing: Who Should Do What? *British Medical Journal* 307 (6899): 309–11.

Lavandeira, A. 2002. Orphan Drugs: Legal Aspects, Current Situation. *Hemophilia* 8 (3): 194.

Martin, D.K., and P.A. Singer. 2000. Priority Setting and Health Technology Assessment: Beyond Evidence-based Medicine and Cost-effective Analysis. In *The Global Challenge of Health Care Rationing*, edited by A. Coulter and C. Ham, 133–35. Philadelphia: Open University Press.

Martin, D.K., P.A. Singer, et al. 2001. Priority-setting Decisions for New Cancer Drugs: A Qualitative Case Study. *Lancet* 358 (9294): 1676–81.

Maryam, B., et al. 2007. Agalsidase-Beta Therapy for Advanced Fabry Disease. *Annals of Internal Medicine* 146 (2): 77–86.

Maxwell, J.A. 1996. *Qualitative Research Design: An Interactive Approach*. London: Sage Publications.

Mignani, R., and L. Cagnoli. 2004. Enzyme Replacement Therapy in Fabry's Disease: Recent Advances and Clinical Applications. *Journal of Nephrology* 17: 354–63.

Raphael, S., K. Jeffrey, et al. 2001. Enzyme Replacement Therapy in Fabry Disease: A Randomized Controlled Trial. *Journal of the American Medical Association* 285 (21): 2743–49.

Sabin, J. 1998. Fairness as a Problem of Love and the Heart: A Clinician's Perspective on Priority Setting in Health Care. *British Medical Journal* 317 (7164): 1000–1007.

Schiffmann, R., M. Ries, et al. 2006. Long-term Therapy with Agalsidase Alfa for Fabry Disease: Safety and Effects on Renal Function in a Home Infusion Setting. *Nephrology Dialysis Transplantation* 21 (2): 345–54.

Shani, S., M.I. Siebzhner, et al. 2000. Setting Priorities for the Adoption of Health Technologies on a National Level—the Israeli Experience. *Health Policy* 54 (3): 169–85.

Strauss, A., and J. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd ed. Thousand Oaks: Sage Publications.

Wilcox, W.R., and M. Banikazemi. 2004. Long-Term Safety and Efficacy of Enzyme Replacement Therapy for Fabry Disease. *American Journal of Human Genetics* 75 (1): 65–74.

Zimran, A., D. Elstein, et al. 1995. Replacement Therapy with Imiglucerase for Type 1 Gaucher's Disease. *Lancet* 345 (8963): 1479–80.

Appendix 1. Availability of Cerezyme and Fabrazyme in Canada

Formulary	Cerezyme	Fabrazyme		
	Listed on Formulary	Other Means of Access	Listed on Formulary	Other Means of Access
Alberta Health & Wellness	No	Special authorization process	No	Post market study
BC Pharmacare	No	No exception drug status	No	Under review
Manitoba Pharmacare	No	Exception drug status benefit	No	Post market study
New Brunswick Prescription Drug Program	No	Special authorization	No	Post market study
Newfoundland & Labrador Prescription Drug Program	No	Special authorization process may cover it	No	No exceptions yet
Northwest Territories Health Benefits Program	No	Exception drug coverage	No	Post market study
Nova Scotia Pharmacare	No	Exception drug status process	No	Exception drug status process
Nunavut Health Benefit Program	No	Exceptional circumstances	No	Exceptional circumstances
Ontario Drug Benefit Program	No	Special drug program	No	Post market study
PEI Drug Cost Assistance Program	No	High cost drug program	No	High cost drug program
Régie de l'assurance maladie du Québec	No	Can appeal with process led by referring physician	No	Can appeal with process led by referring physician
Saskatchewan Drug Plan	No	Not even by exception	No	Not even by exception
Yukon Pharmacare	No	Will only be reviewed on needs basis	No	Reviewed on needs basis

Appendix 2. Interview Guide for Committe Members

Sample Questions
Recommendation Committees: 1. How are recommendations made regarding funding of expensive orphan drugs? 2. Is there a distinction made between drugs which are life saving, orphan, and/or QOL? 3. Who was involved in the decision-making process and what was the extent of their involvement? Specifically, which stakeholders were involved? 4. What considerations do you feel are important in making the reimbursement recommendations for orphan drugs? 5. How were recommendations communicated/disseminated? 6. What happens if someone disagrees with a decision? Is there a formal process that people including the general public can challenge the decision? 7. Do you think it is a fair process? 8. What do you think could be done to improve this process? 9. How do you think the process affects innovation in the area of biotechnology? 10. Do you think priority setting affects innovation in the area of biotechnology? How?

Equity in Prescription Drug Use among Older Ontarians

Sara Allin

Abstract

This paper focuses on equity for older Ontarians in the context of the province's public prescription drug program. Its aims are twofold: (1) to compare two sources of information on prescription drug utilization, claims, and survey data; and (2) to assess the extent of inequity in the use of medicines. The analyses draw on the Canadian Community Health Survey (CCHS) that is linked to pharmacy claims data from the Ontario Drug Benefit (ODB) database. There appears to be relatively limited concordance of reported medicine use and prescription claims data, which may relate to the phrasing of the questions on medicine use in the CCHS and to non-adherence. I find little evidence of inequity in the population of older Ontarians: poorer people appear to use more drugs than those with higher income, even after adjusting for health status and demographics. These findings are discussed in the context of the data limitations and the implications for future research.

Introduction

Prescription drugs play an increasingly important role in the prevention and treatment of disease. However, the financing of prescriptions drugs in Canada differs from that of most other medical services. Ambulatory prescription drugs are excluded from the protections of universal access to care granted by the Canada Health Act. As a result of this exclusion, provincial coverage for ambulatory prescription drugs and the amount that patients need to pay to access these drugs varies across the country, depending on the structure of the provincial plans (Demers et al. 2008). The lack of universal coverage and the fact that patients are often required to cover some of the cost of these drugs raise a number of concerns about equity of access in each province. Like most other provincial pharmaceutical programs, the Ontario Drug Benefit (ODB) program does not provide

universal coverage: comprehensive coverage is available to the population groups who are in greatest need and have least ability to pay. While it is clear then that there is inequitable access to public coverage between people who are eligible for ODB coverage and those who are not, the question whether inequities in access still exist for those who have public coverage for prescription drugs has not been investigated.

This paper has two main aims. The first is to assess the extent of inequity in pharmaceutical usage for people eligible for ODB coverage. Although the program provides coverage to a number of populations, this study focuses on the population group that makes up the majority of the ODB program recipients: individuals aged 65 years and older.[1] For the purpose of this research, equity is defined as the receipt of services or medicines on the basis of need, and the payment of services on the basis of ability to pay (Evans 1983). Need is determined on the basis of the expectation of protecting, promoting, or restoring health (Birch and Abelson 1993; Culyer and Wagstaff 1993; Evans 1983; Evans 1992; Giacomini et al. 2004). Inequity can be assessed empirically using the concentration curve, which indicates how concentrated the use of health services (or expenditure) is along the distribution of income in a population (O'Donnell et al. 2008). Inequity in utilization would arise if the use of medicines is concentrated among higher- or lower-income groups, and inequity in finance would be indicated by a concentration of out-of-pocket payments among those with lower income.

The article's second aim is to examine the comparability of two data sources on prescription drug utilization data available in Ontario for the purpose of evaluating equity in prescription drug usage. Data on prescription drug utilization is available from community surveys and administrative data sources. Each of these data sources offers advantages and disadvantages as a means of evaluating equity in drug utilization. There is also the possibility to link these two data sources at the individual level. In order to address the first aim of the paper, the researcher needs to first determine which data source is the most appropriate for evaluating equity in the public drug program.

This paper is organized in three parts. First, I provide a brief introduction to the financing of prescription drugs, with a focus on the public drug program in Ontario. Second, I describe the analysis of comparability of survey and administrative sources of medicine use.

Third, I outline an assessment of inequity in the use of medicines using linked survey and administrative data. The paper concludes with a summary of the findings along with some suggestions for future research.

Financing Prescription Drugs

The role of prescription drugs in the health system has increased markedly in the past 20 years. For example, as a proportion of total health spending in Canada, pharmaceuticals constituted 9.5% in 1985, compared to 17% in 2007 (Canadian Institute for Health Information 2008). Pharmaceuticals currently represent the largest category of health spending in Canada after hospitals (Canadian Institute for Health Information 2008). The rising total cost for prescription drugs has been attributed to increased utilization (accounting for over half of the rise in spending) in addition to changes in therapeutic choice, and less so to increases in drug prices (Morgan 2004).

Pharmaceuticals can also represent significant costs to individuals. Even though the majority of Canadians have some form of insurance for prescription drugs, the average Canadian family is estimated to spend over $1200 out-of-pocket (on top of any drug insurance premium) per year on prescription drugs (Commission on the Future of Health Care in Canada 2002). The costs of medicines not only create a financial burden for poorer households, but they may also present barriers to access. There is extensive evidence demonstrating that people are sensitive to the price of medicines (Gemmill et al. 2008; Lexchin and Grootendorst 2004). To protect population groups who are less able to pay, provincial public drug insurance programs in Canada provide coverage for defined populations, primarily those aged 65 or over and those receiving social assistance (although income-based programs for which the entire population is eligible are in place in some provinces). These income- and age-based programs have been enacted to protect those without employer-based prescription drug insurance, coverage that exists for the majority of the working-age population. In light of the extensive evidence that individuals are sensitive to the price of medicines, public drug programs aim to ensure equitable access to medicines and a more equitable distribution of the burden of costs (Evans 2005).

An example of such a program is the Ontario Drug Benefit (ODB) program, which funds about half of the total cost of prescription

medications in Ontario. It covers about 2.8 million Ontario residents at an annual cost of approximately $3.8 billion. Eligible beneficiaries of the program include individuals aged 65 and over (this group makes up the majority of ODB program recipients), those on social assistance, families with high drug costs relative to their income (covered by the Trillium Program), residents of long-term care facilities, and recipients of home care.

About two-thirds of the recipients of the ODB program are individuals who are aged 65 years or older. Within this group there are two programs with different cost-sharing arrangements. The default category that an individual enters upon turning 65 is the "high-income senior" program, which is associated with a $100 annual deductible and a $6.11 copayment per drug that is dispensed. If an individual provides documentation of low-income status,[2] he or she becomes eligible for the low-income senior program, which has no deductible and may include a $2 copayment per drug that is dispensed. The relatively low threshold for eligibility for the low-income senior category raises some concern over equity whereby the out-of-pocket payments may deter the appropriate use of medicines among this low-income group.

Among those aged 65 years and older, some with higher income may also hold private prescription drug coverage. The ODB program is the payer of first resort; private insurers cover part of the costs of drugs that are not included in the ODB formulary as well as the ODB deductible (Paterson et al. 2008). The proportion of the population with such supplementary insurance is not known precisely, but is estimated to be around 20% of those aged 65 years and older (Paterson et al. 2008). Although the aim of the ODB program is to allow for more equitable drug access across the provincial population, there has been relatively scant attention paid to evaluating the performance of the ODB program, in particular, the extent to which it achieves an equitable distribution of medicines and an equitable distribution of the burden of payment for the population for which it provides coverage.

Data on Prescription Drug Use: Assessing Comparability

Investigations of equity in the pharmaceutical sector require information on medicine use at the individual level. Two of the most common sources of information on medicine use are survey data and prescription drug claims data. The few studies that have examined the non-need correlates of medicine use have relied on survey data

(Ballantyne et al. 2005; Grootendorst 1995; Zhong 2007). The comparison of self-reported use of hospital and physician services with administrative data has been extensive and shows relatively high comparability (Cleary 1984; Glandon et al. 1992; Jobe et al. 1990; Marquis et al. 1976; Roberts et al. 1996), even in the 65-years-and-older population (Raina et al. 2002; Wallihan et al. 1999). There may be differences in how an individual interprets a health care contact and how that contact is recorded in administrative databases, such as in the classification of a hospital admission versus an outpatient visit. However, a great number of difficulties arise in the efforts to compare different data sources of prescription drug utilization.

The main difference between survey data of medicine use and pharmacy claims data is that the former measures drugs that are actually consumed by the patient, whereas the latter measures drugs that are dispensed. There are many reasons why a drug may not be consumed after it has been dispensed. One reason is that the drug may be prescribed on an "as needed" basis and the patient has not yet needed the drug. Another is that the patient may not adhere to the treatment plan. Non-adherence could arise for numerous reasons: the patient may not remember to take the drug; he may start taking the drug but then discontinue use because the symptoms are reduced or relieved, or because he experiences side effects; or she may decide not to take the drug for other reasons (for instance, she may give it or sell it to another person).

Although surveys are designed to gather information on the medicines that have been taken by the respondent, the ability of survey respondents to accurately recall their medicine use varies according to the design and implementation of the survey (Gama, Correia, and Lunet 2009). Surveys that ask respondents about the details of the drugs that they are taking, such as the names, dosage, etc., for example by checking in their medicine cabinet, show high comparability with pharmacy claims data (Johnson and Vollmer 1991; Klungel et al. 2000). However, surveys that include questions that are more open ended or ask respondents to estimate expenditures have shown less comparability.

One study from the Netherlands found that self-reported utilization of medicines in a three-month period (measured simply as the use of at least one medicine in that time period) was less comparable to administrative data than for hospitalization and physiotherapy

(Reijneveld and Stronks 2001). Studies of Medicare beneficiaries in the United States also found discrepancies in the reporting of medicine use in surveys with information from pharmacies. One found that 24% of survey respondents did not report prescription drug expenditures, in spite of pharmacy records showing purchases (Berk et al. 1990). Among those who reported any expenditures, actual expenditures were under-reported on average by 23%. More accurate reporting was associated with higher education, having private insurance, and younger age groups (Berk et al. 1990). Another study found an under-reporting of medicine use by about 18% and of expenditure by 17%; also, nearly a quarter of the sample over-reported their use of medicines (Poisal 2003). A review of the evidence from 1980–1997 on accuracy of patient self-reports of health care use found 13 studies of medicine use and stated that the ability of patients to recall their drug use is unsatisfactory (Evans and Crawford 1999).

The comparability of self-reported medication use and pharmacy claims data is made difficult by the inability to ascertain whether inaccurate reporting stems from recall problems or from non-adherence. The two data sources have different strengths and weaknesses for research into equity in the use of medicines. Self-reported utilization may be inaccurate due to problems with recall and they may include over-the-counter (OTC) medicines and drugs that are purchased privately, although the survey questions refer to medicines that are actually consumed. Claims data of dispensed medicines are not biased by recall, they contain information on expenditure, and they are limited to the prescription drugs that are purchased within the public drug program; however, the patient may not consume all drugs that are dispensed (in some cases because the drugs were not needed at the time).

Description of the Data

This study makes use of two data sources in Ontario that are linked at the individual level: the Canadian Community Health Survey (CCHS) and the drug claims database of the Ontario Drug Benefit (ODB) program. The ODB database contains the information that pharmacists submit on the medication that is dispensed, including the drug name, dosage form and strength, the date, quantity, and duration of the dispensation. Levy et al. audited 50 pharmacies in Southern Ontario and found extremely high reliability of the coding of drug type, date, quantity, and duration of the dispensed drugs in the ODB

claims database (2003).[3] The database that I include in this study includes the drug information for the 25 most commonly prescribed drugs.

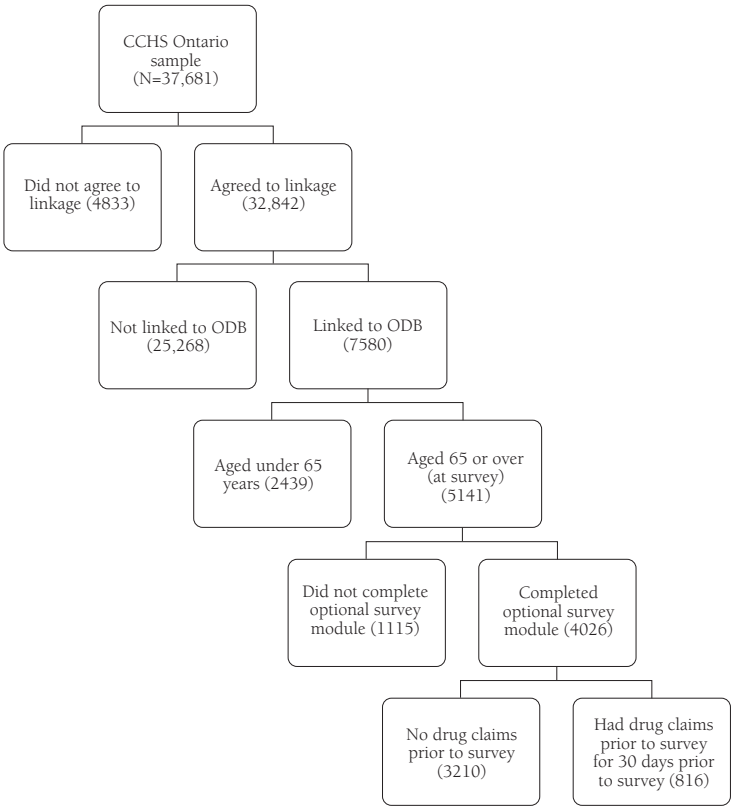
The CCHS is a cross-sectional survey produced by Statistics Canada of 133,300 individuals from 136 health regions across Canada. The CCHS targets persons aged 12 years and older who are living in private dwellings. Excluded populations are those living on Indian Reserves or Crown lands, residents of institutions, full-time members of the Canadian Armed Forces, and residents of some remote regions. The first wave of the survey was collected between 2000 and 2001. It includes a comprehensive set of information on economic, social, demographic, occupational, and environmental correlates of health, and numerous questions on health care utilization. I draw on the Ontario sample of the survey. An optional survey module on medication use was administered in 29 out of the 37 public health regions in Ontario.

The CCHS data for Ontario are linked to the ODB database through the specialized Ministry of Health and Long-Term Care link files. The focus is on those aged 65 years and older who live in private dwellings because the ODB program is the primary payer for this population for all prescription medicines included in the formulary, and the sample is representative of this older population.

To compare reported and dispensed drugs, the cohort for the study included a subset of individuals who had at least one drug dispensed in the 30-day period prior to their survey interview. Figure 1 illustrates how the sample was selected. Starting with 37,681 individuals who were in the Ontario sample of the CCHS, I ended up with 816 individuals who agreed to be linked to administrative data sources, had some prescription drug claims in the ODB program (within a six-year period), were aged 65 years or older at the time of the survey, completed the optional survey module on medicine use, and had at least one drug claim in the 30 days prior to their survey interview date (the reference period for the survey question).[4]

The three drug categories that are included in the ODB database and also referred to in the CCHS survey questions on medicine use are for "blood pressure," "heart," and "stomach remedies"; the medicine survey questions and the corresponding ODB drug names are listed in Table 1. For the remainder of the medicines referred to in the survey, there were no clear matches with the drugs in the ODB database (these

Figure 1. Number of People Excluded in Each Step of the Procedure to Create the Sample



include pain relievers, medicine for asthma, cough/cold, allergies, diabetes, sleeping pills). The sample was selected in order to address the question of whether individuals who had prescriptions dispensed (under the ODB program) reported this usage in the CCHS.

Empirical Analysis of Comparability

I compare the information on prescription claims in the ODB database with self-reported use of medicines in the CCHS, and then estimate the factors associated with under-reporting. Table 1 outlines the drugs that are investigated. Using blood pressure medications as an example, individuals in the survey were asked whether they had taken any medication for blood pressure in the past 30 days. Four drugs in the ODB database were categorized as hypotensives. Among those individuals who had at least one of the hypotensive drugs in the 30-

day period according to ODB, I first observe the proportion that reported having used medication for blood pressure in the survey, and second I estimate the individual level characteristics that are associated with discrepancies between self-report and administrative records.

Table 1. Matched Survey Question and Claims Data for Three Drug Categories Sample

CCHS Survey question: In the past month, did you take…?	Drugs included in ODB database (top-25 prescribed drugs)	Therapeutic drug subcategory
"Medicine for blood pressure"	Enalapril, Maleate, Nifedipine, Ramapril	Hypotensive drugs
"Medicine for the heart"	Nifedipine, Diltiazem HCL, Amlodipine	Cardiac drugs
	Simvastatin, Atorvastatin, Rosuvastatin	Antilipemic drugs
	Enalapril Maleate, Nifedipine, Ramapril	Hypotensive drugs
"Stomach remedies"	Omeprazole, Pantoprazole, Rabeprazole, Lansoprazole	Gastrointestinal drugs

For three subgroups, those who had drug claims for blood, heart, or stomach medicines, I calculate the proportion of under-reporting as the percentage of those with at least one of the relevant prescription drug claims who did not self-report their use in the survey. Then I model the probability of "under-reporting" (i.e., not reporting use in the survey) on a set of individual level characteristics using a probit equation.

The dependent variable is equal to 1 if the individual did not report having used the relevant medication (blood pressure/heart/stomach remedies) in the past month; it is equal to 0 if the individual reported use. Independent variables include demographic, health-related, and socio-economic variables. Health status is measured by general self-assessed health in five categories and ranges from excellent to poor, the number of reported chronic conditions, whether the individual had two or more medicines in the past month (compared to one). Age is measured in five categories: 65–69, 70–74, 75–79, 80–84, and 85 years and older. I measure socio-economic status by the drug program for low-income seniors, the level of education (less than secondary

education, secondary, some post-secondary, and post-secondary), and immigration status.

Findings from the Comparability Study

Almost all of those included in the cohort of individuals who had a drug claim in the 30-day period prior to the survey interview reported having taken some kind of medicine in the past month. However, the reporting of the use of drugs for blood pressure, heart, and stomach remedies was less comparable. About 16% of those who had taken at least one blood pressure medication according to the ODB database did not report this in the survey, 58% of those who were taking a cardiovascular drug did not report taking heart medicines, and 42% of those who had claims for gastrointestinal medications did not report taking stomach remedies in the survey (Table 2).

Table 2. Comparison of Self-reported and Dispensed Medicines

Self-reported medication use in the past month (CCHS)		
	N	**% total sample**
Blood pressure	572	70.10
Heart	320	39.22
Stomach	179	21.94
Medicines dispensed in the past month (ODB)		
	N	**%**
Blood pressure	338	41.42
Heart	671	82.23
Stomach	144	17.65
Under-reporting (dispensed but did not self-report)		
	N	**%**
Blood pressure	54	15.98
Heart	388	57.82
Stomach	61	42.36

For those who had any blood pressure medications dispensed in the past month, we modeled the likelihood of not reporting this use on a set of demographic, health, and socio-economic variables (Table 3). The only factors that are significantly associated with under-reporting

the use of blood pressure medicines are gender, education, and being born in Canada. Taking the marginal effects calculated from probit coefficients, it appears that women and those with a secondary education (compared with less than secondary education) have about a 10% reduced likelihood of under-reporting (not reporting taking blood pressure medications among those who had positive use in the ODB). Those who were born in Canada have an 8% higher likelihood of under-reporting. As for heart medicines, there is a substantial degree of under-reporting. Unlike with blood pressure medications, women have an 18% increased likelihood of under-reporting heart medicine use, while people in poorer self-reported health and with more chronic conditions have significantly lower likelihood of under-reporting. The model of self-reported stomach remedies is not significant, although poorer self-assessed health and being in age category 80–84 reduces the likelihood of under-reporting.

Overall, there appears to be relatively low comparability of the two data sources of medicine use. It is possible that some people who had drugs dispensed did not actually take them during the 30-day period. This is more likely with the stomach remedies (these drugs are mostly used to treat gastric reflux disease) than with blood pressure medicines, since the latter ought to be taken daily. It is also possible that some people do not know what blood pressure medication is or that they even have a diagnosis of hypertension (Muhajarine et al. 1997; Tu et al. 2007). The findings are consistent with previous studies that show relatively low concordance between survey and registry data of medicine use (Berk et al. 1990; Evans and Crawford 1999; Reijneveld and Stronks 2001) with a higher rate of reporting among the better educated (Reijneveld and Stronks 2001).

However, there are some important limitations with this analysis. The small sample sizes in this analysis limits its power, and the restriction of the drug data to the top-25 prescribed drugs limits the ability to generalize these findings to the entire population of medicine users. Further studies are needed to explore the comparability of different data sources and the association between education and recall of drug use. This analysis suggests that one must be cautious when interpreting self-reported medicine use from the CCHS. Furthermore, analyses of equity that are based on self-reported data—in particular, if the studies include information on the types of drugs that are consumed and the questions do not include details of the drugs or

Table 3. Factors Associated with the Under-reporting of Medicine Use among Those with Prescription Drug Claims

	Blood pressure		Heart medicine		Stomach remedies	
	ME	SE	ME	SE	ME	SE
Age 70–74	0.095	0.068	0.044	0.058	-0.173	0.125
Age 75–79	0.054	0.061	0.008	0.057	-0.068	0.125
Age 80–84	0.033	0.074	-0.008	0.066	**-0.237**	0.117
Age 85+	0.136	0.101	-0.048	0.083	-0.213	0.152
female	**-0.104**	0.042	**0.172**	0.043	0.041	0.096
prog3 (low-income senior)	-0.062	0.040	0.021	0.047	0.061	0.100
sah2 (very good)	-0.055	0.077	-0.056	0.105	**-0.349**	0.148
sah3 (good)	0.042	0.099	-0.165	0.101	-0.282	0.190
sah4 (fair)	0.052	0.100	**-0.243**	0.101	-0.285	0.204
sah5 (poor)	0.229	0.159	**-0.344**	0.100	-0.229	0.210
Number of chronic conditions	-0.018	0.010	**-0.079**	0.011	-0.032	0.023
educ2 (secondary)	**-0.096**	0.039	0.085	0.058	0.090	0.134
educ3 (some post-secondary)	-0.088	0.058	0.027	0.091	0.346	0.164
educ4 (post-secondary)	-0.048	0.040	0.008	0.050	-0.056	0.114
Canada born	**0.076**	0.037	-0.038	0.044	-0.136	0.105
total2 (2 or more drugs in past 30 days)	-0.017	0.040	-0.028	0.042	0.056	0.092
Number of observations	338		671		144	
Pseudo R2	0.1197		0.1382		0.0889	
Chi2	**35.55**	0.0033	**126.27**	0	17.44	0.3575

Note: ME is marginal effect, SE is standard error, bold is significant at p<0.05; sah is self-assessed health

requests for respondents to check their medicine cabinets—may yield different findings than those based on claims data. The finding that higher educated individuals who had drugs dispensed are more likely to report their medicine use in surveys suggests that studies that rely on survey data may overestimate the education effect on pharmaceutical usage and bias estimates toward "pro-rich" inequity.

Assessment of Equity in the Use of Medicines

To assess equity in the pharmaceutical sector, researchers need accurate information on medicine use, socio-economic status, and other individual characteristics. In light of the apparent misreporting of medicine use in survey data, the inclusion of over-the-counter drugs in these data, and the lack of expenditure information, claims data are likely to be the more appropriate measure of utilization. In this section, I describe the concentration index approach to measuring inequity in the use of prescribed drugs and in out-of-pocket expenditure, and report the results from a preliminary analysis based on linked survey and administrative data.

There has been little assessment of the extent of inequity in the use of and expenditure on medicines within Ontario's public drug program. The few studies to date have relied on survey data (Ballantyne et al. 2005; Grootendorst 1995; Zhong 2007), which, as outlined above, may not be appropriate for at least two reasons. First, if the limited comparability of survey and administrative data relate to recall difficulties, and this recall is more accurate among higher socio-economic groups, then this would bias the estimate of socio-economic inequity upwards. Second, over-the-counter medicines are often costly. Therefore, the distribution of OTC medicine use is likely to be more concentrated among higher-income groups. The inclusion of OTC drugs in analyses of equity in the ODB program could lead to an overestimation of inequity.

With linked administrative and survey data, one can estimate inequity in the use of and expenditure on prescription drugs within the public drug program. Survey data include important information on individuals' socio-economic status, their level of health, and other characteristics that affect the use of health care services and of medicines. Administrative data of prescription drug use include information on expenditure, which can be separated into the costs borne by the province (through the public drug plan) and those that are borne by the patient (in the form of out-of-pocket payments).

Empirical Analysis

One approach (and the only one based on an explicit social welfare function) to the assessment of inequity is with the concentration index, which indicates how concentrated the use of medicine is along

the distribution of income in the population (O'Donnell et al. 2008; Wagstaff and van Doorslaer 2000). This method derives from the literature on income inequality. Similar to the Lorenz curve, which describes the distribution of income in a population, the concentration curve describes the relationship between the cumulative proportion of the population ranked by income (on the x-axis) and the cumulative proportion of health services or medicine use (on the y-axis). And like the Gini index, which provides a measure of income inequality, the concentration index is a measure of income-related inequality in use of health services or medicines and it is estimated as twice the area between the concentration curve and the line of perfect equality (the diagonal). When utilization is more concentrated in the upper end of the income distribution, after adjusting for differences in need as measured by health status, there is "pro-rich" inequity. It is also possible for the level of needs-adjusted utilization to be concentrated among the lower-income groups; in the literature this is referred to as "pro-poor" inequity. Such "pro-poor" inequity could be understood as an over-utilization among the poorer groups, which could be problematic in the case of medicines due to the increased risk of adverse events that is associated with poly-pharmacy. The concentration index could also be used to provide evidence of inequity in the financing of prescription drugs, whereby a positive concentration of out-of-pocket payments would suggest that the payment system is progressive, since higher-income individuals are paying proportionately more out-of-pocket than those with lower income.

This empirical research can thus serve as an assessment of inequity in Ontario's public drug program. To the extent that, after adjusting for differences in need, medicine use is greater for those with higher socio-economic status, and that out-of-pocket expenditure is higher for lower-income individuals, there is some evidence of inequity in the public drug program.

For this analysis I make use of five years of ODB drug claims data (of the 25 most commonly prescribed drugs) for individuals aged 65 years and older, linked to the CCHS from 2001.[5] The final sample consists of 6016 individuals over the period 2001–2006 (yielding a total of 23,517 observations). The dependent variables of interest are the total number of claims that an individual made per year (this covers all ODB claims and not just the top-25 prescribed drugs), the total expenditure on prescription drugs, and

the total out-of-pocket expenditure on prescription drugs. Because these variables are highly skewed, I took the natural logarithm of each as the dependent variables (but because there are not many zeros there is no real need for a two-step model). The independent variables are derived from the survey data (from 2001) and are separated into those that relate to need for medicines, including age, sex, and health status, and those considered unrelated to need, including income, education, marital and immigrant status, overweight, lifestyle factors (current or past smoker, drinking heavily once per week, eats five portions of fruit and vegetable per day), and year dummies.

Three random effects regression models are run to estimate the effects of need and non-need variables on medicine use and expenditure, accounting for the panel nature of the data. Then I calculate the concentration indices of inequality, as the covariance between the income rank in the population and expenditure, to measure how concentrated is use/expenditure along the distribution of income. I calculate inequity as the extent of inequality that remains after adjusting for differences in need across the income distribution.[6]

Results of Preliminary Analysis of Equity in Medicine Use

There appears to be an association between individuals' income and their use of medicines, the total expenditure incurred, and their out-of-pocket expenditure (Table 4). Seniors with higher income have fewer prescription drug claims on average, incur fewer costs, and spend more out-of-pocket than those with lower income.

The results of the random effect panel analyses confirm that socio-economic factors affect the use of medicines (Table 5). After adjusting for age, self-reported health status, and lifestyle factors, individuals in the low-income senior program make 20% more drug claims than those in the high-income senior program. Higher educated individuals make 6.4% fewer claims and incur 5.3% less cost on prescription drugs than those with less than a secondary school education. Canadian-born individuals incur 6.4% more costs on prescription drug medicines. The strongest predictors of medicine use, as expected, are health status and age: people who are older, with worse self-reported health, more chronic conditions, and limitations in activity make more drug claims than relatively younger and healthier individuals. Total expenditure is related to worse health but younger

Table 4. Mean Expenditure, Out-of-pocket (OOP) Expenditure, and Number of Claims (and Standard Deviation, SD) by Income Quintile and by Program Type

	Total expenditure ($)	OOP expenditure ($)	Total claims (number)
	Mean (SD)	Mean (SD)	Mean (SD)
Individual income quintile			
Lowest-income quintile (mean income $8,494.58)	850.92 (785.31)	43.71 (63.75)	48.19 (91.49)
Q2 (mean $14,602.44)	955.14 (873.84)	49.69 (58.31)	41.88 (48.65)
Q3 (mean $19,926.83)	837.60 (780.81)	70.90 (74.38)	34.78 (42.22)
Q4 (mean $26,970.55)	831.57 (741.51)	75.29 (71.31)	29.40 (34.35)
Highest-income quintile (mean $50,838.90)	724.40 (642.75)	71.55 (65.73)	25.63 (30.33)
ODB Program			
Low-income senior	884.68 (794.98)	20.55 (36.67)	49.75 (77.04)
High-income senior	812.40 (749.76)	78.34 (70.18)	29.36 (34.16)

Note: the total number of claims includes repeat prescriptions and refills, and covers all claims in the ODB; total and out-of-pocket expenditure is for the top-25 prescribed drugs

age. The concentration indices of inequality in the use of and total expenditure on medicines are negative and significant, indicating a significant concentration in the lower end of the income distribution. Health status is worse, on average, among lower-income individuals; therefore, after adjusting for health and age, the concentration index reduces, but remains negative and significantly different from zero. For total expenditure, the index of inequity is very close to zero, which implies an equitable distribution of expenditures.

The analysis of out-of-pocket expenditures confirms that wealthier individuals spend more out-of-pocket than those with lower income (Table 5). I find a significant association between income and out-of-pocket payments after adjusting for demographic and socio-economic characteristics. The significant difference in out-of-pocket expenditure is between the low-income seniors (in the program for the lowest-income group) and the rest of the population. This is not

Table 5. The Marginal Effects of Demographic, Health, and Socio-economic Factors on the Number of Claims, Total Expenditure on Prescription Drugs, and Out-of-pocket Expenditure, and Indices of Inequality, 2001–2006

	Total claims (ln)		Total expenditure (ln)		OOP expenditure (ln)	
	ME	SE	ME	SE	ME	SE
Demographic and health variables						
Age 70–74	**0.035**	0.014	-0.004	0.017	**-0.095**	0.019
Age 75–79	**0.102**	0.018	-0.003	0.021	**-0.096**	0.023
Age 80–84	**0.200**	0.022	-0.013	0.026	-0.023	0.026
Age 85+	**0.260**	0.028	**-0.064**	0.033	-0.014	0.033
Female	-0.030	0.021	**-0.149**	0.025	**-0.085**	0.022
sah2 (very good)	**0.160**	0.034	**0.090**	0.041	-0.038	0.037
sah3 (good)	**0.305**	0.034	**0.223**	0.040	0.026	0.036
sah4 (fair)	**0.459**	0.037	**0.278**	0.045	0.058	0.040
sah5 (poor)	**0.578**	0.047	**0.264**	0.057	0.064	0.051
limit2 (severe limitation)	**0.137**	0.024	**0.027**	0.029	-0.001	0.026
Number of chronic conditions	**0.102**	0.005	0.078	0.006	**0.029**	0.006
Socio-economic and lifestyle factors						
inc2 (2nd quintile)	-0.001	0.032	0.026	0.039	-0.023	0.035
inc3	0.032	0.034	0.028	0.041	0.016	0.037
inc4	-0.029	0.035	0.046	0.042	0.054	0.038
inc5 (highest quintile)	-0.055	0.036	-0.007	0.043	0.029	0.039
High education	**-0.064**	0.019	**-0.053**	0.023	-0.065	0.021
Born in Canada	0.031	0.021	**0.064**	0.025	0.015	0.022
Married	**-0.086**	0.020	-0.003	0.025	-0.001	0.022
Current or past smoker	0.014	0.021	-0.029	0.025	0.002	0.023
Overweight	0.010	0.019	**0.052**	0.023	0.032	0.020
Eats veg/fruit 5 per day	-0.006	0.018	0.018	0.022	0.004	0.020
Heavy drinking	**-0.078**	0.032	**-0.105**	0.038	-0.023	0.034
Low-income program	**0.195**	0.025	0.029	0.030	**-1.459**	0.028
Year dummies						
2002	**0.063**	0.011	**0.335**	0.014	**0.187**	0.018
2003	**0.137**	0.011	**0.460**	0.014	**0.328**	0.018
2004	**0.199**	0.011	**0.585**	0.014	**0.433**	0.018
2005	**0.243**	0.011	**0.698**	0.014	**0.492**	0.018
2006 (Jan-March)	n.a.		**-0.457**	0.015	**-1.150**	0.019
Constant	**2.278**	0.052	**5.477**	0.062	**3.747**	0.057
Indices of inequality	Index	95% CI	Index	95% CI	Index	95% CI
Unadjusted inequality index	**-0.032**	(-0.035, -0.028)	**-0.004**	(-0.004, -0.002)	0.053	(0.049, 0.057)
Adjusted inequality index (Inequity)	**-0.015**	(-0.018 -0.012)	**-0.003**	(-0.005, -0.001)	**0.054**	(0.05, 0.058)

Note: bold is significant at p<0.05; sah is self-assessed health

surprising given that the public drug program nearly eliminates cost sharing for individuals below a low-income threshold (they have no deductible and may have to pay $2 copayment for each drug that is dispensed). This analysis is supported by the concentration indices that are positive, and significantly different from zero; there is a greater concentration of out-of-pocket payments among the higher-income seniors. While this program is successful at protecting the lowest-income individuals from the burden of costs, the overall progressivity of the program is limited since payments are not related to income for the remainder of the population.

Discussion

One of the objectives of the public drug program in Ontario is to ensure equitable access to medicines. This paper examined the comparability of two sources of prescription drug use data for the purpose of evaluating equity in prescription drug use, and presented a preliminary assessment of inequity within Ontario's public drug program. It appears that the comparability of the CCHS and ODB data of medicine use varies across the different drug categories but with relatively high rates of "under-reporting." In light of this finding, and the fact that the survey does not include information on the drug costs incurred, claims data were used to analyze the socio-economic effects on drugs dispensed in the public program.

The study found a significant association between income and the use of medicines, along with a negative index of inequity, which suggests that there is little evidence of inequity in favour of the rich. On the contrary, the elevated use of medicines among lower-income groups, even after controlling for health status, raises some concern associated with the potential increased risk of adverse events and of contraindications associated with polypharmacy (Kroenke and Pinholt 1990). These findings contrast those of Zhong (2007), who found a slight "pro-rich" inequity in the total number of drugs used among the 65-years-and-older population in Ontario (with an index of 0.02) and a slightly greater degree of inequity among the working-age population. This latter finding is expected since, within this population, there are some individuals who do not have either public or private insurance. However, it is possible that recall errors in self-reported medicine use, along with the inclusion of over-the-counter medicines, explain the Zhong (2007) findings for the older population. It could also be that this study includes only the top-25 prescribed drugs and not all drugs.

In this study, I find that people with lower income make more drug claims than those with higher income, but this income effect is not seen when utilization is measured in terms of total expenditure. Since the total expenditure is restricted to the top-25 drugs, whereas the total number of claims refers to all drugs dispensed, the patterns of prescription drug use by different income groups warrant further attention. There are some additional hypotheses that are worth exploring, such as whether perhaps people with higher income demand more expensive, brand-name drugs (drugs that would be paid for through the public system if they are included in the provincial formulary). Also, higher-income patients may be more likely to persuade their doctors to prescribe limited use drugs, drugs that are not included in the formulary but may be accessed under special circumstances. Disaggregated analyses by different therapeutic categories could shed some light on these patterns of medicine use. Finally, the patterns of out-of-pocket spending could be investigated further to determine whether cost sharing poses a financial barrier to access among poor individuals whose income is not low enough to be eligible for cost-sharing exemptions.

Acknowledgements

This study makes use of data that were provided by the Ontario Ministry of Health and Long-Term Care as part of a grant entitled: "Aging, Social Capital, and Utilization of Health Services in Ontario—A Longitudinal Study," led by Professor A. Laporte. I would like to thank Audrey Laporte, Michel Grignon, Carolyn Tuohy, Paul Grootendorst, and the participants of the McMaster Centre for Health Economics and Policy Analysis Polinomics workshop for their helpful comments. I also gratefully acknowledge funding from the Lupina Foundation and the Canadian Health Services Research Foundation.

Notes

1. For the 65-years-and-older population group, the ODB database includes information for almost the entire population of this age unlike for other groups where coverage can change over time (e.g., with movement in and out of paid jobs) and because of eligibility requirements (Levy et al. 2003).

2. An individual is considered to be low income if he or she has a net annual income below $16,018 if single, and below $24,175 if part of a couple.

3. Similar studies have not been conducted in the rest of the province. The southern Ontario region has a greater proportion of chain pharmacies than for the province as a whole, so it is possible that the reliability of these data may differ to some extent throughout the province (Levy et al. 2003).

4. There is a possibility of response bias being introduced at either of the two points of selection: whether the survey respondent agrees to have his or her records linked (and providing the Ontario health card number), and completion of the optional survey module on drug use. A previous study found no substantive difference in socio-economic, health, and demographic characteristics between those who answered the optional drug module and those who did not. Furthermore, since individuals did not decide to respond, but rather the decision was made for health regions, there is no individual-level selection in effect (Zhong 2007). Finally there appears to be no significant differences in health, demographic, socio-economic, and health utilization indicators between the full Ontario sample of the CCHS and the sample that agreed to record linkage.

5. The analysis included only those older people who were in private residence and not residents of long-term care institutions, since the CCHS covers only the private resident population.

6. The process of needs adjustment is not straightforward and requires measurable indicators of need. The literature has relied mostly on self-reported health as a proxy for need, whereby, on average, individuals who report worse health are assumed to have greater need for health services. This approach has advantages and disadvantages. On the one hand, numerous studies have found that self-reported health is a strong predictor of mortality and health service use (Idler and Benyamini 1997). On the other hand, there may be bias in reporting of self-assessed health and self-reported health conditions related to age and possibly also to socio-economic status (Bago d'Uva et al. 2008).

References

Bago d'Uva, T., O. O'Donnell, and E. van Doorslaer. 2008. Differential Health Reporting by Education Level and Its Impact on the Measurement of Health Inequalities among Older Europeans. *International Journal of Epidemiology* 37 (6): 1375–83.

Ballantyne, P.J., J.C. Victor, J.E. Fisher, and J.A. Marshman. 2005. Factors Associated with Medicine Use and Non-use by Ontario Seniors. *Canadian Journal on Aging* 24 (4): 419–31.

Berk, M.L., C.L. Schur, and P. Mohr. 1990. Using Survey Data to Estimate Prescription Drug Costs. *Health Affairs* 9 (3): 146–56.

Birch, S., and J. Abelson. 1993. Is Reasonable Access What We Want? Implications of, and Challenges to, Current Canadian Policy on Equity in Health Care. *International Journal of Health Services* 23 (4): 629–53.

Canadian Institute for Health Information. 2008. *Drug Expenditure in Canada, 1985–2007*. Ottawa: Statistics Canada.

Cleary, P.D. 1984. The Validity of Self-reported Physician Utilization Measures. *Medical Care* 22 (9): 796–803.

Commission on the Future of Health Care in Canada. 2002. *Building on Values: The Future of Health Care in Canada*. Ottawa: Government of Canada.

Culyer, A.J., and A. Wagstaff. 1993. Equity and Equality in Health and Health Care. *Journal of Health Economics* 12 (4): 431–57.

Demers, V., M. Melo, C.A. Jackevicius, J. Cox, D. Kalavrouziotis, S. Rinfret, et al. 2008. Comparison of Provincial Prescription Drug Plans and the Impact on Patients' Annual Drug Expenditures. *Canadian Medical Association Journal* 178 (4): 405–9.

Evans, C., and B. Crawford. 1999. Patient Self-reports in Pharmacoeconomic Studies. Their Use and Impact on Study Validity. *Pharmacoeconomics* 15 (3): 241–56.

Evans, R.G. 1983. Health Care in Canada: Patterns of Funding and Regulation. *Journal of Health Politics, Policy and Law* 8 (1): 1–43.

———. 1992. The Canadian Health Care Financing and Delivery System: Its Experiences and Lessons for Other Nations. *Yale Law and Policy Review* 10 (2): 362–96.

———. 2005. The Economics of Private Insurance. In *Access to Care, Access to Justice: The Legal Debate Over Private Health Insurance in Canada*, edited by C.M. Flood, K. Roach, and L. Sossin, 347–68. Toronto: University of Toronto Press.

Gama, H., S. Correia, and N. Lunet. 2009. Questionnaire Design and the Recall of Pharmacological Treatments: A Systematic Review. *Pharmacoepidemiology and Drug Safety* 18 (3): 175–87.

Gemmill, M., S. Thomson, and E. Mossialos. 2008. What Impact do Prescription Drug Charges Have on Efficiency and Equity? Evidence from High-income Countries. *International Journal of Equity in Health* 7:12.

Giacomini, M., J. Hurley, I. Gold, P. Smith, and J. Abelson. 2004. The Policy Analysis of "Values Talk": Lessons from Canadian Health Reform. *Health Policy* 67 (1): 15–24.

Glandon, G.L., M.A. Counte, and D. Tancredi. 1992. An Analysis of Physician Utilization by Elderly Persons: Systematic Differences between Self-report and Archival Information. *Journals of Gerontology* 47 (5): S245–52.

Grootendorst, P. 1995. A Comparison of Alternative Models of Prescription Drug Utilization. *Health Economics* 4 (3): 183–98.

Idler, E., and Y. Benyamini. 1997. Self-rated Health and Mortality: A Review of Twenty-seven Community Studies. *Journal of Health and Social Behavior* 38 (1): 21–37.

Jobe, J.B., A.A. White, C.L. Kelley, D.J.Mingay, M.J. Sanchez, and E.F. Loftus. 1990. Recall Strategies and Memory for Health-care Visits. *Milbank Quarterly* 68 (2): 171–89.

Johnson, R.E., and W.M. Vollmer. 1991. Comparing Sources of Drug Data about the Elderly. *Journal of the American Geriatrics Society* 39 (11): 1079–84.

Klungel, O.H., A. de Boer, A.H. Paes, R.M. Herings, J.C. Seidell, and A. Bakker. 2000. Influence of Question Structure on the Recall of Self-reported Drug Use. *Journal of Clinical Epidemiology* 53 (3): 273–77.

Kroenke, K., and E.M. Pinholt. 1990. Reducing Polypharmacy in the Elderly. A Controlled Trial of Physician Feedback. *Journal of the American Geriatrics Society* 38 (1): 31–36.

Levy, A.R., B.J. O'Brien, C. Sellors, P Grootendorst, and D. Willison. 2003. Coding Accuracy of Administrative Drug Claims in the Ontario Drug Benefit Database. *Canadian Journal of Clinical Pharmacology* 10 (2): 67–71.

Lexchin, J., and P. Grootendorst. 2004. Effects of Prescription Drug User Fees on Drug and Health Services Use and on Health Status in Vulnerable Populations: A Systematic Review of the Evidence. *International Journal of Health Services* 34 (1): 101–22.

Marquis, K.H., M.S. Marquis, and J.P. Newhouse. 1976. The Measurement of Expenditures for Outpatient Physician and Dental Services: Methodological Findings from the Health Insurance Study. *Medical Care* 14 (11): 913–31.

Morgan, S. 2004. Drug Spending in Canada: Recent Trends and Causes. *Medical Care* 42 (7): 635–42.

Muhajarine, N., C. Mustard, L.L. Roos, T.K. Young, and D.E. Gelskey. 1997. Comparison of Survey and Physician Claims Data for Detecting Hypertension. *Journal of Clinical Epidemiology* 50 (6): 711–18.

O'Donnell, O., E. van Doorslaer, A. Wagstaff, and M. Liondelow. 2008. *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation*. Washington, DC: The World Bank.

Paterson, J.M., A. Suleilman, J.E. Hux, and C. Bell. 2008. How Complete Are Drug History Profiles That Are Based on Public Drug Benefit Claims? *Canadian Journal of Clinical Pharmacology* 15 (1): e108–e116.

Poisal, J.A. 2003. Reporting of Drug Expenditures in the MCBS. *Health Care Financing Review* 25 (2): 23–36.

Raina, P., V. Torrance-Rynard, M. Wong, and C. Woodward. 2002. Agreement between Self-reported and Routinely Collected Health-care Utilization Data among Seniors. *Health Services Research* 37 (3): 751–74.

Reijneveld, S.A., and K. Stronks. 2001. The Validity of Self-reported Use of Health Care across Socioeconomic Strata: A Comparison of Survey and Registration Data. *International Journal of Epidemiology* 30 (6): 1407–14.

Roberts, R.O., E.J. Bergstralh, I. Schmidt, and S.J. Jacobsen. 1996. Comparison of Self-reported and Medical Record Health Care Utilisation Measures. *Journal of Clinical Epidemiology* 49 (9): 989–95.

Tu, K., N.R.C. Campbell, Z.-L. Chen, K.J. Cauch-Dudek, and F.A. McAlister. 2007. Accuracy of Administrative Databases in Identifying Patients with Hypertension. *Open Medicine* 1 (1): E18–E26.

Wagstaff, A., and E. van Doorslaer. 2000. Equity in Health Care Finance and Delivery. In *Handbook of Health Economics* edited by A.J. Culyer and J.P. Newhouse, 1803–62. Amsterdam: North-Holland.

Wallihan, D.B., T.E. Stump, and C.M. Callahan. 1999. Accuracy of Self-reported Health Services Use and Patterns of Care among Urban Older Adults. *Medical Care* 37 (7): 662–70.

Zhong, H. 2007. Equity in Pharmaceutical Utilization in Ontario: A Cross-section and Over Time Analysis. *Canadian Public Policy* 33 (4): 487–506.

A New Chapter in the Interprofessional Collaborative Literature: Developing a Framework to Examine Professional Culture on a Family Health Team

Jennifer Beales, Murat Kristal,
Christine Papoushek, and Zubin Austin

Abstract

Family health teams (FHTs) have been created to respond to the mandate of collaborative patient-centre care, yet barriers such as Professional Culture (PC) pose a challenge to attaining an interprofessional collaborative (IPC) ideal. The paucity of literature on PC confounds our understanding of how this concept impacts IPC and the evolving FHT. A goal of this study is to explore how PC manifests itself on an FHT and within FHT culture, in an effort to develop a framework to examine the dynamics of cultural conflict on an FHT. Qualitative data were collected using in-depth semi-structured focus groups (n=5). Participants were recruited from medicine, nursing, and allied health professions at an FHT and Diabetes Education Centre in a large academic teaching hospital in urban Canada. Discussions were audio-taped and transcribed verbatim. Transcripts were coded and analyzed using a modified directed content analysis approach. Analysis identifies three themes (Professional World Views; FHT World Views; and Resource Utilization) that generate a framework and depict the tension between these domains. Future ethnographic studies of FHTs would elucidate the concept of PC and be well suited to further develop this framework.

Introduction

Interprofessional collaboration (IPC) has become progressively fashionable—a ubiquitous buzzword—supported and used in academic and governmental circles alike. This comes as no surprise given the cultural shift that we are witnessing in the delivery of health care. The cultural shift, which advocates patient centredness, is constructed on the tenet that IPC among health care providers will improve patient care. In this way, IPC has been identified as a vehicle to deliver the mandate of patient-centred care. Endorsed by politicians, educators, and academics alike, this concept seems likely to hold sway for the foreseeable future. Moreover, a burgeoning body of literature on IPC is evidence that this concept is of growing academic interest.

Canadians have long exchanged ideas regarding the affordability and sustainability of our health care system. In an effort to deal with these concerns, reports from Commissioner Roy Romanow, and the Federal/Provincial/Territorial 2003 First Ministers' Accord on Health Care Renewal addressed the importance of IPC to improve patients' quality of care and to effect change in our health care system (Romanow 2002; National Steering Committee on Patient Safety 2002). In this emerging paradigm, the health care system strives to make optimal use of existing health care practitioners' knowledge and skills, while government and educational arenas work on diminishing the Health Human Resource (HHR) shortage (Ontario 2005b, p. 3). Literature suggests that quality of care can be improved when skills, knowledge, and experience are effectively coordinated between professional groups (Reeves and Lewin 2004; Reeves et al. 2003; Schmitt 2001; Way et al. 2000; Sorrells-Jones 1998; Aiken et al. 1998; Zwarenstein et al. 1997). Moreover, the interprofessional collaborative ideal in health policy is widely supported in political agendas internationally (Kahn 2004; Leathard 2003; Willumsen and Breivik 2003; Romanow 2002; National Steering Committee on Patient Safety 2002).

In reaction to the HHR shortage, the Ontario government has created the Family Health Team (FHT) Initiative, which has established 152 FHTs across Ontario. Designed around specific community needs, these teams may consist of doctors, nurse practitioners, and nurses, in addition to chiropractors, rehabilitation workers, social workers, dieticians, pharmacists, physician specialists, and mental health workers (Ontario 2005a, 2005c; Romanow 2002). In this light, an

FHT is a valuable setting where one might examine a diverse group of health professionals who have been brought together under one umbrella that supports the notion of collaborative patient-centred care and espouses the ideal of interprofessional collaboration.

Despite a number of publications that outline best practices in collaborative teamwork, and how to foster collaborative environments (EICP 2005; Oandasan et al. 2004; Health Canada 2004; Way et al. 2000), minimal research and theory exists about how health care teams function (Phillips 1999; Cott 1998; Opie 1997) and/or whether they operate in a true collaborative fashion. This is cause for concern given that FHTs (and professionals) are expected to operate in interprofessional harmony so as to deliver the mandate of patient-centred care. The literature on IPC outlines barriers and challenges when applying the IPC ideal in health care settings. These barriers have been identified as follows: individual values; learning about interprofessional differences and respective roles; internalizing a common purpose and goal; creating norms and values that shape social and professional behaviour and activities; trust (Ontario 2005a, p. 8); fear of change; different professional agendas (Ontario 2005a, p.10); power imbalances; and the walls of professional identity and territoriality (Oandasan et al. 2004). What these barriers seemingly have in common is a shared origin that relates to what happens when unique professional cultures collide in an interprofessional team setting. Thus, while the literature on IPC is expanding, it fails to address how health care professionals possessing diverse professional ideals, or professional world views, will negotiate their roles, and collaborate as a team.

Literature does suggest that each team will develop its own culture (Ontario 2005a, p. 3; Atwal 2005; Goodenough 1970), but what this culture will look like remains unknown. What remains unclear is how professional culture influences the collaborative practice of health care professionals. Moreover, the type of culture that evolves when health professionals from different backgrounds come together under the auspices of IPC is also uncertain. For these reasons, an examination of the concept of professional culture in the context of evolving FHT culture is timely and relevant. Furthermore, given our limited understanding of professional culture and how it hampers or facilitates IPC, a new chapter in the interprofessional collaborative literature is warranted and necessary to broaden our collective knowledge base.

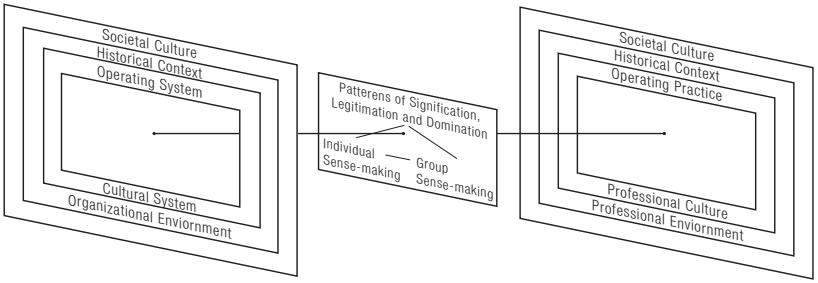
Using a qualitative approach, this study aims to explore the concept of professional culture, examine the manner in which it manifests itself in the day-to-day workings of an FHT, and ascertain whether it assists or impedes the mandate for interprofessional collaborative patient-centred care in the current FHT environment. In speaking to these goals, we sought to generate a preliminary framework that illustrates the interconnected and tightly woven variables to explain professional culture in the context of an interprofessional FHT.

Context

While the literature, as it is currently constituted, recognizes that the culture of health care professions is distinct (Hall 2005), less has been published that explores or explains the differences between and among the many professional cultures found in health care circles (Austin et al. 2007, p. 84). The literature has yet to examine professional culture in the context of an FHT, nor has it considered how professional culture will influence FHT culture (and the conditions under which this takes place). Moreover, there is scant empirical research examining how professional culture impedes or facilitates the interprofessional collaborative process in an interprofessional team environment. Given these gaps in the literature, we turned to the organizational behaviour literature to frame our investigation of professional culture in the context of an FHT.

Bloor and Dawson's provisional model examining professional culture in an organizational context (see Figure 1) was useful in establishing a foundation from which we could begin to understand how professional culture maps onto an interprofessional health team culture in general and FHT culture in particular. Here the authors succinctly depict how "professionals, individually and collectively, make sense of and structure their worlds and how organizational structures and practices in turn impinge upon professional understanding and action" (1994, p. 281). As a starting point, this model lends itself to the study of professional culture in the context of evolving FHT culture, since it takes into consideration the societal culture, the historical context, and political environment in which professions and organizations evolve and exist. Moreover, the model describes the manner in which professionals (who have been trained in different "silos" and possess dissimilar world views) use interpretive schemes (much like a frame of reference, a paradigm [Covey 1989], or Hall's [2005] cognitive map) to make sense of, and

Figure 1. A Provisional Model for Understanding Professional Culture in Organizational Context. Bloor and Dawson (1994)



interpret, their social surroundings and interactions with others in the organization.

These schemes constitute a collection of experiences from which one's world view and interpretation of everyday life are drawn. In essence, these schemes provide a lens through which one sees the world. In the context of organization life and culture, new interpretive schemes are created, tested, and revised when professionals interact with co-workers. They begin to internalize the values, norms, expected behaviours and codes of conduct, and beliefs espoused by the organization. Referred to as "signification,"[1] these schemes are embraced through learned experiences. However, an individual's previous interpretive schemes (internalized through professional training and socialization) also influence the manner in which professionals make sense of everyday life and work within the organization. The authors also consider how elements of legitimation[2] and domination[3] (the political process involved in having certain group interests and values recognized and absorbed into the culture of an organization [Bloor and Dawson 1994, p. 278]) inform the larger picture of professional culture in an organizational context. The authors also acknowledge that individual and group behaviours and characteristics, as well as elements of hierarchy and control, inform professional culture and organizational culture. We believe these distinctions in professional culture in an organizational context may be applied to the experience of health care professionals in an FHT setting. For these reasons, Bloor and Dawson's (1994) model was an ideal springboard to guide our investigation and allow us to explore professional culture in the context of an FHT.

For the purpose of our investigation we deemed "culture" to be inclusive of one's world view (or cognitive map). Culture inherently shapes the world view of a professional by providing her/him with "schemes of interpretation," which are called upon to make sense of day-to-day social surroundings and interactions with other professional team members.

Research Objectives

The primary objective of this research was to map professional culture in an FHT and generate a preliminary conceptual framework to examine professional culture and the manner in which it operates in team-based health care settings.

Methods

Study Background

"The Identification and Development of a Collaborative Practice Model in Managing Type 2 Diabetes at the Family Health Centre, Hospital A [name of hospital anonymized]" was an exploratory qualitative study. The study was designed first to explore those factors that health care professionals deemed facilitators or barriers to IPC. Next, the researchers sought to uncover the opportunities and challenges associated with the implementation of a collaborative practice model in managing type 2 diabetes. Lastly, the experiences and attitudes of primary health care providers (physicians, nurses, and allied health care professionals) regarding collaborative practice were investigated. Data presented here were collected in February and March of 2007.

Objectives

Following the development of a collaborative practice model in managing type 2 diabetes (Papoushek and Beales 2008), researchers (JB, CP, and ZA) revisited the raw data. Due to a paucity of literature on professional culture in general, and within the context of IPC in particular, the underlying area of interest in this secondary analysis was to address this knowledge gap and examine professional culture in the context of evolving FHT culture.

Two research questions emerged:

1) What happens when people from different professional backgrounds (i.e., cultures) come together under the auspices of IPC—how do they negotiate their different professional cultural backgrounds?

2) How do these professionals navigate their own professional identities to form FHT culture?

Participants

The Family Health Centre (FHC) and Diabetes Education Centre (DEC) at an academic teaching hospital located in urban Ontario were selected. The primary sampling strategy was non-random convenience (Cresswell 1998). All full- and part-time staff physicians, nurses, and allied health care professionals providing patient care at the FHT and the DEC were invited to voluntarily participate in the study. In total, 42 individuals—including pharmacists, registered dieticians, social workers, registered nurses, cognitive behavioural therapists, and nurse practitioners—participated in five focus groups. Being housed in an academic teaching hospital, all participants had familiarity with the concept and practice of IPC.

Design and Methodology

The primary study was qualitative and exploratory. As the research sought to uncover the nature of health care professionals' experiences, attitudes, and opinions, a qualitative approach was amenable with these goals (Strauss and Corbin 1998). Researchers conducted semi-structured focus groups to capitalize on the dynamic communication between participants.[4] Moreover, this approach was an efficient way to examine the range of opinions and experiences that health care professionals brought to the study (Kitzinger 1995). Field notes were recorded at the conclusion of each focus group discussion.

Focus Groups

Five semi-structured focus groups[5] consisting of 6–12 participants, each lasting approximately 90 minutes, were conducted on issues surrounding the delivery of patient-centred care in an interprofessional collaborative working environment.

All focus groups were carried out under the guidance of a trained moderator (JB), who followed the basic principles for conducting focus groups (see Berg 1998 and Krueger 1994). Using a semi-structured interview guide, participants were asked questions related to IPC. Examples of questions included: (a) how they defined collaboration?; (b) how and when team members worked together?; (c) what factors restricted their ability to collaborate?; and (d) what facilitated their ability to collaborate? A research assistant made detailed handwritten notes during the focus group discussions to

document participants' turn of speech, nonverbal communication cues, and any other information deemed pertinent during the conversations.

Ethics approval was sought by the researchers and provided by the hospital, and focus group participants provided written consent to voluntarily participate in the discussion sessions and have these audio-recorded for data transcription purposes. Participant confidentiality was preserved and maintained by anonymizing all transcripts and field notes, purging them of all identifying characteristics, assigning pseudonyms when and where required, and assuring participants that no identifying information would appear in any published papers.

Data Analysis

Focus groups were audio-recorded and transcribed verbatim, omitting all identifying characteristics. Data were entered, organized, and coded in QSR International's Nvivo 7 (a multi-function qualitative research software program). Data were analyzed using a modified directed content analysis approach (Miller et al. 2008; Hsieh and Shannon 2005) that utilized sensitizing concepts to guide the coding process (Bowen 2006). Researchers consider sensitizing concepts as interpretive devices and starting points for qualitative research studies (Bowen 2006; Patton 1990). Charmaz (2003, p. 259) explains that sensitizing concepts are "those background ideas that inform the overall research problem" yet cautions that "we may use sensitizing concepts *only* as points of departure from which to study the data." Moreover, as these concepts can be used to gain a broader understanding and appreciation of social phenomena (Bowen 2006, p. 8), they lend themselves to a study of professional culture. Accordingly, Hall's (2005) "Culture of Professions," "Professional Socialization and Identity Formation," and "Symbolic Interactionism" were selected to guide our coding process and provide a structure for analysis and interpretation of our findings. In this way, we were able to examine professional culture in the context of an FHT, which was an elemental concept that was not considered in the primary study. To attain unbiased results, we followed Hsieh and Shannon's (2005) suggestion of using an audit trail during data analysis to track the progression of the coding frame and the evolution of coding categories.

Initially, all transcripts were read and instances that appeared representative of our sensitizing concepts *and* professional culture in particular were identified and categorized according to predetermined

codes. New codes were created if text could not be categorized within the preliminary coding frame. As analysis progressed, subcategories emerged that were condensed, revised, and refined into lower-order and higher-order codes according to thematic content. The categorizations and wording of each thematic code were considered and further developed, and any discrepancies were reflected upon and continually referred back to particular illustrations in the transcripts so as to be consistent. Themes were examined to determine whether illustrations of professional culture (captured in the data) constituted barriers to interprofessional collaboration in the FHT and, if so, in what ways. Results were analyzed to determine whether professional culture influenced FHT culture and, if so, how this occurred.

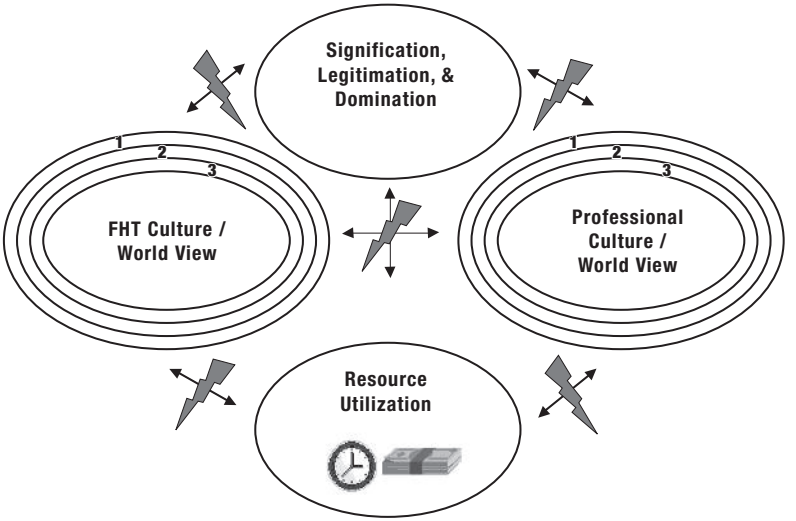
Findings

Analysis of the transcripts highlighted three domains that reflected professional culture in the context of an FHT, which we termed (1) professional culture (professional world view), (2) FHT culture (FHT world view) as related to structure and process, and (3) resource utilization. Further examination of the data exposed the interconnectedness of these domains, and it became apparent that impacts were felt among and between the cultures. Woven within this illustration was the constraint of resources on professional culture and FHT culture (see Figure 2). The data presents an image of tension between professional culture and FHT culture (as one world view crashes against another). Clashes of interprofessional cultures (i.e., the cultures of medicine, pharmacy, and nursing) were observed, together with a tension between professional cultures and FHT culture. Accordingly, a study of professional culture on an FHT cannot be divorced from an examination of FHT culture. Within each domain (professional culture, FHT culture, and resource utilization) major subthemes emerged which are discussed below.

1. Professional Culture (Professional World View)

Analysis of the focus group transcripts brought to light three major themes from which professional culture (professional world view) was evidenced among participants: (1) The recognition of professional hierarchies, (2) notions of professional responsibility, and (3) upholding professional values. These themes were indicative of professional culture and an individual's professional world view and ultimately how the various health care professionals would collaborate with others on the team.

Figure 2. Preliminary Framework to Examine Professional Culture in FHT Context



FHT Culture/World View	Professional Culture/World View
1. Societal Culture	1. Societal Culture
2. Historical Context & Organizational Environment	2. Historical Context & Professional Environment
3. Operating and Cultural System	3. Operating Practice & Professional System
* Subsumed within are the social beings' individual identities	* Subsumed within are the social beings' individual identities

i. Recognition/Perception of Professional Hierarchies

The recognition of a distinct professional hierarchy on the FHT emerged strongly in all focus groups, and these hierarchal perceptions create tension and uneasiness among the practicing health care professionals. For example, one allied health care professional explains:

> Some physicians are more willing to share their power with other members of the team, where others really, they really feel that they need to maintain control and they see other team members as being… no—not subservient because that is the wrong word, but as being underneath—that they are supervising other members of the team (Allied2-FG2).

While this FHT professional clarifies the choice of words, the notion of being underneath or in need of supervision still reflects subservience and suggests that some professionals have yet to embrace the egalitarian relationships espoused by the IPC mandate. Moreover,

this quote suggests that allied health care professionals recognize a professional hierarchal ladder where the physicians are the ones who are looking down on them from their supervisory positions. This notion of power and authority is similarly reflected by a physician:

> I think that if a physician says something to me, I, it wouldn't enter my mind to say "you're sort of overstepping your role or you are sort of outside of your practice" um, but if some other professional said something to me, somehow that comes into that equation (Physician1-FG5).

This notion of power and control is directly related to whose skill sets are perceived to be valued more. Lack of knowledge surrounding the competencies of team members suggests that health professionals may be less willing to share responsibility and trust their fellow team members. Interestingly, the concept of a football game arose in virtually all focus groups, whereby it was recognized that health care professionals had to work together as a team to be able to perform well on "game day." However, one physician's caveat to this point was that "in some places the quarterback needs to run with the ball, and he just has to do it himself" (Phys3-FG1). In this light, it appears that some professionals may pay lip service to the ideals of IPC and support it abstractly, but in reality there are certain times when they may feel exempted from working collectively as a team.

For most physicians, the tension and confusion surrounding what they perceive to be the hesitancy of allied health care professionals taking on more responsibility in the delivery of patient care related to the perceived hierarchal relations on the health care team:

> We dance around the diplomatic roles of not offending the physician or saying things that they might find offensive or change ownership, and to me that is a reflection of limitations of authority…that people think that they have to act in a certain way in order to engage the physician…and it is that hierarchy that I think is the major barrier (Physician1-FG5).

The professional hierarchy is internalized by both physicians and allied health care professionals alike. Most allied health care professionals continue to take the word of the physician as gospel, as they perceive a professional hierarchy that prompts them to feel obliged to defer to the physician's authority. Moreover, allied health care professionals defer to physicians because there is uncertainty surrounding who does what and who is accountable for what on the

FHT. When in doubt, the physician is the most likely candidate for accountability.

ii. Professional Responsibility

The notion of a cultural shift that advocates joint responsibility and shared decision making among FHT members calls for health care professionals to be responsible for varying degrees of patient care as they relate to scopes of practice. Accordingly, allied health care professionals will start feeling a greater degree of power as they carry a heightened degree of professional responsibility. While focus group participants are part of an FHT that supports the mandate of IPC and have been provided with some opportunities to begin internalizing an FHT world view that supports egalitarian relationships, this cultural shift has yet to be realized. One physician explained this situation as follows:

> You get socialized in medical school to become the responsible one, and I think that this is changing, but, for example, if I pick nursing as an example, they are cultured to say that the physician is the one who is responsible. So it is interesting because they defer and we accept (Physician7-FG1).

Most participants agreed that for this change to take place, the FHT would have to clearly define who is responsible for what:

> I mean we all have scopes of practice that we have to abide by, and I think that when there are overlapping responsibilities and accountabilities then the important thing is that those need to be clearly articulated. So whether that is your creation of policies or guidelines, but those need to be there so that if there is ambiguity, there is an agreed upon document or policy that people can go to and say that "this is what we decided as a team, and so this is how and why we do things this way," and so most of the documents won't say it is going to be the physician's responsibility for everything, it will probably be the shared responsibility. I think where the discomfort comes in is that it is not clearly articulated, so we are not really sure and we always default to the physician—and where they are most responsible is their accountability (Allied6-FG2).

Along these lines, all health care professionals must be willing to accept equal degrees of responsibility and share all issues related to patient care. For many physicians, this translated into "taking off some of the slack" that they believed was associated with being a doctor. Allied health care professionals would become more accountable and willing to share the brunt of responsibility if the team

would move beyond the perceived power structure towards power sharing. In theory, this sounds like an ideal solution; however in practice, physicians still feel ultimately responsible:

> I don't like the term gatekeeper, but we were on the hook, and I think that at the present time we are still on the hook. And you know when a lab test, like last night I get phoned about an urgent lab test and I was the one who was phoned and not members of the team. It was me—and that still persists (Phys2- FG1).

> Well, they [physicians] are the ones who sign on the bottom line, and it may be the pharmacist who fills it out and says "hey" and calls them back and says it is wrong, but I still think that most of the responsibility is still on the physician. That is how I feel (Phys3-FG1).

Most allied health care professionals acknowledge that physicians have traditionally held the power in medicine, but concede that all team members must be prepared to take on more responsibility:

> I would say that normally the physician has been the gatekeeper and the holder of power, and more of a one-on-one. So there has to be a willingness of all of the partners on the team to share the power, I think (Allied2-FG2).

While most participants recognized that a cultural change is on the horizon, uncertainties surrounding professional responsibilities and accountabilities constitute real barriers to successful collaboration. Physicians commented on the fact that qualities of professional accountability and responsibility guide their work, whereas other team members do not embrace these qualities to the same degree, especially when confronted with a medical challenge/dilemma. One physician commented:

> Just on the issue of accountability—when everything goes well it is "our patient," when things go sour it is "your patient." That is what I hear, and that is what I am talking about. And I think the teams should be prepared to say that it is still "our patient" when things are not going well (Physician2-FG1).

iii. Professional Values

While being socialized and trained into the culture of a profession, health care professionals internalize professional responsibilities and values which reflect the larger mandate of the profession. In medicine, one of the most profound values is that of the doctor-patient relationship. As health care settings evolve towards shared care, this central relationship could potentially be lost as other health care

professionals forge relationships with the patient. This concern was echoed by physicians:

> One of the principles of family medicine is the relationship that you have with the patient. Do we jeopardize that in some way with collaborative practice? What keeps me hooked into the patient? (Physician2-FG1).

> We talk about ownership and the patients feel like they own you and you own them, and it is "MY PATIENT!" I am going to go see "my patients" and you sort of develop that relationship and that needs…it is a culture change that is a big barrier, and it is very hard to go through that frame of mind (Phys7-FG1).

Most of the allied health care professionals also recognized physicians' sense of patient ownership, which they perceived to be a barrier to collaborating:

> I see the physicians not wanting to share, and there are certain physicians who will hold on to "my patient," and don't want to have anyone else involved (Allied4-FG2).

When values, beliefs, and ways of doing and knowing are so engrained on the cognitive maps of these health care professionals, professional culture may not map perfectly onto FHT culture. Accordingly, one must not assume that IPC will "just happen," as one allied health care professional observed:

> I think it has to be a coordinated, conscious effort. I don't think that you can just put a bunch of people together and expect them to work the best. It has to be facilitated and there has to be time set aside, and there have to be opportunities for people to get to know other peoples' roles and voice their thoughts on how inefficiencies can be improved. It doesn't just happen, to be efficient (Allied6 -FG2).

2. FHT Culture (FHT World View)

Analysis of the focus group transcripts also revealed the manner in which professional culture was impinged upon by the FHT culture. In virtually all focus groups, participants identified elements of FHT culture as being the: (1) FHT environment; (2) various structures and processes that had been clearly defined, established, and were successfully implemented; or conversely (3) various structures and processes that had not been clearly defined nor were they successfully implemented.

i. FHT Environment

Most participants acknowledged that the FHT had cultivated a culture

that recognized and supported the mandate of IPC. As an organization, the FHT had hired a variety of health care professionals with complementary skill sets and practices who were expected to work together as a team. Yet participants commented that the structure of the working environment does not facilitate collaboration given that the FHT has a revolving door of learners, which makes relationship building and understanding of roles and responsibilities that much more challenging:

> It is very hard to have collaborative patient care that is good if people are only in the office half a day a week, and I am going to put that out "that you have to be around to do this" (Physician2-FG3).

> Once you go and have turnover of team members, that speaks to the sustainability of healthy teams, and you never know who is going to be that person in the office two [doors] down from you—who is supposed to be doing that role? Who is in your space every two to three years? How does an effective team continue to function (Physician6-FG1)?

In addition to the turnover of learners who are present on the FHT for a finite amount of time, there are other health care professionals whose time is limited on the FHT. Participants in this study frequently spoke of their frustration of not being able to contact or find a particular health care professional who only worked two or three days per week. Consequently, while FHTs have attempted to create an environment that fosters the IPC ideal, the structure of the environment does little to support a collaborative culture.

ii. Structures and Processes Are Clearly Defined, Established, and Have Been Successfully Implemented

In order for FHTs to cultivate a culture that endorses the mandate of IPC, certain structures and processes must be put in place to facilitate cultural transmission. Values, various artifacts, organizational beliefs, and codes of behaviour among other elements constitute the culture of the FHT. Most participants recognized the value of IPC in the context of patient care and indicated that this notion had been imparted to them through the culture of the FHT. Moreover, many of the participants acknowledged that being collaborative translated into a sharing of power. Along these lines, participants have seemingly internalized and oriented themselves to FHT culture and the IPC ideal. One participant was quite willing to demonstrate his awareness of collaboration. When asked to define collaborative practice his

remark was, "I'm smiling because I know the exact definition" (Physician7-FG1). While his comment did not indicate whether he collaborated with others or not, it did indicate that the FHT had instilled health care professionals with an awareness and appreciation of IPC. In fact, most participants reported that the interprofessional collaborative concept was not a new phenomenon in the FHT environment. Being associated with an academic teaching hospital had exposed them to continuing education sessions and various team-building initiatives. One physician recognized the importance of having a complementary array of skill sets on the team:

> You tap skill sets that are not available in other disciplines. For example, an occupational therapist may have knowledge that physicians don't regarding functional abilities, and physicians historically are the quarterback—as far as they are the only ones who write orders in the hospital chart, and we are moving and we are changing, but certainly we are tapping knowledge that physicians don't have (Physician3-FG1).

The real life experiences modeling IPC, while limited, were enthusiastically supported by professionals. Participants in the study frequently referred to these opportunities as being rewarding, valued, and appreciated among team members. For most participants, these infrequent occasions presented them with a chance to become acquainted with the roles and scopes of practice of other team members. One physician shared his recent experience with a pilot program, championing the learning potential when health care professionals adopt a synergistic approach to patient care:

> We have just started this initiative of doing an interprofessional rapid consult where people could come with the patient and go through an hour-and-a-half or two-hour session, and, as I did in the first day, learn more about insulin distribution systems or dietary records or things in an hour and a half that I couldn't have learned otherwise in many other forums (Physician1-FG5).

iii. Structures and Processes Are Not Clearly Defined nor Have They Been Successfully Implemented

As evidenced, when certain structures and processes are in place, the FHT clearly cultivates an environment that supports the IPC ideal. Unfortunately, other elements impede health care professionals' ability to successfully collaborate. Most participants commented on the fact that collaborative meetings to discuss patient care were not a regular

occurrence on the FHT. Moreover, there had been no designated clinic time to learn individual roles, competencies, or skill sets. Once again the analogy of the football team was raised when one participant stressed the importance of having time to practice and learn the skill sets of others on the team:

> Anyone who has played sports will tell you—who has a job in the professionals—that when they go to practice, it's a football team from Monday to Friday, and that is their work. That is where they put in all their effort and that is the hard part, and Sunday—game day—is the fun time. That is their play time. So I think for this to work, the practice will be for all of us to get to understand the roles and capabilities of each other and that will be work, and only when those barriers are overcome will we be able to change the culture (Physician8-FG1).

For many participants the need for understanding roles and scopes of practice, in addition to clarifying who is responsible for what, was deemed not only important but essential:

> There has to be time devoted to clinic time together and understanding each other's roles and getting that comfort level and that somehow needs to be built in (Physician2-FG5).

Despite the fact that this FHT was affiliated with an academic teaching hospital (which is indicative of a highly functioning complement of academically inclined health care professionals), and that FHT members were exposed to various collaborative practice learning opportunities, one participant acknowledged that he was unaware of what the other health care professionals' scopes of practice were, even though they worked together:

> I think the other piece is how much do we know? I actually can say honestly I don't know your scope of practice…in order to define that we need to have an opportunity to dialogue and really know (Physician1-FG3).

Participants in the study repeatedly spoke of the lack of designated time to learn roles and scopes of practice, which hindered their ability to successfully collaborate. Moreover, most participants were of the opinion that the FHT required a coordinated approach and formal definition that clearly set out who was responsible for what, as well as the various skill sets:

> We all have scopes of practice that we have to abide by, and I think that if there are overlapping responsibilities and accountabilities then

> the important thing is that those need to be clearly articulated. So whether that is your creation of policies or guidelines, but those need to be there so that if there is uncertainty—there is an agreed upon document or policy that people can go to (Allied6-FG2).

As the FHT lacked an effective process in conveying roles and scopes of practice, this impinged upon health care professionals' ability to communicate and collaborate with one another. This was demonstrated in the previous domain whereby allied health care professionals acknowledged feeling uncomfortable collaborating with other health care professionals due to the fact that their roles and skills sets were not clearly articulated or understood. The previous conditions illustrate how barriers associated with structure and processes of the FHT have a negative affect on FHT professionals and tend to evoke feelings of tension and uneasiness. As a result, health care professionals call upon their professional world views as frames of reference to navigate the murky waters. This suggests that for these FHT professionals, their professional world view is stronger than the imparted FHT world view.

3. Resource Utilization (Time and Money)

Lastly, resources were identified as having the propensity to create tension among FHT members. Resource utilization impinged upon FHT culture and professional culture, and in some ways shaped how health care professionals viewed collaboration. Analysis of the data emphasized two themes within the domain of resources which reflected: (1) Playing the Game, and (2) Compensation Tension.

i. Playing the Game

Across the focus groups there was a theme that reflected the fact that health care professionals acknowledge and internalize the value of IPC and that as a team they should work together to provide patient-centred care. However, their motivation to collaborate was not strictly based on the steadfast belief that their complementary skill sets would improve patient care. Physicians in particular shared that their motivation to collaborate was in part due to resource allocation and the money that the Canadian government has infused into the creation and implementation of FHTs. Along these lines, physicians recognize that in order to be paid, they must play the game, which currently pays to provide collaborative patient-centred care:

> There are rules for engagement. And I think that for those of us who have been in an organization long enough, you know that every 5–10

> years there are different themes that are promoted either at the hospital level, at the provincial level, at the federal level, and as physicians—as players—within that organization, if you are going to survive, and thrive, and grow and expand and continue to provide the care that brought us here in the first place, you have to know where the resources are (Physician6-FG1).

ii. Compensation Tension

Resource tension was also evidenced across the focus group transcripts as participants discussed the fee structure of the varying health care professions on the FHT, and the manner in which they were compensated. Both physicians and allied health care professionals explained that the manner in which physicians are paid poses a limitation to fostering a collaborative culture on the FHT. Physicians remarked that when indirect care or team-building initiatives are not rewarded financially, their time could be better spent elsewhere. Consequently, physicians remove themselves from the FHT culture and are more likely to retreat to the comfort zone of their professional world view where they are rewarded for what they do best. One allied health care professional explained the situation as follows:

> Some of it has to do with sharing of responsibility and the fee structure and how that works, and that is the big barrier that we did not talk about. But when physicians have to bill to earn their money, that does impact us (Allied2-FG2).

> I think money is the barrier—that our current system of rewarding team members is a barrier (Allied1-FG5).

Most allied health care professionals commented on the fact that they perceived compensation to be a barrier to collaboration and explained that because physicians had to bill to earn their money, it was less likely that they would voluntarily participate in IPC meetings. On a larger level, this distinction in how FHT professionals are compensated transmits a silent message throughout the team that the FHT, as an organization, does not value all health care professionals in the same manner—contrary to what the culture espouses.

Another issue that was raised in focus groups had to do with busy schedules and a perceived lack of time to engage in planned IPC team-building sessions. The old adage that time is money seemingly holds true in an FHT environment—predominantly for physicians. The demanding schedules of health care professionals was highlighted in

the transcripts among physicians who were concerned about taking time out of their day to engage in indirect care and team-building efforts:

> Part of my biggest frustration with being a physician is that idea that you spend time on the phone, and go to team meetings, and talk about patients…everyone who is there, excuse me for that, but the majority of you [allied health care professionals] on salary—that is all part of your day, but it is NOT part of mine. So I am taking time off where I am not actually earning income and where my expenses are going to attend that. So, I mean it is a crude example, but if there was support there for indirect care, that doesn't necessarily need to be a paycheque at the end of the month, it could be support to attend a conference, it doesn't need to be support to pay for the whole conference, but support in some form. So support in that way to support my involvement in indirect care would be to me on a smaller scale one way to begin to approach this (Physician2-FG5).

Consequently, physicians recognize a value attached to providing direct patient care (internalized as a result of their professional training and socialization) and feel their time should be similarly rewarded when being present for, or participating in, team-building initiatives or indirect care. This distinction of being different from other members of the team also fails to support the equal-status basis requisite for successful and integrated collaboration.

Discussion

Our analysis provides a contextual description of professional culture in the context of an interprofessional FHT and demonstrates how cultural barriers may be used to understand interprofessional collaborative barriers (Austin et al. 2007; Hall 2005; Irvine et al. 2002; Hong 2001). Moreover, our findings demonstrate how FHT culture and resource utilization has the ability to facilitate or challenge the successful fruition of the collaborative ideal in an FHT environment where diverse professional cultures are present. In addition, we provide insight on the manner in which health care professionals navigate between their professional world view and their FHT world view, and the circumstances under which they may choose one world view over another. To overcome these barriers, we need to understand the distinctions between and among the unique health care professions, but we must also understand the context in which these play out in an interprofessional health team environment. Our framework (see Figure 2) begins to tell this story, provides a base from

which others may investigate this area further, and contributes to the growing knowledge base on IPC.

While the participants in this study were certainly well versed in what interprofessional collaboration was and how it should play out on the team, FHT members were still unsure of who was competent, had authority, and could provide assistance when needed (Ben-Syra and Szyf 1992). Moreover, they acknowledged that a cultural change had not been seen on the FHT. It still functioned in a hierarchal manner with physicians holding the greatest responsibility and making decisions when they saw fit. Data suggest that physicians continue to retain the highest rung on the professional hierarchy. Imparted through cultural transmission, physicians' professional training embeds concepts of power, authority, and hierarchy in the identities of many health care professionals. These notions are then mirrored in their social worlds (Austin et al. 2007; Hall 2005). This professional hierarchy is further sustained when structure and processes of the FHT operating system have not been clearly established.

Bloor and Dawson's (1994) provisional model for examining professional culture in an organizational context proved useful for setting the foundation to examine professional culture in the context of an FHT. Taking into consideration the societal, historical, as well as local environment in which culture exists, we understand that these factors cannot be divorced from this study, and while our findings did not speak specifically to these facets of professional culture or FHT culture, we acknowledge that they exist and should be included in any working framework. While Bloor and Dawson's (1994) model was helpful in guiding our own research investigation, it failed to capture the tension and unease that was seen when FHT and professional world views collided. Moreover, elements of resource utilization were also found to fuel tension among these two domains. Accordingly, our study acknowledges that professional culture is not the only barrier to achieving IPC but that FHT culture also seems to play an important role in this tension, as evidenced by our framework (see Figure 2).

Components of signification, legitimation, and domination were evidenced in our findings. Health care professionals internalize FHT culture through their collection of frames of reference (signification), which take place through exposure to and experiences with the various policies, values, and beliefs of the FHT, as well as interactions among group members. As they amass frames of reference on

their FHT world view, they begin to internalize the group beliefs (legitimation) and their placement on the FHT hierarchal ladder (domination). Physicians' interests have become part of the taken-for-granted social reality that structures health care circles. As this manifestation best serves their interests, it is reproduced (Mumby 1988, p. 67 as cited by Bloor and Dawson 1994, p. 292) and, consequently, we see this mirrored within the FHT setting. Patterns of signification, legitimation, and domination were evidenced in the findings even though health care professionals had adopted frames of reference that suggested they support the IPC ideal. They outwardly recognize its importance and value and understand that it is the future of health care practice. Unfortunately, their FHT frames of reference have not been fully developed to the point where they can call upon them to make sense of their work environments in the face of uncertainty. This causes them to fall back on their professional world views (where their frames of reference are fully developed) to enable them to make sense of their surroundings through their uniprofessional lens.

When conflict and uncertainty are present within FHT culture, the individual's professional world view emerges. During tense times, health care professionals abandon their evolving FHT world views and re-engage their own professional world views as a way of making sense of the particular situation. Consequently, professional dominance and professional hierarchies are perpetuated, maintained, and reinforced, as they are mapped on each health care professional's world view. The experience of discomfort triggers professionals to retreat to the safe zone of their earlier frames of reference (their professional world view), which imparted notions regarding professional dominance and hierarchy. A retreat by health care professionals to their safe zones is analogous to a form of "code switching," whereby they switch from their FHT world view into their professional world view, allowing them to comfortably coexist with and understand the other members of the FHT. In this way, the individual's professional world view is seemingly stronger than that of her/his FHT world view. Creating a culture that espouses egalitarian relationships will be a challenge. The dominance of professional world views, which have been constructed with a clear understanding of power and hierarchy, will be difficult to overcome.

The FHT has yet to provide its health care professionals with enough interprofessional collaborative experiences to be able to map onto

their FHT world views. In addition, the FHT has not established structures or processes that clearly define who does what and who is responsible for what. Additionally, it has not yet fully articulated the various scopes of practice among team members, and consequently professional world views are maintained. As the structures and processes of the FHT operating system have yet to cultivate egalitarian relationships among team members, allied health care professionals continue to defer to the physician who they perceive to hold the ultimate responsibility and accountability. In turn, physicians do little to challenge this situation. While physicians might want allied health care professionals to take more initiative and demonstrate professional responsibility and accountability, they are not willing to provide them with the opportunities to demonstrate their capabilities because they are unaware of the competencies and skill sets of many of their team members. Moreover, they wish to retain their ability to run with the ball when they deem necessary. Although some physicians may feel the need to exercise full power over their patients and practice, given the strong mantra of IPC, these opportunities to quarterback the team may diminish as others on the team are thrown the ball. Ultimately, the success of fostering a true IPC environment will need to reflect an FHT operating system and organizational environment that cultivate frames of reference for these health care professionals when navigating their FHT world view.

Our findings demonstrate a prominent medical voice that is attempting to negotiate its presence on a team that supports the IPC ideal. The prominent voice of medicine depicted in the findings is one that is situated against the backdrop of an interprofessional milieu struggling to define itself in an environment where the professional world view continues to hold sway. While other health care professions' (i.e., nursing and pharmacy) training environments cultivate learning to work as a team, communication, and sharing, medicine is one that imparts independence and competition (Austin et al 2007; Hall 2005). In this way, nurses and pharmacists may have less difficulty internalizing FHT culture as it maps clearly onto their previous professional frames of reference. Physicians may have a harder time doing this (as their frames of reference seemingly do not map easily onto the collaborative FHT culture), and accordingly hold steadfast to their medical frame of reference which imparts dominance, responsibility, self-confidence, independence, and competition (Austin et al 2007; Hall 2005).

Results suggest that academicians, educators, policy-makers, and health team managers who are researching, conceptualizing, and/or implementing mechanisms to impart IPC skills to health care professionals should incorporate components that explicate the cultural distinctions between and among professions (Austin et al. 2007), but also illuminate how these play out in the context of an interprofessional milieu. In terms of FHTs, the operating system and organizational environments must be sensitive to the effects of professional culture (Barker et al. 2005) and provide health care professionals with time, as well as opportunities, to understand and appreciate each other's unique cognitive maps (Hall 2005, p. 190). The uniprofessional foci of most training programs for health care professions also suggests that team members will need experiential opportunities to learn *how* to work together (Clark 2002 as cited by Hall 2005, p. 193). Moreover, the structure and processes of the FHT must not only endorse but implement mechanisms and clear methods (i.e., implementing team development meetings; generating formalized protocols that clearly articulate scopes of practice, skills, authority; clarifying who does what and who takes responsibility; and providing opportunities to develop team relationships that foster trust and mutual understanding of fellow team members' competencies and skills) that prevent health care professionals from retreating to their safe zones. If this can be accomplished, the tension that exists between professional and FHT cultures may be lessened and the IPC ideal may be more fully adopted.

Limitations and Research Recommendations

For the purpose of this paper, we have prioritized the most salient limitations that warrant discussion, cognizant that others exist. First, the study focused on professionals working at one FHT (affiliated with a large academic teaching hospital, in a busy urban centre of Ontario) and, therefore, participants represent one FHT in one geographical location. Nonetheless, as transcripts were analyzed, tensions among the health care professions (which stemmed from FHT culture, professional cultures, and resources) arose repeatedly in each focus group discussion, reflecting tangible issues that are faced by FHT members in the provision of collaborative patient-centred care. Accordingly, our study provides a depth of understanding to professional cultures and FHT culture, and generates themes that might be applicable in a broader context. These themes might be applied in other FHT settings even though the idiosyncrasies of each FHT may be different.

Second, we believe that we maximized the advantages of using secondary data, in that the primary data were collected to examine a facet of the interprofessional collaborative process. Accordingly, a secondary analysis of the data, to examine professional culture in the context of IPC, was tangential to the original purpose of the data collection process. Moreover, as the data collection process was supervised, lead, and carried out by one of our investigators (JB), we are confident in the accuracy of the primary source data. Despite these initiatives, however, we were constrained by the parameters of the original primary study. In order to preserve the confidentiality and anonymity of the focus group participants, these participants were assigned identifying labels of "physician" or "allied" (allied health care professional), which ultimately limited our interpretation of the data, as these identifiers did not fully capture nor adequately reflect the true diversity of professional cultures found on the FHT.

Literature clearly presents cultural differences among health care professions, and the distinction of these differences within the group labeled "allied" (allied health care professionals) could not be captured to the extent that the data would have allowed. In this way, the data clearly identify the cultural differences and similarities between the "physician" and "allied" groups. However, we recognize that that the similarities and differences among the different health care professionals subsumed under the "allied" group are undoubtedly correspondingly distinct (Irvine et al. 2002).

Finally, this qualitative study does not place variables of social significance such as race, class, ethnicity, and gender under a lens. These variables constitute additional overlapping social identities that health care professionals must negotiate and these, too, ultimately inform their distinct cognitive maps and overall world views (Austin et al. 2007). To this end, while certainly not exhaustive, our study lays the groundwork for tomorrow's research on professional culture in the context of IPC and evolving FHT culture. Moreover, the study is potentially transferable across other FHT settings and establishes a springboard for further research.

Conclusion

In many ways the history of health care in Canada reflects the evolution of health professions (Beales and Austin 2006) and, arguably one might add, the evolution of professional cultures. As we see a new chapter being written that endorses patient-centredness and

teamwork, what remains unclear is how distinct professional cultures will respond to this wider cultural shift in the delivery of health care. This paper offers fresh insight on the concept of professional culture and evolving FHT culture and does so in the context of IPC. A framework is provided that may help guide future research and empirical studies to garner a greater understanding of professional culture, providing academicians with the bricks and mortar to build on the foundation that has been laid today.

Our findings from this exploratory study suggest that the connections between professional culture, FHT culture, and IPC are worthy of further investigation. While the limitations identified above suggest that there are further strides to be made in this evolving area, our research provides an initial lens through which academics may begin to visualize and conceptualize how diverse professional cultures will respond to this wider cultural shift in the delivery of health care. Our framework begins to bridge the gaps of our current knowledge and understanding of professional culture and evolving FHT culture, and starts to shed light on the complex web of variables that interconnect professional cultures and FHT culture, and ultimately influence collaboration on the team. In this light, while professional culture has been identified as a barrier to IPC in the literature, we believe that with a better understanding of how professional culture shapes and is shaped by FHT culture, we—including educators, institutional leaders, government officials, hospital leaders, and practitioners—will be better equipped to develop, support, and sustain the IPC ideal in the context of providing patient-centred care.

Acknowledgements

The authors would like to extend warm thanks to the [name of institution anonymized] Allied Health Fund (with matching funding from the Department of Family and Community Medicine, Hospital A [name of hospital blinded for anonymity], the Ontario Graduate Scholarship Program, and the Lupina Foundation (Comparative Program on Health and Society) at the Munk Centre for International Studies at the University of Toronto for providing their support. In particular, the authors would like to thank Rishma Walji, Zahava Rosenberg-Yunger, and Linda Muzzin for their reviews and constructive feedback, Joseph Cosentino for his valuable edits and suggestions, and our focus group participants.

Notes

1. Signification relates to the manner in which organizational members call upon schemes of interpretation (as a way of understanding and making sense of organizational life) to internalize the mandate of the organization, what their role is, what their place is, and how to speak and interact with other members within the organization. These schemes are learned through socialization with other members of the organization (Van Maanen and Schein 1979 as cited by Bloor and Dawson 1994, p. 278), and frame how artifacts, patterns of behaviour, values, and basic organizational assumptions are respected by the organization.

2. Legitimation relates to the manner in which dominant groups have their values, interests, and goals accepted as being legitimate within an organization (Bloor and Dawson 1994, p. 279).

3. Domination refers to the formation of pecking orders within an organization, which identify who does what (Bloor and Dawson 1994, p. 278).

4. Berg (1998) contends that "when focus groups are administered properly, they are extremely dynamic. Interactions among and between group member stimulate discussions in which one group member reacts to comments made by another…The resulting synergy allows one participant to draw from another or to brainstorm collectively with other members of the group" (p. 101).

5. The first two focus groups were homogeneous groups (Group 1, physicians only; and Group 2, allied health care professionals only). The next three focus groups were heterogeneous groups. The homogenous focus groups provided the participants an opportunity to be open about their attitudes, experiences, opinions, and beliefs among their peers without feeling influenced by perceived power structures. The heterogeneous groups provided the participants an opportunity to share experiences, ideas, and opinions, and brainstorm together in a synergistic forum.

References

Aiken, L., D. Sloan, and J. Sochalski. 1998. Hospital Organization and Outcomes. *Quality in Health Care* 7 (4): 222–26.

Atwal, A. 2005. Do All Health and Social Care Professionals Interact Equally: A Study of Interactions in Multidisciplinary Teams in the United Kingdom. *Scandinavian Journal of Caring Sciences* 19 (3): 268–73.

Austin, Z., P. Gregory, and C. Martin. 2007. Negotiation of Interprofessional Culture Shock: The Experiences of Pharmacists Who Become Physicians. *Journal of Interprofessional Care* 21 (1): 83–93.

Barker, K., C. Bosco, and I. Oandasan. 2005. Factors in Implementing Interprofessional Education and Collaborative Practice Initiatives: Findings from Key Informant Interviews. *Journal of Interprofessional Care* 19, Suppl. 1: 166–76.

Beales, J., and Z. Austin. 2006. The Pursuit of Legitimacy and Professionalism: The Evolution of Pharmacy in Ontario. *Pharmaceutical Historian* 36 (2): 22–27.

Ben-Syra, Z., and M. Szyf. 1992. Status Inequality in the Social Worker-Nurse Collaboration in Hospitals. *Social Science and Medicine* 34 (4): 365–74.

Berg, B. 1998. *Qualitative Research Methods for the Social Sciences*, 3rd ed. Toronto: Allyn and Bacon.

Bloor, G., and P. Dawson. 1994. Understanding Professional Culture in Organizational Context. *Organization Studies* 15 (2): 275–95.

Bowen, G.A. 2006. Grounded Theory and Sensitizing Concepts. I*nternational Journal of Qualitative Methods* 5 (3): 1–9.

Charmaz, K. 2003. Grounded Theory: Objectivist and Constructivist Methods. In *Strategies for Qualitative Inquiry*, 2nd ed., edited by N.K. Denzin and Y.S. Lincoln, 249–91. Thousand Oaks, CA: Sage.

Clark, P.G. 2002. Evaluating an Interdisciplinary Team Training Institute in Geriatrics: Implications for Teaching Teamwork Theory and Practice. *Educational Gerontology* 28 (6): 511–28.

Cott, C. 1998. Structure and Meaning in Multidisciplinary Teamwork. *Sociology of Health and Illness* 20 (6): 848–73.

Covey, S. 1989. *The Seven Habits of Highly Effective People*. New York: Simon and Schuster.

Cresswell, J.W. 1998. *Qualitative Inquiry and Research Design: Choosing among Five Traditions*. Thousand Oaks, CA: Sage.

EICP. Enhancing Interdisciplinary Collaboration in Primary Care (EICP) Initiative. 2005. *Health Human Resources Task Group Snapshot*. Ottawa: Enhancing Interdisciplinary Collaboration in Primary Care (EICP) Initiative.

Goodenough, W.H. 1970. *Description and Comparison in Cultural Anthropology*. Chicago: Aldine.

Hall, P. 2005. Interprofessional Teamwork: Professional Cultures as Barriers. *Journal of Interprofessional Care*. 19, Suppl. 1: 188–96.

Health Canada. 2004. *Interprofessional Education for Collaborative Patient-centred Practice*. Ottawa: Health Canada.

Hong, G.Y. 2001. Front-line Care Providers' Professional Worlds: The Need for Qualitative Approaches to Cultural Interfaces. *Forum: Qualitative Social Research* 2 (3): 6.

Hsieh, H.F., and S.E. Shannon. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15 (9): 1277–88.

Irvine, R., I. Kerridge, J. McPhee, and S. Freeman. 2002. Interprofessionalism and Ethics: Consensus or Clash of Cultures? *Journal of Interprofessional Care* 16 (3): 199–210.

Kahn, N. 2004. The Future of Family Medicine: A Collaborative Project of the Family Medicine Community. *Annals of Family Medicine* 2 (1): 3–32.

Kitzinger, J. 1995. Introducing Focus Groups. *British Medical Journal* 311 (7000): 299–302.

Krueger, R. 1994. *Focus Groups: A Practical Guide for Applied Research*, 2nd ed. London: Sage Publications.

Leathard, A. 2003. *Interprofessional Collaboration: From Policy to Practice in Health and Social Care*. London: Brunner-Routledge.

Miller K.L., S. Reeves, M. Zwarenstein, J. Beales, C. Kenaszchuk, and L.G. Conn. 2008. Nursing Emotion Work and Interprofessional Collaboration in General Internal Medicine Wards. *Journal of Advanced Nursing* 64 (4): 332–43.

Mumby, D.K. 1988 *Communication and Power in Organizations: Discourse Ideology and Domination*. Norwood, NJ: Ablex.

National Steering Committee on Patient Safety. 2002. *Building a Safer System: A National Integrated Strategy for Improving Patient Safety in Canadian Health Care*. Ottawa: National Steering Committee on Patient Safety.

Oandasan, I., D. D'Amour, M. Zwarenstein, K. Barker, M. Purden, M. Beaulieu, S. Reeves, L. Nasmith, C. Bosco, L. Ginsburg, and D. Tregunno. 2004. *Interdisciplinary Education for Collaborative Patient-centred Practice*. Ottawa: Health Canada.

Ontario Ministry of Health and Long-Term Care. 2005a. *Guide to Collaborative Team Practice*. Toronto: Ministry of Health and Long-Term Care.

———. 2005b. Supporting Providers to Work Collaboratively. In *Laying the Foundation for Change: A Progress Report on Ontario's HHR Initiative*. Toronto: Ministry of Health and Long-Term Care.

———. 2005c. McGuinty government announces creation of 31 new family health teams; Second wave of teams will provide thousands of Ontarians with more access to doctors and nurses. News Release Communiqué, December 9. Toronto: Ministry of Health and Long-Term Care.

Opie, A. 1997. Thinking Teams Thinking Clients: Issues of Discourse and Representation the Work of Health Care Teams. *Sociology of Health and Illness* 19 (3): 259–80.

Papoushek, C, and J. Beales. The Identification and Development of a Collaborative Practice Model in Managing Type 2 Diabetes. Forthcoming manuscript.

Patton, M. 1990. *Qualitative Evaluation and Research Methods*, 2nd ed. Newbury Park, CA: Sage.

Phillips, J. 1999. Interprofessional Collaboration for Low Incomes and Immigrant Families: A Case Study. PhD diss., University of Texas at Austin.

Reeves, S. and S. Lewin. 2004. Interprofessional Collaboration in the Hospital: Strategies and Meanings. *Journal of Health Services Research Policy* 9 (4): 218–25.

Reeves, S., S. Lewin, J. Meyer, and M. Glynn. 2003. *The Introduction of a Ward-based Medical Team System within a General and Emergency Medical Directorate*. London: City University.

Romanow, R.J. 2002. *Building on Values: The Future of Healthcare in Canada* (Final Report. 1-356). Ottawa: Commission on the Future of Health Care in Canada.

Schmitt, M. 2001. Collaboration Improves the Quality of Care: Methodological Challenges and Evidence from U.S. Health Care Research. *Journal of Interprofessional Care* 15 (1): 47–66.

Sorrells-Jones, J. 1998. Practitioner Response. *Journal of Healthcare Management* 43 (2): 203–5.

Strauss, A.L., and J. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounding Theory*, 2nd ed. Thousand Oaks, CA: Sage Publications.

Van Maanen, John, and Edgar Schein. 1979. Towards a Theory of Organizational Socialization. In *Research in Organizational Behavior*, vol. 1, edited by Barry M. Staw, 209–64. Greenwich, Conn.: JAI Press.

Way, D., L. Jones, and N. Busing. February 2000. Implementation Strategies: Collaboration in Primary Care—Family Doctors and Nurse Practitioners Delivering Shared Care. Discussion Paper. Toronto: Ontario College of Family Physicians.

Willumsen, E., and P. Breivik. 2003. Welfare and Education Policy—How Do These Factors Influence Interprofessional Education? The Norwegian Case. In *Interprofessional Collaboration: From Policy to Practice in Health and Social Care*, edited by A. Leathard, 207–309. New York: Brunner-Routledge.

Zwarenstein, M., W. Bryant, R. Bailie, and B. Sibthorpe. 1997. *Interventions to Change Collaboration between Nurses and Doctors*. (Cochrane Review). In the Cochrane Library, Issue 3. Oxford: Update Software; 1997. Updated in The Cochrane Library, Issue 4, 2000.

Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement

Jason M. Fletcher and Steven F. Lehrer

Abstract

While there is a well-established, large positive correlation between mental and physical health and education outcomes, establishing a causal link remains a substantial challenge. Building on findings from the biomedical literature, we exploit specific differences in the genetic code between siblings within the same family to estimate the causal impact of several poor health conditions on academic outcomes. We present evidence of large impacts of poor mental health on academic achievement. Further, our estimates suggest that family fixed effects estimators by themselves cannot fully account for the endogeneity of poor health. Finally, our sensitivity analysis suggests that these differences in specific portions of the genetic code have good statistical properties and that our results are robust to reasonable violations of the exclusion restriction assumption.

1. Introduction

One of the most controversial debates in academic circles concerns the relative importance of an individual's innate qualities ("nature") versus environmental factors ("nurture") in determining individual differences in physical and behavioural traits.[1] For many years, researchers in the social sciences could only examine the relative importance of a multitude of environmental factors on various individual outcomes, as data on genetic variation between individuals was unavailable. Yet, with the decoding of the human genome, this limitation no longer exists, and recent years have been characterized by substantial amounts of research in the biomedical literature examining whether specific point mutations in genetic code (aka

single nucleotide polymorphisms [SNPs]) between dizygotic twins (among other family-based samples) are associated with specific diseases and outcomes. Findings from these studies have not only led to new drug discoveries but also improved diagnostic tools, therapies, and preventive strategies for a number of complex medical conditions.[2] As clinical researchers identify unique genetic bases for many complex health behaviours, diseases, and other outcomes, opportunities arise for social scientists to exploit this knowledge and use differences in specific sets of genetic information to gain new insights into a variety of questions.

In this paper, we exploit differences in genetic inheritance among children within the same family to estimate the impact of several poor health conditions on academic outcomes via a family fixed effects instrumental variables strategy. Understanding the consequences of growing up in poor health for adolescent development has presented serious challenges to empirical researchers due to endogeneity that arises from both omitted variables and measurement error problems pertaining to health.[3] Empirical research that has attempted to estimate a causal link have either used a within-family strategy (i.e., Currie and Stabile 2006; Fletcher and Wolfe 2008a, 2008b; Fletcher 2008) or instrumental variables approach (i.e., Ding et al. 2006, 2009; Behrman and Lavy 1998; Norton and Han 2008; Glewwe and Jacoby 1995), and in general researchers find large negative impacts of poor health on academic outcomes.[4] Our empirical strategy combines both elements and identifies the causal impact of health on education by exploiting exogenous variation in genetic inheritance among both siblings and dizygotic twins.

Differences in genetic inheritance occur at conception and remain fixed between family members at every point in the life cycle, irrespective of all nurture investments an individual faces (even those that occur in utero).[5] Since a great deal of variation in characteristics and outcomes is found within families, exploiting the genetic processes that affect development (but are not self-selected by the individuals themselves) presents a potential strategy to identify differences within families.[6] However, it is worth stating explicitly that this identification strategy relies on assumptions regarding how specific genetic markers affect health and academic outcomes in adolescence. As the biomedical literature has not reached a consensus on how specific genetic markers operate, concerns could exist that, despite no detectable evidence in the biomedical literature,[7] the

specific genetic markers we use in our analysis are not only related to poor health in adolescence but also to genetic factors that directly impact education outcomes. In our analysis, we examine the sensitivity of our empirical results to the degree in which the exclusion restriction assumption is potentially violated, finding that our main results are not sensitive to the plausibility of the instruments at reasonable levels. Since nearly every social, behavioural, and health outcome has a unique genetic basis, this identification strategy can potentially shed light on a large number of questions.[8]

Our empirical analysis reaches three major conclusions. First, we find that the impact of poor mental health outcomes on academic achievement is substantial. Our preferred estimates examine the relationship with a sample consisting only of same-sex dizygotic twins, and they indicate that inattention leads on average to a one standard deviation decrease in academic performance.[9] The significant negative impacts of inattention on academic performance remain large and significant if we examine the relationship using other family-based samples.

Second, we conduct a variety of specification tests which indicate that family fixed effects estimators by themselves cannot fully account for the endogeneity of poor health. This indicates that the commonly observed differences in health and education outcomes between full biological siblings should not be treated as random in empirical analyses.

Third, we find that differences in specific portions of the genetic code have desirable properties to identify the impact of poor health on education within families, as there are statistically significant correlations with each endogenous health variable that are consistent with the biomedical literature. In addition, sensitivity analyses indicate that our results are robust to reasonable violations of the exclusion restriction assumption.[10]

The rest of the paper is organized as follows. In section 2, we provide an overview of the data we employ in the study. We also review the scientific literature linking the genes in our dataset to health behaviours and health outcomes. The empirical framework that guides our investigation and our identification strategy is described in section 3. The empirical results are presented and discussed in section 4. A concluding section summarizes our findings and discusses directions for future research.

2. Data

This project makes use of the National Longitudinal Study of Adolescent Health (Add Health), a nationally representative longitudinal dataset.[11] The dataset was initially designed as a school-based study of the health-related behaviours of 12-to-18-year-old adolescents who were in grades 7 to 12 in 1994/95. A large number of these adolescents have subsequently been followed and interviewed two additional times, in both 1995/96 and 2001/2. To develop our identification strategy, we use a specific subsample of the respondents for which DNA measures were collected during the 2001/2 interview and for which there were multiple family members in the survey. This specific subsample is composed of monozygotic twins, dizygotic twins, and full biological siblings, and includes information on 2,101, 2,147, and 2,275 individuals who completed the survey at each interview point. Excluding those individuals for whom there is incomplete education, health and DNA measures for multiple family members reduces the sample to 1,684 individuals.

The dataset contains information on a number of health conditions, including depression, ADHD, and obesity. Depression is assessed using 19 responses to the Center for Epidemiologic Studies Depression Scale (CES-D), a 20-item self-report measure of depressive symptoms. Items on the CES-D are rated along a four-point Likert scale to indicate how frequently in the past week each symptom occurred (0 = never or rarely; 3 = very often). The sum of these items is calculated to provide a total score, where higher scores indicate a greater degree of depressive symptoms. To determine whether an individual may be depressed, we followed findings from earlier research with adolescent samples (Roberts, Lewinsohn, and Seeley 1991) and use specific age and gender cut-offs. We also use adult-based cut-offs to capture a broader measure of depressive symptoms in our analyses. The primary indicator of childhood ADHD symptoms is taken from an 18-question retrospective rating collected during the third data wave. Since there is evidence that the effects of ADHD may vary by whether the symptoms are of the inattentive or hyperactive type,[12] we examine the effects of these different domains as well as the clinical measure of ADHD of any type. Finally, overweight and obesity are calculated from each individual's self-reported height and weight applied to age and gender specific definitions obtained from the Center for Disease Control.

While concerns may exist regarding the use of self-reports to construct indicators for health measures such as ADHD or obesity, we believe this is a limited concern for our study. Not only are we using an instrumental variables approach, but past research with this data (Goodman et al. 2000) indicates that there is a strong correlation between measured and self-reported height (0.94), and between measured and self-reported weight (0.95). There is no evidence that reporting errors are correlated with observed variables such as race, parental education, and household income. Further, several reviews have concluded that childhood experiences are recalled with sufficient accuracy to provide useful information in retrospective studies (e.g., Kessler et al. 2005).

Regarding academic outcomes, the data contains information on GPA and an age standardized score on a common verbal test.[13] The data also provides a rich set of information on environmental and demographic variables (i.e., family income, gender, parental education, family structure, etc.) that are used as control variables in our analysis. Finally, the restricted Add Health data allows community-level variables from the Census Bureau and school input variables from the NCES common core of data to be matched to the individuals in the dataset to serve as additional controls.

Summary statistics on our sample are provided in Table 1. Household income for the full sample (column 1) is slightly higher than U.S. averages and the majority of mothers have attended college. Both the sibling and twins subsamples respectively presented in columns 2 and 3 appear gender balanced. With the sole exception of race variables, there are few differences in any of the summary statistics between the subsample of siblings and twins. While the mean verbal test score for each sample approximates the national average, the standard deviation of test scores is slightly smaller than those obtained with nationally representative samples.[14] Unlike the education and demographic variables that are similar to those obtained from nationally represented surveys, the incidence of poor mental health outcomes differ. On the one hand, roughly 8% of the sample is coded with ADHD, which exceeds the 6% national average. On the other hand, adolescents classified as being depressed in our sample is lower than the 1999 estimate of the fraction of the adolescent population being clinically depressed (12.5%) from the U.S. Department of Health and Human Services. Similarly, both obesity rates and rates of being overweight rates fall slightly below the national average for this period. Only the

Table 1: Summary Statistics

Variable	Full Sample	Sibling Sample	Twin Sample
Test Score	100.552 (13.564)	100.794 (13.324)	100.107 (13.984)
AD	0.050 (0.218)	0.049 (0.215)	0.056 (0.229)
HD	0.049 (0.215)	0.052 (0.223)	0.043 (0.203)
ADHD	0.077 (0.266)	0.077 (0.266)	0.078 (0.268)
Depression	0.062 (0.241)	0.067 (0.251)	0.052 (0.223)
Obesity	0.072 (0.258)	0.081 (0.272)	0.060 (0.238)
Age in Initial Data Collection	17.03 (1.687)	17.054 (1.700)	16.990 (1.667)
Male	0.489 (0.500)	0.479 (0.500)	0.504 (0.500)
African-American	0.169 (0.375)	0.131 (0.338)	0.234 (0.424)
Hispanic	0.141 (0.348)	0.140 (0.348)	0.145 (0.352)
Family Income (*$1,000)	46.807 (40.158)	45.206 (30.734)	49.828 (53.873)
Mother's Education	13.200 (2.203)	13.166 (2.105)	13.232 (2.356)
Parental Age	41.850 (5.337)	41.382 (5.017)	42.527 (5.750)
Observations	1684	1068	629

Note: Standard deviations in parentheses.

separate diagnoses of AD and HD fall within standard ranges observed with adolescent samples.

Table 2 documents the well-known positive association between good health and educational outcomes. Individuals classified as depressed and obese have significantly lower (one sided t-tests) verbal test scores. Surprisingly, individuals classified to have HD score higher on average than those who are not coded with this disorder.

2.1 Genetic Data

The DNA samples were drawn in the third collection and were genotyped for six candidate polymorphisms.[15] The specific markers

Table 2: Summary Statistics on Peabody Verbal Test Score Performance by Health Disorder and Health Behaviour

Variable	Full Sample	Sibling Sample	Twin Sample
Depression	92.00 (14.19)	94.03 (13.53)	91.63 (15.87)
No depression	101.03 (13.38)	101.23 (13.16)	100.70 (13.73)
T-statistic	5.705	4.44	3.66
ADHD	100.19 (12.336)	101.5 (12.167)	98.06 (12.44)
No ADHD	100.58 (13.664)	100.68 (13.40)	100.40 (14.09)
T-statistic	0.312	-0.527	1.13
HD	102.18 (11.550)	103.11 (11.77)	100.39 (11.09)
No HD	100.49 (13.657)	100.62 (13.38)	100.22 (14.10)
T-statistic	-1.112	-1.34	-0.06
AD	98.45 (12.41)	99.56 (11.92)	96.84 (13.11)
No AD	100.66 (13.62)	100.81 (13.38)	100.42 (14.01)
T-statistic	1.456	0.646	1.46
Obese	98.00 (12.755)	98.84 (13.22)	96.02 (11.50)
Not obese	100.74 (13.68)	100.91 (13.31)	100.48 (14.08)
T-statistic	2.14	1.37	1.86
Overweight	100.798 (13.44)	99.70 (14.42)	97.32 (13.76)
Not overweight	98.92 (14.22)	100.92 (13.12)	100.61 (13.97)
T-statistic	1.92	1.02	1.89
Smoke cigarettes	100.12 (12.22)	100.65 (11.93)	99.27 (12.69)
Does not smoke cigarettes	100.71 (13.97)	100.79 (13.73)	100.57 (14.38)
T-statistic	0.757	0.14	1.01

Note: Most cells present the mean verbal test score and standard deviations in parentheses for individuals by health category.

that have been collected in this study were selected based upon a large and growing body of research showing a strong correlation between

their variation and health outcomes such as obesity, ADHD, and depression, controlling for other relevant factors. It is important to state that these health outcomes are polygenic—they are affected by many mutations at many genetic loci (including many that are not collected in the study) as well as the environment an individual encounters throughout her life (as well as possible gene-environment interactions).[16] However, only an individual's genetic make-up is both assigned at conception prior to any interaction with the environment and remains invariant to all nurture investments over the life cycle, eliminating concerns related to reverse causality.

The set of genetic markers we use in our analysis includes the dopamine transporter (DAT), dopamine D4 receptor (DRD4), serotonin transporter (5HTT), monoamine oxidase A (MAOA), dopamine D2 receptor (DRD2), and cytochrome P4502A6 (CYP2A6) gene. Mutations in the coding of these genes, not the genes themselves, are believed to impact multiple health outcomes and behaviours. Scientists hypothesize that these point mutations distort cell functions and/or processes, leading to the higher propensities for specific disorders. It is important to state explicitly that individual point mutations can have phenotypic effects of any strength, including quite mild effects, and it is likely that each genetic marker has pleiotropic effects.[17]

The genetic markers collected in the Add Health study are primarily linked to the transmission of two specific neurotransmitters in the primitive limbic system of the brain: dopamine and serotonin.[18] The scientific hypothesis of how these genetic markers predispose individuals to poor health is that these genetic markers each impact the synaptic level of dopamine and serotonin, which provides larger signals of pleasure from the limbic system and leads individuals to forego other basic activities.[19] The specific markers are believed to achieve these impacts as follows: individuals with the A1 allele variants of the DRD2 gene have fewer dopamine D2 receptors than those with the A2 allele, thereby requiring larger consumption of substances to achieve the same level of pleasure. The DAT and 5HTT genes code for proteins that lead to the reuptake of dopamine and serotonin respectively. For each of these genes, longer lengths are believed to affect the speed at which production of these proteins occur. The MAOA gene product is primarily responsible for the degradation of dopamine, serotonin, and norepinephrine in several regions of the brain. A SNP of this gene is believed to have decreased

productivity of this protein, thereby increasing the risk for a number of poor outcomes. Individuals with a longer version of the DRD4 gene are more inclined to partake in additional novelty or sensation-seeking activities to achieve similar levels of reward as those with shorter variants. The CYP2A6 gene is primarily located in the liver and affects the rate of metabolism for tobacco, drugs, and other toxins. Once these compounds are broken down, they travel in the bloodstream to the brain where they generally lead to neurotransmitters being released. Finally, in our analysis we will not only consider the SNPs by themselves but also allow for gene-gene interactions, which may also have potentially powerful effects.[20] We present and discuss the genetic characteristics of our sample and unconditional relationships with poor health outcomes in the results section of the paper.

3. Empirical Framework

The empirical framework that underlies our analysis involves the estimation of a system of equations generated from a simple extension to the model developed in Ding et al. (2009). We assume that in each period, altruistic parents select inputs to maximize the household indirect utility function after receiving noisy signals of their children's health status, health behaviours, and ability endowment. Subsets of these inputs enter both an education production function and health production function, generating stocks of human capital for each child. The parents provide children who have different abilities and health outcomes with different inputs where, in equilibrium, the marginal returns to investments in schooling of one child is equated to the marginal returns to investments in health in their sibling.

First, consider a linear representation of the child's education production function, which translates a set of inputs into human capital as measured by a score on an achievement test as

$$A_{ijfT} = \beta_0 + \beta_0 X_{iT} + \beta_2 H_{iT} + \beta_3 Q_{iT} + \beta_4 N_{iT} + \nu_f + \varepsilon_{ijfT} \quad (1)$$

where A_{ijfT} is a measure of achievement for child i in family f, in school j in year t, the vector X contains individual and family characteristics (child gender, race, parental education, birth order, family income, and family structure),[21] the vector H consists of variables that capture health measures, the vector Q contains school quality variables, the vector N contains information on community and neighborhood inputs, ν_f is an unobserved family effect, and ε_{ijfT} is

an idiosyncratic error term. Notice that H_{iT} is directly included as an input to the education production function. We hypothesize that there are several possible channels under which health status potentially affects academic performance. First, it may affect the physical energy level of a child, which determines the time (including classroom attendance) that can be used for learning. Second, it affects the child's mental status that may have a direct impact on academic performance. Lastly, a child's health status may affect the way a child is treated by teachers, parents, and peers, which can, in part, shape the learning environment that is encountered.

The major empirical challenge in estimating equation (1) is that the health vector (H_{iT}) is likely to be endogenous.[22] That is, individuals with a higher health "endowment" could obtain improved academic performance because of genetic characteristics or parental investments that are also unobserved to the analyst. The inclusion of family fixed effects (ν_f) in equation (1) directly accounts for family factors that are unobserved to the researcher but that are common across siblings and may be related to both individual health and education outcomes. This allows the researcher to simultaneously control (assuming constant impacts between family members) for many parental characteristics/behaviours and some genetic factors. However, it does not provide any guidance as to why, within a twin or sibling pair, the subjects differ in explanatory characteristics such as health status. Thus, estimating equation (1) using a family fixed effects approach may overcome biases from correlations between the health vector and the family effect ν_f, but it may not completely solve the endogeneity problem, as correlations may remain the health variables and the error term (i.e., $Cov(H_{iT} - \overline{H}_f,\ \varepsilon_{ifjT} - \overline{\varepsilon}_f) \neq 0$).

Supplementing the family fixed effects strategy with instrumental variables can potentially overcome the endogeneity bias arising from $Cov(H_{iT} - \overline{H}_f,\ \varepsilon_{ifjT} - \overline{\varepsilon}_f)$. We propose to use exogenous variation from the "genetic lottery" between family members to identify the impact of poor health on measures of achievement. In the first stage equation, we explain differences in health outcomes between family members using differences in the coding of specific genetic markers between family members as an instrumental variable, while controlling for other individual and family characteristics that affect health and education outcomes. Formally, the first stage presents a linear representation of the child's health production function

$$H_{ifT} = \gamma_0 + \gamma_1 X_{iT} + \gamma_2 G_i^H + \gamma_3 Q_{iT} + \gamma_4 N_{iT} + \nu_f + \varepsilon_{ifjT}, \quad (2)$$

where G_i^H is a vector of genetic markers that may provide endowed predispositions to the current state of health status.

Our identification relies on the assumption that the vectors of genetic markers that impact health outcomes (G_i^H) are unrelated to unobserved components (ε_{ifjT}) of the achievement equation. While there might not be any existing evidence that the markers considered in this study have any impact on the education production process, it remains possible. Additionally, our strategy is valid as long as this set of genetic markers only affects A_{ifjT} via the health outcomes we consider, and not through some other channel. Using multiple genetic instruments also allows the use of over-identification tests of the validity of our choice of instruments. Finally, an additional advantage of our identification strategy is that there are no concerns regarding reverse causality, as these genetic markers are assigned at conception, prior to any health outcome or selection of any parental choice input to the health production function (even in utero).

We not only estimate the system of equations (1) and (2) via fixed effects instrumental variables methods, but also consider family fixed effects estimation of equation (1) as well as both OLS and instrumental variables estimation of the system of equations described above where $\nu_f = O$. Estimates from these alternative approaches are used to conduct specification tests that can shed light on the source of the endogeneity in estimating the impact of poor health on academic outcomes.

In the analysis, we consider two different health vectors that consist of multiple health problems. The first health vector includes depression, overweight, and ADHD. The second health vector includes depression and overweight but decomposes ADHD into being inattentive (AD) or hyperactive/impulsive (HD). We make this distinction as ADHD is often denoted by AD/HD since, as defined in the American Psychiatric Association's Diagnostic and Statistical Manual, it encompasses the "Inattentive Type," marked by distractibility and difficulty following through on tasks, as well as the "Hyperactive Type," which includes excessive talking, impulsivity, and restlessness. It is not uncommon for people to be diagnosed with the "Combined Type," showing a history of both features, but *ex ante* we would imagine that inattention and hyperactivity could have different impacts on academic performance as well as other human capital outcomes.

Finally, to examine the robustness of our results, we consider including an individual's birth weight (both linearly and up to a

quartic) as an additional control variable(s) in equations (1) and (2).[23] An individual's birth weight can be viewed as an imperfect proxy for an individual's initial stock of health capital. While birth weight is known to have a large genetic component (e.g., Lunde et al. 2007) it is well established to differ even among monozygotic twins. Royer (2009) presents evidence that birth weight differences between twins have impacts on educational attainment, and Christensen et al. (2001) demonstrate that differences in birth weight also affect health later in life between twins. Accounting for differences in birth weight can capture additional differences in both genetic factors and prenatal environments between full biological siblings.

3. Results

3.1 Genetic Associations

Our empirical identification relies on the validity of the genetic lottery to serve as a source to identify the impact of adolescent health on education outcomes. Statistically, for the genetic markers to serve as instruments, they must possess two properties. First, they must be correlated with the potentially endogenous health variables. Second, they must be unrelated to unobserved determinants of the achievement equation.

Prior to describing our instrument set and conducting formal tests, we present some summary information regarding the data that motivates the notion that these markers and their two-by-two polygenic interactions are good candidates to serve as instruments for adolescent health outcomes. Table 3 contains the conditional mean, standard deviation, and odds ratio of alternative poor health outcomes for individuals that possess a particular marker. For each genetic marker, we use at most three discrete indicators that are defined by specific allelic combinations.[24] For each poor health outcome and behaviour, there is at least one gene in which a specific SNP exhibits a higher propensity. Statistically different odds ratios in Table 3 are denoted with an asterisk. For depression, individuals with the A2A2 allele of the DRD2 gene and two seven-repeats of the DRD4 gene have significantly lower odds. For ADHD, individuals with two four-repeats of the MAOA gene have greater odds and individuals with one four-repeat of the MAOA gene have lower odds. These relationships also show up for inattention (AD) and hyperactivity (HD). For obesity, those with no repeats of the DAT1 gene have substantially lower odds.

Table 3: Relationship between Genetic Markers and Health Outcomes

Gene	Variant	ADHD	AD	HD	Obese	Depression	Smoking
DRD2	A1A1	0.076 (0.266) [0.987]	0.038 (0.192) [0.734]	0.053 (0.224) [1.103]	0.061 (0.240) [0.822]	0.053 (0.225) [0.840]	0.220 (0.416) [0.879]
	A1A2	0.071 (0.257) [0.876]	0.054 (0.225) [1.130]	0.038 (0.191) [0.671]+	0.072 (0.259) [1.014]	0.071 (0.257) [1.280]	0.237 (0.426) [0.967]
	A1A2	0.081 (0.273) [1.136]	0.049 (0.216) [0.963]	0.056 (0.229) [1.398]+	0.073 (0.260) [1.041]	0.057 (0.231) [0.827]+	0.246 (0.431) [1.071]
SLC6A4	Two short alleles	0.058 (0.234) [0.700]	0.032 (0.176) [0.576]*	0.038 (0.191) [0.726]	0.067 (0.250) [0.912]	0.076 (0.265) [1.328]	0.223 (0.417) [0.882]
	One short/one long alleles	0.084 (0.278) [1.218]	0.058 (0.234) [1.362]	0.051 (0.221) [1.111]	0.072 (0.259) [1.017]	0.054 (0.226) [0.781]	0.230 (0.421) [0.900]
	Two long alleles	0.077 (0.267) [1.016]	0.050 (0.218) [0.998]	0.052 (0.221) [1.097]	0.074 (0.262) [1.047]	0.064 (0.244) [1.049]	0.265 (0.442) [1.222]*
DAT1	No ten repeats	0.065 (0.247) [0.823]	0.032 (0.178) [0.621]	0.043 (0.204) [0.872]	0.032 (0.178) [0.416]+	0.054 (0.227) [0.856]	0.194 (0.397) [0.745]
	One ten repeat	0.088 (0.284) [1.279]	0.059 (0.236) [1.324]	0.059 (0.236) [1.381]	0.078 (0.268) [1.147]	0.062 (0.242) [1.017]	0.241 (0.428) [1.005]
	Two ten repeats	0.071 (0.257) [0.822]	0.046 (0.210) [0.832]	0.043 (0.204) [0.754]	0.072 (0.259) [1.005]	0.062 (0.241) [1.016]	0.244 (0.430) [1.057]
DRD4	No seven repeats	0.082 (0.274) [1.125]	0.052 (0.223) [1.172]	0.051 (0.219) [1.128]	0.073 (0.260) [1.039]	0.066 (0.249) [1.256]	0.242 (0.429) [1.025]
	One seven repeat	0.070 (0.255) [0.866]	0.047 (0.212) [0.919]	0.045 (0.208) [0.896]	0.068 (0.252) [0.917]	0.058 (0.235) [0.920]	0.242 (0.428) [1.006]
	Two seven repeats	0.044 (0.207) [0.546]	0.029 (0.170) [0.567]	0.044 (0.207) [0.898]	0.088 (0.286) [1.263]	0.015 (0.121) [0.219]*	0.209 (0.410) [0.827]
CYP	Main SNP	0.076 (0.265) [0.822]	0.049 (0.215) [0.604]	0.049 (0.216) [1.275]	0.073 (0.260) [1.433]	0.061 (0.239) [0.769]	0.237 (0.426) [0.687]+
	No four repeats	0.075 (0.264) [0.973]	0.046 (0.209) [0.875]	0.050 (0.217) [1.025]	0.075 (0.264) [1.074]	0.069 (0.254) [1.198]	0.235 (0.424) [0.953]
MAOA	One four repeat	0.046 (0.209) [0.507]***	0.028 (0.165) [0.477]**	0.030 (0.172) [0.546]*	0.061 (0.239) [0.795]	0.081 (0.273) [1.491]*	0.218 (0.414) [0.848]
	Two four repeats	0.093 (0.291) [1.547]**	0.064 (0.245) [1.735]**	0.057 (0.233) [1.420]+	0.075 (0.264) [1.100]	0.047 (0.212) [0.616]**	0.256 (0.437) [1.169]

*Note: Each cell presents the conditional mean, the standard deviation in round parentheses and the odds ratio for outcomes (excluding BMI) in square parentheses. ***, **, *, +, denote the Null of homogeneity of odds across markers by genotype from a chi-squared test is rejected at the 1%, 5%, 10%, and 15% level respectively. The tests were conducted with the same sample used to construct Table 1.*

The significant correlations between the SNPs and the health outcomes are also consistent with the scientific hypotheses outlined in section 2. Each of the health disorders we consider in this paper is believed to have a large genetic component and be polygenic.[25] To date, the scientific literature has not identified a unique depression, ADHD, or obesity gene. Concerns could exist that the genetic markers we use in our analysis are not only related to poor health in adolescence but also to genetic factors that directly impact education outcomes. To examine this concern, we first present evidence that there are no direct links between the inheritance of the specific genetic markers in our study with other portions of the genetic codes. Second, we present over-identification tests of our instrument sets. Last, we use a procedure developed in Conley, Hansen, and Rossi (2007) to examine the sensitivity of our estimates to the degree in which the exclusion restriction assumption is violated.

Regarding whether the inheritance of different portions of the genetic code are correlated, we examine the extent to which genetic linkages occurs in our sample.[26] We examined cross-tabulations of different genetic combinations for both the full sample as well as by the first and second family member in the data. We constructed the sample of single family members based on their relative age, since one could expect linkages within families. Whether Mendel's law of independent assortment is violated can only be tested across families. We conducted tests for homogeneity of odds ratios (results available from the authors upon request) to see whether possessing a polymorphism in one genetic marker increases the odds of possessing a specific polymorphism in a different genetic marker. We did not find any evidence indicating a systematic relationship between markers of any two of the genes for either sample that contains only one family member, lessening concerns regarding linkage.[27] This was not a surprise as linkage was highly unlikely due to the location of these markers on the genome. Additionally, using maps of the location between the specific genetic markers in our study and those which have been hypothesized to be linked to education outcomes (Plomin et al. 2006, see endnote 8 for more details), we find no evidence that they are located closely on the genome, suggesting that linkage in inheritance is unlikely. To construct the instrument set, we included only genetic markers or their interactions that had statistically significant (at the 2% level) differences in the odds ratio of suffering from one of the four conditions.[28] It is unlikely that the majority of

these unconditional relationships are due to chance, and we also considered whether the direction of the odds ratio was biologically plausible. We do not vary our instrument set across samples so that any observed difference in terms of health effects is not the result of the selection of different instrument sets that vary based on genetic similarity between family members. It is worth repeating that these genes are pleiotropic and cannot credibly account for the majority of the variation in these health disorders. Thus, even if two siblings had the same markers for many of these six genes, this would neither guarantee that they suffer from the same disorders nor that these particular genes would affect the siblings in a similar fashion.

3.2 Estimates of the Empirical Model

We now examine whether poor health is related to academic outcomes in adolescence. Table 4 presents estimates of equation (1) for the full sample. In the odd columns, results are presented for the first health vector, which includes depression, overweight, and ADHD. The even columns decompose the classification of ADHD into being inattentive (AD) or hyperactive/impulsive (HD) in the health vector. The first four columns of Table 4 present OLS and family fixed effects, which assume either that health is exogenous or that health is only correlated with the family-specific component of the residual.

We find that depression is strongly negatively correlated with academic performance. However, the estimated magnitude diminishes by over 50% when family fixed effects are included in the specification. While the impacts of depression in the OLS specifications are fairly large relative to the other health variables, they remain approximately half of the estimated magnitude of the race variables. In addition to depression, the two other mental health conditions enter the equation in a significant manner. AD is strongly negatively correlated and HD is positively correlated with academic performance when family fixed effects are not included. Despite the evidence in Table 2 that overweight and obese students score significantly lower than non-overweight and non-obese students, this state of health does not significantly affect verbal test scores in any of the specifications in Table 4, which is consistent with Kaestner and Grossman (2008). The OLS results also indicate that both African-Americans and Hispanics score substantially lower on the verbal test than Caucasian and Asian students, the children who are older in their families perform slightly better than their siblings, and parental

Table 4: Estimates of the Achievement Equation for the Full Sample

Estimation Approach	OLS		Family Fixed Effects		Instrumental Variables		Family Fixed Effect Instrumental Variables	
AD	N/A	-3.447 (1.307)**	N/A	-2.202 (1.483)	N/A	-18.351 (11.354)	N/A	-26.026 (13.011)*
HD	N/A	2.305 (1.306)+	N/A	1.810 (1.542)	N/A	24.807 (15.031)+	N/A	2.553 (12.896)
ADHD	-1.263 (0.987)	N/A	-0.250 (1.167)	N/A	-7.845 (11.104)	N/A	-6.924 (15.811)	N/A
Depression	-4.318 (1.333)**	-4.282 (1.333)**	-2.083 (1.249)+	-2.079 (1.247)+	-10.046 (17.953)	-12.282 (14.992)	-10.854 (15.186)	-3.627 (13.882)
Obesity	-0.468 (0.750)	-0.460 (0.747)	-0.007 (0.893)	0.051 (0.893)	3.335 (7.661)	3.179 (7.333)	-5.210 (9.875)	4.630 (8.072)
Age	5.483 (3.263)+	5.439 (3.259)+	1.191 (3.658)	0.886 (3.657)	4.659 (3.829)	3.836 (3.970)	1.015 (6.065)	1.431 (5.580)
Age squared	-0.165 (0.096)+	-0.163 (0.096)+	-0.029 (0.107)	-0.019 (0.107)	-0.141 (0.115)	-0.109 (0.118)	-0.023 (0.175)	-0.018 (0.164)
Male	1.240 (0.595)*	1.204 (0.594)*	-0.609 (0.691)	-0.618 (0.689)	1.668 (1.076)	0.730 (0.837)	-0.155 (1.157)	0.003 (1.037)
African-American	-9.245 (0.852)**	-9.270 (0.850)**			-9.461 (1.130)**	-9.354 (1.083)**		
Hispanic	-7.185 (0.944)**	-7.156 (0.942)**			-7.755 (1.668)**	-6.887 (1.571)**		
Sibling	0.482 (0.623)	0.436 (0.623)			0.237 (0.934)	0.097 (0.972)		
Birth order	-1.236 (0.311)**	-1.249 (0.311)**	-1.647 (0.780)*	-1.616 (0.779)*	-1.240 (0.398)**	-1.335 (0.406)**	-1.813 (1.187)	-0.818 (1.143)
Family Income	0.021 (0.006)**	0.020 (0.006)**			0.021 (0.008)**	0.020 (0.008)*		
Maternal Years of Education	1.139 (0.153)**	1.134 (0.153)**			1.301 (0.371)**	1.068 (0.344)**		
Parents Age	0.266 (0.062)**	0.262 (0.062)**			0.249 (0.080)**	0.229 (0.083)**		
Parents Married	0.082 (0.733)	0.110 (0.733)			-0.007 (0.953)	0.250 (1.034)		
Observations	1684	1684	1684	1684	1684	1684	1684	1684

*Note: Corrected standard errors in parentheses. ***, **, * denote statistical significance at 1%, 5%, 10% level respectively..*

education and family income are positively correlated with test scores. There does not appear to be any evidence indicating that gender differences exist once family fixed effects are controlled.

Instrumental variable and family fixed effects IV estimates of the impacts of poor health on education are presented in the last four columns of Table 4. The IV estimated impacts of depression, AD, and HD are very large relative to the OLS results, and the latter two are marginally significant. As to the size of the impact, the results indicate that both depression and inattention lead to substantial decreases in test scores whereas HD leads to a marked increase. The inclusion of family fixed effects leads the IV point estimate of HD and depression to become statistically insignificant in both health vectors. Notice in the last column that the magnitude of the coefficient on depression and HD diminishes substantially as we add the family fixed effects into the IV analysis. Only the IV fixed effects estimate of AD remains statistically significant once we account for family fixed effects. It also increases by over 40% in magnitude. Focusing on the fixed effects IV specification in column 8 as a benchmark, the point estimate indicates that suffering from inattention would lead to roughly a 26-point decline in academic performance. We note that the parameters in Table 4 are reduced-form estimates. Since we have instrumented for poor health outcomes, we make the causal assertion that AD significantly decreases verbal tests scores, while a range of other demographic variables excluding race, birth order, and maternal education have at best a tenuous impact on test score performance.[29]

Attenuation bias due to measurement error in the AD and HD variables could account for some of the difference between the OLS and instrumental variable estimates in Table 4. Recall that these classifications are based on answers to retrospective questions, which are thought to be recorded with error. By including statistical controls for common family influences, the fixed effects strategy only uses information within families, attenuating the variance in the regressors. Thus, measurement error imposes a degradation in the signal to noise ratio and a variable measured with error will be severely biased toward zero. Interestingly, only the estimates on two health conditions, HD and depression, become smaller when family fixed effects are accounted for when estimating equation (1), suggesting this is not the explanation for the large difference in the impact of AD.

The estimates from Table 4 can also be used to examine the source of the endogeneity in the health variables. Tests of joint significance of the family effects are statistically significant for all specifications. This indicates that one should account for family-specific heterogeneity. Random effect estimates (not reported) were used to conduct Hausman tests of the endogeneity of the health variables, and the results suggest fixed effects indeed removes some of the endogeneity. We next examined whether accounting for family fixed effects eliminates the need to treat the health vector as endogenous by testing the Null hypothesis that the IV estimates and the fixed effects IV estimates are similar using a Hausman-Wu test. If the Null is accepted, this would suggest there are efficiency gains from conducting family fixed effects estimates. For both health vectors, we can reject the Null of IV and IV/FE coefficient equality, suggesting that the family fixed effects do not fully remove the sources of endogeneity that bias estimates of the impacts of poor health.

Similarly, we conducted Hausman tests between the simple OLS and IV estimates. In the event of weak instruments (as well as overfitting), the fixed effects IV estimates would be biased towards the OLS estimates. We can reject the Null of exogeneity of health outcomes for each health vector with each sample at the 5% level.

3.2.1 Testing the Validity of the Instruments

We considered several specification tests that examine the statistical performance of the instruments for each health equation and sample. Since our IV estimates are over-identified, we use a J-test to formally test the over-identifying restrictions. This test is the principal method to test whether a subset of instruments satisfy the orthogonality conditions. The smallest of the p-values for these tests is 0.29, providing little evidence against the over-identifying restrictions.[30]

In order to further examine whether these genetic markers are valid instruments, we considered several specification tests to be used with multiple endogenous regressors. First, we used the Cragg-Donald (1993) statistic to examine whether the set of instruments is parsimonious (i.e., the matrix is of full rank) and has explanatory power. Second, in order to examine whether weak instruments are a concern, we calculated the test statistic proposed by Stock and Yogo (2005).[31] To demonstrate the strength of the instruments, we considered the most difficult test with our data is using the full set of genetic instruments. That is, since using a large number of

instruments or moment conditions can cause the estimator to have poor finite sample performance, we will demonstrate results using the full set of genetic instruments and their polygenic interactions. Our preferred instrument sets are a subset, and one could argue that we achieved strong results in those contexts since we dropped redundant instruments, thereby leading to more reliable estimates.[32]

The critical value for the Stock and Yogo (2005) test is determined by the number of instruments, endogenous regressors, and the amount of bias (or size distortion) one is willing to tolerate with an IV estimator. With the full set of instruments, the critical value increases substantially and we find that the Cragg-Donald statistic is 45.73 and 46.11 in health vector 2 with and without family fixed effects respectively, which exceeds the critical value.[33] This suggests that even with this large set of instruments, the estimator will not perform poorly in finite samples and that, with or without family fixed effects, we can reject the Null hypothesis, suggesting an absence of a weak instruments problem. We also considered more traditional F-statistics with our preferred set to test for the joint significance of the full set of instruments in each first stage equation.[34] The first stage F-statistics indicate that in each equation the full set of instruments is jointly significant in both the specifications that include and exclude family fixed effects. We also examined the partial R-squared for each outcome and they ranged between 2.3% and 5.1%, which fit our prior belief that, since these disorders are polygenic, it would be unlikely that these genes would account for more than 5% of the variation in the disorders.

To examine the sensitivity of both our IV and family fixed effect IV estimates to the degree in which the exclusion restriction assumption is potentially violated, we considered the local to zero approximation sensitivity analysis proposed in Conley, Hansen, and Rossi (2007). This analysis involves making an adjustment to the asymptotic variance matrix, thereby directly affecting the standard errors. While the variance matrix continues to account for the usual sampling behaviour, Conley, Hansen, and Rossi (2007) suggest including a term that measures the extent to which the exogeneity assumption is erroneous.[35] The amount of uncertainty about the exogeneity assumption is constructed from prior information regarding plausible values of the impact of genetic factors on academic performance that are obtained from the reduced form. We successively increased by 5% increments the amount of exogeneity error from 0% to 90% of the

reduced form impacts. At levels below 40% of the reduced form impacts, our results are robust as inattention continues to have a statistically significant negative impact on verbal test scores. Our full set of results become statistically insignificant only if the extent of deviations from the exact exclusion restrictions are assumed to be above 60% of the reduced form impacts. Since there does not exist any scientific evidence that these specific markers directly affect academic achievement, the sensitivity analysis indicates that the levels at which our results are sensitive to the exclusion restriction assumption appear highly implausible. The sensitivity analysis suggests that our quantitative results are robust to potentially mild and moderate violations of the exogeneity assumption, further increasing our confidence in Table 4.

3.3 Robustness

In order to demonstrate the robustness of our empirical findings, we replicated the analysis on various subsets of the data based on family relationships, zygosity, and gender as well as additional controls for health endowments. We considered these family relationship breakdowns as the inclusion of family fixed effects would ensure that only the dizygotic twins and siblings identify the fixed effect IV estimates of $\beta 2$. The measure of genetic relatedness does not differ in theory between dizygotic twins and full siblings since dizygotic twins, coming from different eggs, are as genetically similar as any other non-twin sibling and have a genetic correlation of approximately half that of monozygotic twins. However, the inclusion of family fixed effects also imposes an equal environment assumption on the family members. That is: (1) family inputs that are unobserved to the analyst do not differ between family members, and (2) these factors have the same impact on achievement between relations. This assumption of equal impacts from family factors is more likely to be satisfied with data on twins than siblings as one could imagine that (1) parents make differential time-varying investments across siblings, and (2) the impacts of particular family factors may differ for children of different ages. In addition, sibling models do not effectively deal with endogeneity bias that could result from parents adjusting their fertility patterns in response to the (genetic) quality of their earlier children.[36]

While one could imagine that data on the subsample of twins would provide the strongest robustness check, we imposed an additional sample restriction that the pairs (or trios) of children are of the same

gender. It is more likely that parents will make the same investments in the children who are most similar.[37] We replicate the above analysis only on the subsample of twins of the same gender, and the results from all four estimation approaches are presented in Table 5.

Table 5: Estimates of the Achievement Equation for the Sample of Twins of the Same Gender

Estimation Approach	OLS		Family Fixed Effects		Instrumental Variables		Family Fixed Effect Instrumental Variables	
AD	N/A	-5.957 (2.297)**	N/A	-3.049 (2.552)	N/A	-4.292 (6.218)	N/A	-14.991 (7.475)*
HD	N/A	2.061 (2.592)	N/A	-0.172 (2.749)	N/A	-4.213 (8.633)	N/A	-15.994 (10.828)
ADHD	-4.538 (1.812)*	N/A	-2.155 (2.153)	N/A	-6.643 (14.245)	N/A	-18.075 (6.473)**	N/A
Depression	-3.184 (2.969)	-3.306 (2.928)	0.738 (2.493)	0.734 (2.498)	-7.181 (17.247)	-4.161 (15.283)	-12.229 (21.557)	-11.27 (17.456)
Obesity	-2.853 (1.427)*	-2.93 (1.421)*	0.007 (1.81)	0.059 (1.81)	-3.379 (9.682)	-3.25 (8.718)	-3.884 (6.880)	-1.61 (6.261)
Male	3.597 (1.127)**	3.483 (1.125)**			3.641 (1.670)*	3.619 (1.515)*		
African-American	-8.318 (1.463)**	-8.311 (1.463)**			-8.464 (2.009)**	-8.345 (1.970)**		
Hispanic	-6.894 (1.757)**	-6.93 (1.735)**			-6.895 (2.733)*	-6.974 (2.643)**		
Family Income	0.012 (0.004)**	0.013 (0.004)**			0.012 (0.007)	0.012 (0.007)+		
Maternal Years of Education	1.275 (0.240)**	1.249 (0.240)**			1.233 (0.363)**	1.26 (0.346)**		
Parents Age	0.184 (0.099)+	0.184 (0.099)+			0.197 (0.134)	0.187 (0.134)		
Parents Married	-1.659 (1.263)	-1.657 (1.268)			-1.795 (1.652)	-1.776 (1.680)		
Observations	469	469	469	469	469	469	469	469

*Note: Corrected standard errors in parentheses. ***, **, * denote statistical significance at 1%, 5%, 10% level respectively.*

Notice the OLS estimates (column 2) suggest a substantially larger role for ADHD (column 1) and AD (column 2), whose magnitude is

nearly twice as large as that for the full sample presented in Table 4. On average, inattention leads to a six-point decline in verbal test scores. Depression no longer enters the equation in a significant manner, though the magnitude is similar, and the impact of being overweight on academic performance leads to a small decrease in academic performance that is statistically significant at the 10% level. None of the health variables enter the equation in a significant manner once we either include family fixed effects or use traditional IV analysis. However, once we account for family fixed effects and also instrument the health conditions, AD continues to enter the equation in a significant manner. On average, a child with AD scores almost 14 points lower. ADHD also now enters significantly in these specifications and HD now enters in a marginally significant manner, but the sign of the coefficient has changed. The large impact of both AD and HD are identified from dizygotic twin pairs, which differ in these classifications, but this is the only specification in which the impacts of AD and HD enter in a significant manner and are not significantly different. While neither depression nor obesity enter the equation in a statistically significant manner, it is important to stress that we have a very small sample size in which we are able to identify effects and approximately 60% of the twin pairs are monozygotic, leading to larger standard errors. However, the coefficient estimates for depression and overweight are practically identical in magnitude and sign to those presented in Table 4. Additionally, tests of the validity of the instrument continue to suggest that this set of genetic markers has good statistical properties and Hausman tests between columns 2 and 6 of Table 5 reject the exogeneity of the health vector.

We believe that the estimates in Table 5 present the strongest possible robustness check for our empirical evidence of causal impacts of poor mental health on academic achievement as the family members are of the same age, race, and gender. With the exception of health and education outcomes, the only other measures contained in our data for which there are different values within children in these families are genetic markers. As noted above, these results are also robust to including birth weight controls. The fixed effect IV estimates presented in the last column continue to suggest that poor mental health impacts academic performance, whereas our physical health measure has no significant impact.

Since one must always be cautious in attributing external validity to an analysis with twins' data, we replicate the analysis that corresponds

to Table 4 where we utilize only the subsamples of siblings. As discussed above, the equal family environment assumption is inconsistent with many models of family behaviour[38] and the likelihood that the assumption is valid is higher with the subsample of twins (of the same gender) versus siblings.[39] However, results with the siblings sample are likely of increased external validity, so there is a clear trade-off. In the sibling sample (results available from the authors upon request), the AD condition continues to lead to a significant decrease in test scores and none of the other health variables enter the equation in a significant manner in the family fixed effects and IV analyses. Lastly, in this subsample, the instrument set continues to have good first stage properties, the p-values of the over-identification tests are above 0.35, Hausman tests suggest that the health vector should be treated as endogenous, and family fixed effects by themselves do not remove all of the potential biases.

As a final robustness check of our main results, we consider including an individual's birth weight (both linearly and up to a quartic) as an additional control variable(s) in equation (1). By directly accounting for differences in birth weight we could potentially control for additional differences in both genetic factors and prenatal environments between full biological siblings. We find that our full set of results (available upon request) from Tables 4 through 7 is robust to both of these specifications. In particular, inattention continues to negatively impact academic performance and specification tests reject family fixed effects estimators in favour of family fixed effect IV estimators.

3.4 Comorbidity and Measurement Error

In our study, we used a rich vector of health outcomes in part to ensure that the exclusion restriction property of the instrument holds. Using only a single health outcome to proxy for health could lead to different results, since health disorders and risky health behaviours are known in the medical literature to be more common among individuals with one particular disorder than among the remaining population. Table 6 demonstrates the substantial presence of comorbidities in our sample. Column 1 of Table 6 displays the number of individuals (and marginal distribution) in each wave that smoke or have been classified with either AD, HD, ADHD, obesity, or depression. Across each row, we present the number of individuals (and conditional frequency) who also engage in smoking or suffer

Table 6: Relationship between Health Behaviours and Health Outcomes during Adolescence

Behaviour	Total Number	Nothing Else	Also ADHD	Also AD	Also HD	Also Obese	Also Depressed	Also Smokes
Full Sample								
Nothing	975 [58.24]	***	***	***	***	***	***	***
ADHD	129 [7.66]	67 (51.94)	------	------	------	16 (13.22)	11 (8.53)	46 (35.66)
AD	84 [4.99]	40 (47.62)	------	------	37 (44.05)	11 (13.10)	8 (9.52)	33 (39.29)
HD	82 [4.87]	41 (50.00)	------	37 (45.12)	------	11 (13.41)	5 (6.10)	30 (36.59)
Obese	121 [7.19]	69 (57.50)	16 (12.40)	11 (9.09)	11 (9.09)	------	14 (11.57)	32 (26.67)
Depression	104 [6.18]	48 (46.15)	11 (11.93)	8 (7.69)	5 (4.81)	14 (13.46)	------	44 (42.31)
Smokes Cigarettes	404 [24.08]	297 (73.51)	46 (11.39)	33 (8.17)	30 (7.43)	32 (7.92)	44 (10.89)	------

Note: Each cell contains the number of individuals diagnosed with the respective row and column combination. The conditional frequency of dual diagnoses is presented in round parentheses. The marginal probability of being diagnosed with each outcome is presented in square [] parentheses.

other poor health outcomes. Not only are adolescents with ADHD more likely to smoke, but they also have a higher rate of being classified as either depressed or obese than their cohorts (one sided t-tests). This result is not unique to ADHD, as we find that individuals with any of these health disorders are significantly more likely to have a second disorder. In addition, those with any health disorder are more likely to smoke cigarettes.

The majority of the empirical literature that estimates the impact or association of health with socio-economic outcomes generally includes only a single explanatory measure, such as obesity, smoking, or birth weight in their analysis. We considered what would happen to the sign, significance, and magnitude of the estimated impact of each specific disorder if we followed the usual practice and did not control for comorbidities in the achievement equation. It is reasonable to hypothesize that in OLS and family fixed effects strategies, omitted variable bias would arise, since many of the neglected health

conditions would be correlated with both the included health condition as well as verbal test scores. Further, in these specifications, IV or family fixed effects IV estimates may not overcome these biases, unless a subset of the genetic instruments are known to be scientifically unique to that included health condition to ensure the plausibility of the exclusion restriction assumption. Excluding significant comorbid conditions potentially leads to problems not only with sets of genetic markers as instruments, but makes it equally difficult to imagine that any nurture or environmental factor could break the statistical association between those included and excluded to the estimating equation measures of poor health.[40] In our application, there may be a concern that the genetic markers used in the above analysis may also be associated with health measures not available in the data. An exhaustive survey of PubMed indicates two potential disorders: schizophrenia and Tourette's syndrome. However, each of these disorders has low prevalence rates and low discordance rates within families. Thus, we do not believe that this is a major issue with either the IV or fixed effects IV specification reported earlier, but it remains an empirical question.

Table 7 presents OLS, family fixed effect, IV, and fixed effects IV estimation of equation (1) where the health vector includes only a single specific disorder at a time.[41] Thus, each entry in Table 7 refers to the point estimate of that specific health outcome on verbal achievement, controlling for the same set of observed controls as in Table 4. The empirical estimates of several disorders differ from that obtained using the full health vector reported in Table 4. In the OLS regressions reported in Table 7, HD no longer enters significantly and the magnitude of the impact of AD is substantially smaller. The fixed effects results in Table 7 are very similar to those obtained in Table 4, which could suggest that there are limited sets of twins/siblings that are discordant for multiple health problems. Interestingly, the impact of depression does not vary substantially between Table 7 and Table 4 in the OLS and fixed effects analysis.

The IV estimates in Table 7 differ greatly and it could be concluded that each health variable (with the exception of AD) has a significant impact on academic performance. Depression is negatively and significantly related to verbal test scores, but the estimated impact of hyperactivity changes signs from that reported in Table 4. ADHD is highly negatively related to test scores and enters in a significant manner at the 15% level. The estimated impact of being overweight

Table 7: Estimates of the Achievement Equation Where We Include Only a Single Health Condition by Itself

Estimation Approach	OLS	Family Fixed Effects	Instrumental Variables	Family Fixed Effects and Instrumental Variables
AD	-2.275 (1.176)+	-0.737 (1.352)	-0.904 (6.040)	-15.050 (9.790)
HD	1.106 (1.142)	1.356 (1.408)	13.510 (9.600)	-7.353 (8.846)
ADHD	-1.208 (0.981)	0.317 (1.142)	3.304 (7.077)	-12.303 (8.532)
Depression	-4.473 (1.285)**	-2.193 (1.209)+	-23.265 (11.010)*	-5.742 (8.625)
Obesity	-0.846 (0.741)	-0.06 (0.877)	7.879 (5.308)	-6.887 (4.328)
Estimates from Specifications Which Only Include AD and HD Separate Diagnoses.				
AD	-3.289 (1.289)*	-1.424 (1.457)	-19.900 (12.456)	-17.164 (11.401)
HD	2.495 (1.302)+	1.912 (1.519)	31.573 (14.986)*	7.415 (12.557)

*Note: Corrected standard errors in parentheses. Each cell of the table corresponds to a separate regression. The dependent variable of the regression differs by row. Columns reflect different estimation approaches as denoted in the first row. Regressions control for the same set of non-health inputs as in Table 5, including student demographics, parental characteristics and home environment variables. ***, **, * denote statistical significance at 1%, 5%, 10% level respectively.*

now becomes significant at the 15% level and leads to a seven-point increase in test scores on average when estimating equation (1) using IV analysis. Regarding the preferred fixed effects IV specifications from Table 7, we would conclude that AD and ADHD each has a negative and significant impact on academic performance. The sign of the estimated impact on HD changes from negative to positive. Interestingly, the addition of family fixed effects leads the estimated signs of the impacts of ADHD, HD, and obesity to change signs when instruments are also employed. Similar to Table 4, the estimated impact of depression decreases substantially when family fixed effects and instrumental variables are used to estimate equation (1). Finally, sensitivity analysis for all IV and family fixed effects IV estimates in Table 7 indicate that they are extremely sensitive to the degree in

which the exclusion restriction assumption is potentially violated. None of the results remain significant at very low levels of exogeneity error (5–10% of the reduced form impacts), confirming that ignoring comorbid conditions leads to the exclusion restriction assumption becoming implausible.

Overall, this investigation clearly demonstrates that controlling for comorbid conditions is an important issue to credibly estimate the impact of specific health conditions on educational outcomes. We find that there are numerous differences in the estimated impacts of mental health disorders when estimating equation (1) by OLS, IV, and family fixed effects with IV, depending on whether one comorbid conditions are accounted for in the specifications. To summarize, constructing an appropriate health vector presents an additional challenge for empirical researchers, as the omission of comorbid conditions could lead to either biases in coefficient estimates or invalidate exclusion restriction assumptions.

4. Conclusions

Numerous studies have reported that within families, siblings and twins are often radically different in personality traits, health, education, and labour market outcomes. Researchers have traditionally examined whether different environmental factors account for the development of these differences within families but have concluded that these factors can only account for a limited amount of the variation in outcomes within families. Each time a new sibling is conceived, a genetic lottery occurs and roughly half of the genes from each parent are passed on to the child in a random process. With recent scientific discoveries (most notably the decoding of the human genome), it is now possible to collect data that provides a precise measure of specific genetic markers, permitting researchers to directly explore a variable that empirical researchers traditionally viewed as unobserved heterogeneity. In this paper, we exploit variation within siblings and twins from the genetic lottery to identify the causal effect of several poor health conditions on academic outcomes via a family fixed effect/instrumental variables strategy.

We find evidence of large impacts from poor mental health to lower academic performance. Inattention leads on average to a one standard deviation decrease in performance on a verbal tests within families. Our results indicate that, while researchers should treat health as an endogenous input when estimating education production functions,

family fixed effects estimators by themselves cannot fully remove the endogeneity bias. We present evidence that differences in genetic inheritance have desirable properties to identify the impact of poor health on education within families as there are, consistent with the biomedical literature, statistically significant correlations with each endogenous health variable, and sensitivity analyses indicate that our results are robust to reasonable violations of the exclusion restriction assumption. Lastly, our results underscore the challenge facing empirical researchers interested in identifying the impact of specific health conditions that arises due to comorbidities.

The quantitative and qualitative patterns of our empirical results are robust to not only multiple sample definitions, including the restriction to using only dizygotic twins of the same gender, but also the inclusion of an individual's birth weight. A potential limitation of this study deals with external validity. It is important to consider whether our analysis of family members can be generalized to larger populations of interest.

We believe that there is substantial potential from using data on genetic markers in social science research. As the scientific literature is developing an ever-increasing understanding of how genetic inheritance relates to individual (health) outcomes, this knowledge can be used to refine searches for potential genetic markers to serve as instrumental variables. Genetic markers have a great deal of conceptual validity as instruments for many (health) outcomes since (1) the markers are inherited at conception prior to any interaction with the environment, eliminating concerns related to reverse causality; (2) a large body of literature exists that documents robust correlations between specific markers and individual (health) outcomes; (3) studies of genetic inheritance and measures of genetic distance from maps of the human genome are available to investigate whether genetic linkage is a valid concern; and (4) most genes are pleiotropic so that a predisposition can be viewed as a form of inherited encouragement. In addition, researchers could investigate the sources of pleiotropy by examining how different environmental disturbances affect gene expression and how that relates to a variety of economic outcomes. In summary, we believe that integrating biological findings into the social sciences has the potential to not only address open research questions but also help develop policies that can promote human capital development. However, unlike biological measures such as height, weight, blood pressure, blood

alcohol content, cholesterol levels, or hormones whose measures are influenced by behavioural inputs, genetic markers are time-invariant and cannot be modified by environmental influences. However, within families, any differences in the inheritance of specific markers present the opportunity for additional experiments in "nature."

Acknowledgements

We are grateful to Ken Chay, Dalton Conley, Weili Ding, Ted Joyce, Robert McMillan, John Mullahy, Matthew Neidell, Jody Sindelar, and participants at the 2007 NBER Summer Institute, Northwestern University, Brown University, CUNY, McGill University, University of Calgary, Tinbergen Institute, Institute for Fiscal Studies, Warwick University, University of Calgary, 2008 AHEC Conference at the University of Chicago, 2008 SOLE meetings, Yale Health Policy Colloquium, University of British Columbia, University of Connecticut, University of Saskatchewan, University of Tennessee, University of Toronto, and Simon Fraser University for comments and suggestions that have improved this paper. We are both grateful to the CLSRN for research support. Lehrer also wishes to thank SSHRC and CPHS New Faculty Fellowship for additional research support. We are responsible for all errors. This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524, USA (addhealth@unc.edu).

Notes

1. This debate has been traced back to thirteenth-century France and the field of quantitative behavioural genetics basically compares trait similarities across individuals that systemically differ in the genetic or environmental influences they have in common (e.g., identical vs. fraternal twins, adoptive vs. biological children) to decompose the variation of quantitative traits, and their covariances with other traits, into genetic and environmental (co)variance components. Within economics, Cesarini et al. (2008, 2009) utilize these methodologies to demonstrate that preferences for co-operative behaviour, risk, and giving have a significant genetic component. The relative importance of nature and nurture is of particular relevance for public policy. For example, consider education policy. If nurture factors drive the success of children in school,

inequality in educational opportunity may well come from sources such as failing capital markets, suggesting that specific policies could reduce future inequalities in schooling. However, if inequality in educational opportunity reflects the distribution of innate ability among the population, there are fewer opportunities to design policies that can reduce future inequality. That being said, the notion that nurture inputs are more easily susceptible to policy remediation relative to nature is a non sequitur.

2. For example, see Johnson (2003), Kelada et al. (2003), Goldstein et al. (2003), Zerhouni (2003), and Merikangas and Risch (2003).

3. Using similar methodologies, economists have begun to explore whether specific genetic loci are associated with financial risk preferences (e.g., Dreber et al. 2009; Benjamin et al. 2009).

4. Grossman and Kaestner (1997) and Strauss and Thomas (1998) present surveys of the literature of the impact of health on, respectively, education and income. The majority of empirical studies discussed in the surveys report correlational relationships.

5. Several other studies that use alternative empirical approaches are worth noting. Kremer and Miguel (2004) randomly assign health treatments to primary schools in Kenya and find that health improvements from the clinical treatment significantly reduced school absenteeism but did not yield any gains in academic performance. Bleakley (2007) uses a quasi-experimental strategy that exploits different timing at which cohorts were exposed to a large-scale public health intervention against hookworm in childhood. He finds that the treatment boosted health and was associated with larger gains in income and higher rates of return to schooling later in life.

6. Genes consist of two alleles, and a child randomly inherits one of the two alleles from each parent at the time of conception. The child's genome consists of approximately 3.2 billion base pairs, along which there are 9.2 million candidate SNPs (International HapMap Consortium 2005), which are specific locations where a mutation in the genetic code is known to occur in the population. This variability in the genetic code may influence an individual's susceptibility to various developmental outcomes, such as developing an illness. In other words, our empirical strategy exploits these differences in the coding of a specific marker between full siblings and can intuitively be viewed as an experiment in "nature."

7. Ding et al. (2006, 2009) was the first empirical study within economics to explicitly use differences in genetic information across individuals as an instrumental variable in estimating the effects of poor health on high school grade point average (GPA). More recently, Norton and Han (2008) use genetic information to attempt to estimate the impact of obesity on employment. Neither study exploited variation in genetic inheritance within families (the "genetic lottery"), which we show to be important empirically and improves the plausibility of the exclusion restriction.

8. Plomin et al. (2006) and de Quervain and Papassotriopoulos (2006) present recent surveys on which genes are believed to be directly associated with intelligence and memory ability respectively. Using maps of the location between these genes and the specific genetic markers in our study, we find no evidence that they are located closely on the genome, suggesting that linkage in inheritance is unlikely. Researchers have found no direct links between several of the genes in this study and intelligence (i.e., Moises et al. 2001) or cognitive ability (e.g., Petrill et al. 1997), and we hypothesize that, if a link exists, it operates through specific health measures.

9. These ideas are not new, having been discussed in Harrison (1970) and Allen (1970).

10. Similarly large negative impacts of poor health on measures of later cognitive achievement have been found in studies that exploit shocks to an individual's prenatal conditions such as in utero exposure to the flu (Almond 2006) and low levels of radiation (Almond, Edlund, and Palme 2008).

11. The importance of the sensitivity analysis should not be understated, since poor health conditions often occur simultaneously and it is hard to identify a unique source of genetic or environmental variation to identify the impact of specific disorders due to the potential presence of unmeasured comorbid conditions. As we discuss in the results section, in our context, the main threats are schizophrenia and Tourette's syndrome, health measures which were not collected in the dataset. We argue that this concern is unlikely to be a serious threat to our main results as schizophrenia does not manifest itself among adolescents and Tourette's syndrome is extremely uncommon, with current estimates indicating that it affects approximately 0.5 to 3 people in 1,000.

12. Add Health selected schools in 80 communities that were stratified by region, urbanicity, school type (public, private, or parochial), ethnic mix, and size. In each community, a high school was initially selected, but, since not all high schools span grades 7–12, a feeder school (typically a middle school) was subsequently identified and recruited. In total, there are 132 schools in the sample. Additional details on the construction of the sample are provided in Harris et al. (2003).

13. For example, Babinski et al. (1999), Ding et al. (2009), and Fletcher and Wolfe (2008a) present empirical evidence of different impacts from these two diagnoses.

14. The test is an abridged version of the Peabody Picture Vocabulary Test-Revised and consists of 78 items. The test was administered at the beginning of the in-home interview and first involves the interviewer reading a word aloud. The respondent then selects the illustration that is the closest match to the word from four simple black-and-white illustrations. The test is arranged in a multiple-choice format. See http://www.agsnet.com/assessments/technical/ppvt.asp for details.

15. Complete details of the sampling and laboratory procedures for DNA extraction, genetic typing, and analysis are provided in an online document prepared by Add Health Biomarker Team, available at http://www.cpc.unc.edu/addhealth/files/biomark.pdf/. Note that the method to genotype varies across markers and different assays were conducted. In addition, to reduce coding errors, genotypes were scored independently by two individuals. To control for potential genotyping errors, any analysis that is questionable for routine problems (i.e., poor amplification, gel quality, software problems, etc.) is repeated.

16. More recently, evidence indicates that differences within families, even among identical twins, can exist because of epigenetic factors. Epigenetics refer to natural chemical modifications that occur in a person's genome shortly after conception and that act on a gene like a gas pedal or a brake, marking it for higher or lower activity. For instance, identical twins have different fingerprints. The general pattern of their fingerprints is determined by genetic factors and is initially identical; however, the exact pattern changes in utero based on when and how each twin touches the amniotic sac (Jain et al. 2002).

17. Pleiotropy refers to the heterogeneous impacts that a difference in specific genetic marker occurs. Intuitively, the operation is similar to a "power grid," as a single-gene mutation may also affect the expression of other genes, which together leads to changes in behaviours and outcomes.

18. The effect of a neurotransmitter comes about by its binding with receptor proteins on the membrane of the postsynaptic neuron. As long as the neurotransmitter remains in the synapse, it continues to bind its receptors and stimulate the postsynaptic neuron. In the brain, dopamine and serotonin function as a neurotransmitter as they are commonly believed to provide individuals with feelings of enjoyment. Caplin and Dean (2008) and Caplin et al. (2009) have recently developed formal neuroeconomic models that are consistent with specific neuroscientific hypotheses that respectively explain how dopamine affects individual decision making and belief formation.

19. The limbic system is highly interconnected with the region of the brain associated with reward and pleasure. This region was initially discovered in Olds and Milner (1954), who reported that, if given the choice of food versus stimulation by electrodes of the neurons within this region of the brain, rodents ended up dying from starvation and exhaustion rather than lessening the stimulation of their pleasure centre. Recent studies using mice whose genes have been mutated to affect dopamine and serotonin production have confirmed that these markers affect basic activities.

20. For example, Dremencov et al. (2004) present evidence that the SNPs of the 5HTT gene interacts with genes that release dopamine and suggest this channel could impact the speed at which certain pharmaceutical treatments become effective. Similarly, since many addictors stimulate dopamine release in the nucleus accumbens, it is likely that the rate of metabolism of these drugs (which is in part determined by the CYP2A6 gene) interacts with the DRD2 genes.

21. *Ex ante*, one could hypothesize that parental education and family income are positively associated with measures of academic performance. In genetic studies, controlling for ethnicity and race are important as it has been hypothesized that there are differences in allele frequencies across race and ethnic groups (e.g., Cooper et al. 2003). Within families, birth order effects could exist as higher rank children are more likely to have older parents at birth, which could affect the amount of time invested by parents. Similarly, across families, higher rank children are more likely to be born into larger families, which can also capture family size effects.

22. An equally important challenge occurs in measuring the health vector from omitted variables. If the researcher omits comorbid conditions, biased estimates of the impacts of poor health on academic outcomes will be recovered. This empirical challenge is discussed in detail in section 3.4 of the text.

23. It is well documented by many authors that better health early in life is associated with higher educational attainment (e.g., Grossman 1975; Perri 1984) and that more educated individuals in turn have better health later in life (e.g., Grossman and Kaestner 1997; Cutler and Lleras-Muney 2007).

24. The DAT genotypes are classified with indicator variables for the number of ten-repeat alleles (zero, one, or two). The MAOA genotype is classified with indicator variables for the number of four-repeat alleles (zero, one, or two). Similarly, the DRD4 genotype is classified with indicator variables for the number of seven-repeat alleles (zero, one, or two). The DRD2 gene is classified as A1/A1, A1/A2, or A2/A2 where the A1 allele is believed to code for reduced density of D2 receptors. The SLC6A4 gene is classified as SS, SL, or LL where S denotes short and L denotes long. Finally, we include indicator variables for the two possible variants of the CYP gene. We organize the genetic data reported in the empirical table in order of the raw number of individuals who possess each particular marker within that gene from lowest frequency to most common.

25. Polygenic refers to a phenotype that is determined by multiple genes. For example, the ninth annual Human Obesity Gene Map released in 2006 identified more than 300 genes and regions of human chromosomes linked to obesity in humans. Several of the genetic markers contained in Add Health are listed, but one should reasonably expect that they only account for a limited amount of variation in the health outcomes.

26. Examining whether genetic linkages occur is an active area of study as it presents a test of whether Mendel's law of independent assortment is supported. This law suggests that different genes are inherited independently of each other, and scientists have essentially concluded that there is an independent assortment of chromosomes during meiosis. However, alleles that are in close proximity on the same chromosome may be inherited as a group. Studies finding small links in genetic assortment have been obtained from samples consisting only of family members. However, there appears to be evidence that different groups of alleles are transmitted together across families when many of these studies and samples are examined jointly. Thus, violations are not systematic.

27. As discussed in the preceding endnote, this result is consistent with a large amount of evidence presented in the scientific literature.

28. Recall that Table 3 demonstrated that significant correlations do indeed exist between health outcomes and the genetic markers in our data. To construct the instrument set, we considered two alternative strategies. First, we followed Klepinger, Lundberg, and Plotnick (1999), who used forward stepwise estimation to select a subset of these markers and their interactions. This implementation is identical to Ding et al. (2006, 2009) and this approach has the advantage of making it easier to replicate the study. The scientific literature provides some (arguably weak) guidance for selecting particular markers, as the evidence tends to be inconsistent across studies, which tend to use very small unrepresentative clinical samples. We examined the robustness of our results by using the complete set of the markers in our study. The general pattern of IV and fixed effects IV results are robust to the instrument set for the full sample. The first stage properties are particularly weak for the full set of markers and their two by two interactions, yet the partial R-squared for that instrument set is substantially larger than studies using dates of birth in the labour economics literature. Finally, at the request of a seminar participant, we considered five other strategies based on either stepwise regression using different criteria or retaining those markers with significant relationships at the 5% level. Again the pattern of results was fairly consistent. These results are available from the authors upon request.

29. While the estimated effect for AD is quite large (approximately two standard deviations in the test score) in comparison to the estimated effects of depression and obesity, the effect size differences are consistent with differences in the typical age of onset of the health outcomes. For AD and HD, symptoms occur at a young age, typically during elementary school or earlier. In contrast, the age of onset for symptoms of depression is typically during middle adolescence. There is also emerging evidence that children seem to outgrow HD symptoms to some extent but not AD symptoms.

30. Many of the p-values are large and exceed 0.5. P-values are computed from Sargan tests of the joint Null hypothesis that the excluded instruments are valid instruments for the health variables in the achievement equation. Similarly with other instrument sets that we explored, we found evidence of large p-values above 0.2.

31. This is an F-statistic form of the Cragg and Donald (1993) statistic and requires an assumption of i.i.d. errors, which is more likely to be met in the specifications with family fixed effects. We are not aware of any studies on testing for weak instruments in the presence of non-i.i.d. errors.

32. We did conduct Kleibergen and Paap (2006) tests for the preferred instrument set reported in Table 5 and can reject the Null hypothesis at the 10% level. This suggests the matrix is of full rank and, while over-identified, the set does provide identification of the health variables.

33. For health vector 1, the results are 48.03 and 51.62.

34. The F-statistics also suggest that our empirical results in Table 5 are not driven by the instruments performing well in certain health equations and not in others.

35. Essentially, the procedure involves estimates of the second stage equation with the instrumented health vector where the instruments are additionally included in the specification. If the exclusion restriction assumption is satisfied, the coefficients on the instrument are not identified. To conduct the analysis, we assume a prior distribution for the estimated impact of these coefficients. In our analysis, the impacts are distributed $N(0,\delta 2)$, where δ is the q% percentage of the reduced form impact obtained from an OLS regression of academic achievement on the instruments and exogenous factors. We vary q to conduct our sensitivity analysis.

36. A large empirical literature has documented that subsequent fertility decisions are influenced by prior birth outcomes. For example, Angrist and Evans (1998) and Preston (1985), among others, have established that fertility decisions are influenced by sex composition of existing children as well as past neonatal or infant mortality.

37. For example, birth order, birth spacing, and sex composition have been shown to affect differential levels of investment by parents into children (e.g., Hanushek 1992; Black, Devereux, and Salvanes 2005; Conley and Glauber 2005).

38. See Rosenzweig and Wolpin (2000) for a discussion.

39. Results for the full subsample of twins (n=617) are available upon request. There are few differences in the significance and magnitude of the impacts from health variables.

40. For example, Chou et al. (2004) and Gruber and Frakes (2006) examine whether higher cigarette prices affected relative prices, thereby reducing smoking but increasing obesity. The former study finds evidence and the latter examines the robustness and suggests that many of the results are implausible.

41. The results reported in this subsection are robust to examining only the same-sex twin subsample.

References

Allen, G. 1970. Within Group and between Group Variation Expected in Human Behavioral Characters. *Behavior Genetics* 1 (3–4): 175–94.

Almond, D. 2006. Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy* 114 (4): 672–712.

Almond, D., L. Edlund, and M. Palme. 2008. Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden. *Quarterly Journal of Economics* 124 (4): 1729–72.

Angrist, J.D., and W. Evans. 1998. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review* 88 (3): 450–77.

Babinski, L.M., C.S. Hartsough, and N.M. Lasbert. 1999. Childhood Conduct

Problems, Hyperactivity-Impulsivity, and Inattention as Predictors of Adult Criminal Activity. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 40 (3): 347–55.

Behrman, J.R., and V. Lavy, 1998. Child Health and Schooling Achievement: Association, Causality and Household Allocations. CARESS Working Papers 97-23, University of Pennsylvania.

Benjamin, D., C. Chabris, E.L. Glaeser, and D. Laibson. 2009. Genetic Influences on Economic Outcomes. Paper presented at the annual meeting of the American Economic Association, San Francisco.

Black, S., P. Devereux, and K. Salvanes. 2005. The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *Quarterly Journal of Economics* 120 (2): 669–700.

Bleakley, H.C. 2007. Disease and Development: Evidence from Hookworm Eradication in the American South. *Quarterly Journal of Economics* 122 (1): 73–117.

Caplin, M., and M. Dean. 2008. Dopamine, Reward Prediction Error, and Economics. *Quarterly Journal of Economics* 123 (2): 663–701.

Caplin, M., M. Dean, P. Glimcher, and R. Rutledge. 2009. Measuring Beliefs and Rewards: A Neuroeconomic Approach. Mimeo, New York University.

Cesarini, D., C.T. Dawes, J.H. Fowler, et al. 2008. Heritability of Cooperative Behavior in the Trust Game. *PNAS* 105 (10): 3721–26.

Cesarini, D., C.T. Dawes, M. Johannesson, et al. 2009. Genetic Variation in Preferences for Giving and Risk Taking. *Quarterly Journal of Economics* 124 (22): 809–42.

Chou, S.Y., M. Grossman, and H. Saffer. 2004. An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics* 23 (3): 565–87.

Christensen, K., A. Wienke, A. Skytthe, et al. 2001. Cardiovascular Mortality in Twins and the Fetal Origins Hypothesis. *Twin Research* 4 (5): 344–49.

Conley, D., and R. Glauber. 2005. Parental Education Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility. NBER Working Paper w11302, National Bureau of Economic Research.

Conley, T., C. Hansen, and P.E. Rossi. 2007. Plausibly Exogenous. Mimeo, University of Chicago.

Cooper, R.S., J.S. Kaufman and R. Ward. 2003. Race and Genomics. *The New England Journal of Medicine* 348 (12): 1166–70.

Cragg, J.G., and S.G. Donald. 1993. Testing Identifiability and Specification in Instrumental Variables Models. *Econometric Theory* 9 (2): 222–40.

Currie, J., and M. Stabile. 2006. Child Mental Health and Human Capital Accumulation: The Case of ADHD. *Journal of Health Economics* 25 (6): 1094–1118.

Cutler, D., and A. Lleras-Muney. 2007. Education and Health: Evaluating Theories and Evidence. NBER Working Paper w12352, National Bureau of Economic Research.

de Quervain, D.J.F., and A. Papassotiropoulos. 2006. Identification of a Genetic Cluster Influencing Memory Performance and Hippocampal Activity in Humans. *Proceedings of the National Academy of Sciences* USA 103 (11): 4270–74.

Ding, W., S.F. Lehrer, J.N. Rosenquist, et al. 2006. The Impact of Poor Health on Education: New Evidence Using Genetic Markers. NBER Working Paper w12304, National Bureau of Economic Research.

———. 2009. The Impact of Poor Health on Academic Performance: New Evidence Using Genetic Markers. *Journal of Health Economics* 28 (3): 578–97.

Dreber, A., C.L. Apicella, D.T.A. Eisenberg, et al. 2009. The 7R Polymorphism in the Dopamine Receptor D4 Gene (DRD4) is Associated with Financial Risk-Taking in Men. *Evolution and Human Behavior* 30 (2): 85–92.

Dremencov, E., I. Gispan-Herman, M. Rosenstein, et al. 2004. The Serotonin-Dopamine Interaction is Critical for Fast-Onset Action of Antidepressant Treatment: In Vivo Studies in an Animal Model of Depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 28 (1): 141–7.

Fletcher, J.M. 2008. Adolescent Depression and Educational Attainment: Evidence from Sibling Fixed Effects. *Health Economics* 17: 1215–35.

Fletcher, J.M., and B.L. Wolfe. 2008a. Long-term Consequences of Childhood ADHD on Criminal Activities. Mimeo, Yale University.

———. 2008b. Child Mental Health and Human Capital Accumulation: The Case of ADHD Revisited. *Journal of Health Economics* 27 (3): 794–800.

Glewwe, P., and H. Jacoby. 1995. An Economic Analysis of Delayed Primary School Enrollment in a Low-Income Country—the Role of Early Childhood Nutrition. *Review of Economics and Statistics* 77 (1): 156–69.

Goldstein, D.B., S.K. Tate, and S.M. Sisodiya. 2003. Pharmacogenetics Goes Genomic. *Nature Reviews Genetics* 4 (12): 937–47.

Goodman, E., B.R. Hinden, and S. Khandelwal. 2000. Accuracy of Teen and Parental Reports of Obesity and Body Mass Index. *Pediatrics* 106 (1): 52–58.

Grossman, M. 1975. The Correlation between Health and Schooling, in Household Production and Consumption. In *Studies in Income and Wealth. Conference on Research in Income and Wealth*, vol. 40, edited by N.E. Terleckyj. New York: Columbia University Press for the National Bureau of Economic Research.

Grossman, M., and R. Kaestner. 1997. Effects of Education on Health. In *The Social Benefits of Education*, edited by J.R. Behrman, and N. Stacey. Ann Arbor: University of Michigan Press.

Gruber, J., and M. Frakes. 2006. Does Falling Smoking Lead to Rising Obesity? *Journal of Health Economics* 25 (2): 183–97.

Hanushek, E. 1992. The Trade-off between Child Quantity and Quality. *Journal of Political Economy* 100 (1): 84–117.

Harris, K.M., F. Florey, J. Tabor, et al. 2003. The National Longitudinal Study of Adolescent Health: Research Design, available at http://www.cpc.unc.edu/projects/addhealth/design.

Harrison, G.A. 1970. Human Variation and Its Social Causes and Consequences. In *Proceedings of the Royal Anthropological Institute of Great Britain and Ireland*, 1970, 5–13.

International HapMap Consortium. 2005. A Haplotype Map of the Human Genome. *Nature* 437 (7063): 1299–1320.

Jain, A.K., S. Prabhakar, and S. Pankanti. 2002. On the Similarity of Identical Twin Fingerprints. *Pattern Recognition* 35 (2): 653–63.

Johnson, J.A. 2003. Pharmacogenetics: Potential for Individualized Drug Therapy Through Genetics. *Trends Genetics* 19 (6): 60–66.

Kaestner, R., and M. Grossman. 2008. Effects of Weight on Children's Educational Achievement. NBER Working Paper 13764, National Bureau of Economic Research.

Kelada, S.N., D.L. Eaton, S.S. Wang, et al. 2003. The Role of Genetic Polymorphisms in Environmental Health. *Environmental Health Perspectives* 111 (8): 1055–64.

Kessler, R.C., L.A. Adler, R. Barkley, et al. 2005. Patterns and Predictors of Attention-Deficit / Hyperactivity Disorder Persistence into Adulthood: Results from the National Co-morbidity Survey Replication. *Biological Psychiatry* 57 (11): 1442–51.

Kleibergen, F., and R. Paap. 2006. Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics* 133 (1): 97–126.

Klepinger, D.S.L., S. Lundberg, and R. Plotnick. 1999. How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women? *The Journal of Human Resources* 34 (3): 421–48.

Kremer, M., and E. Miguel. 2004. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica* 72 (1): 159–217.

Lunde, A., K.K. Melve, H.K. Gjessing, et al. 2007. Genetic and Environmental Influences on Birth Weight, Birth Length, Head Circumference, and Gestational Age by Use of Population-based Parent-offspring Data. *American Journal of Epidemiology* 165 (7): 734–41.

Merikangas, K.R., and N. Risch. 2003. Genomic Priorities and Public Health. *Science* 302 (5645): 599–601.

Moises, H.W., R.M. Frieboes, P. Spelzhaus, et al. 2001. No Association between Dopamine D2 Receptor Gene (DRD2) and Human Intelligence. *Journal of Neural Transmission* 108 (1): 115–21.

Norton, E.C., and E. Han. 2008. Genetic Information, Obesity, and Labor Market Outcomes. *Health Economics* 17 (9): 1089–1104.

Olds, J., and P. Milner. 1954. Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of Rat Brain. *Journal of Comparative and Physiological Psychology* 47 (6): 419–27.

Perri, T.J. 1984. Health Status and Schooling Decisions of Young Men. *Economics of Education Review* 3 (3): 207–13.

Petrill, S.A., R. Plomin, G.E. McClearn, et al. 1997. No Association between General Cognitive Ability and the A1 Allele of the D2 Dopamine Receptor Gene. *Behavior Genetics* 27 (1): 29–31.

Plomin, R., J.K.J. Kennedy, and I.W. Craig. 2006. The Quest for Quantitative Trait Loci Associated with Intelligence. *Intelligence* 34 (6): 513–26.

Preston, S.H. 1985. Mortality in Childhood: Lessons from WFS. In *Reproductive Change in Developing Countries*, edited by J.G. Cleland and J. Hobcraft. Oxford: Oxford University Press.

Roberts, R.E., Lewinsohn, P.M., and J.R. Seeley. 1991. Screening for Adolescent Depression: A Comparison of Depression Scales. *Journal of the American Academy of Child & Adolescent Psychiatry* 30 (1): 58–66.

Rosenzweig, M.R., and K.I. Wolpin. 2000. Natural "Natural Experiments" in Economics. *Journal of Economic Literature* 38 (4): 827–74.

Royer, H. 2009. Separated at Girth: US Twin Estimates of the Effects of Birth Weight. *American Economic Journal: Applied Economics* 1 (1): 49–85.

Stock, J.H., and M. Yogo 2005. Testing for Weak Instruments in Linear IV Regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by D.W. Andrews and J.H. Stock. Cambridge: Cambridge University Press.

Strauss, J., and D. Thomas. 1998. Health, Nutrition, and Economic Development. *Journal of Economic Literature* 36 (2): 766–817.

Zerhouni, E. 2003. Medicine. The NIH Roadmap. *Science* 302 (5642): 63–72.

Understanding and Addressing Barriers to Adolescent Mental Health Service Retention: An Intervention to Enhance Engagement

Heather Spielvogle

Abstract

Approximately one-fifth of Canadian children experience a mental health problem which may complicate their development and impede their future potential. Child and adolescent mental health services (CAMHS) aim to provide interventions that ameliorate mental health problems and encourage successful transitions to adulthood. Nonetheless, many children and adolescents who may benefit from CAMHS either do not access services or terminate services prematurely. This paper presents a literature review on the effectiveness of psychotherapies and barriers to access and participation commonly found in CAMHS. The purpose of this review is to provide a framework for an engagement intervention to encourage adolescent participation in the therapeutic relationship. The components of this engagement intervention are described and suggestions for clinical practice are outlined. Since the engagement intervention is designed to explore adolescents' perceptions of CAMHS barriers, it is hoped that the intervention may be used to inform CAMHS about these barriers and, subsequently, enhance service delivery.

Introduction

Mental health problems are a significant concern among Canadian children (Health Canada 2002; Ontario Ministry of Children and Youth Services 2006; Waddell et al. 2005). Affecting an estimated 14–21% of Canadian children, mental health problems are associated with economic burden and detrimental outcomes, including criminality, academic failure, teen pregnancy, and suicide (Fergusson, Horwood, and Ridder

2007; Fergusson and Woodward 2002; Kessler et al. 1995; Kessler et al. 1997; Simonoff et al. 2004; Waddell et al. 2002). Interventions provided by CAMHS (e.g., parent training, individual/family psychotherapies, and pharmacotherapy) may provide an opportunity to alleviate child mental health problems and inhibit the related consequences. However, the impact of CAMHS is compromised by low attendance and high drop-out rates among families and children in these settings (Edlund et al. 2002; Kazdin, Holland, and Crowley 1997). In order to improve retention in CAMHS, it is important to understand and address the barriers that contribute to service drop-out.

Among all age groups entering mental health services, adolescent-aged children (i.e., individuals between the ages of 12 and 18) may be the most critical population to engage. Suicide, the most significant outcome of mental health problems, is the second leading cause of death among Canadian adolescents and young adults (Langlois and Morrison 2002). Although interventions which identify suicide risk factors (Aseltine and DeMartino 2004) and prevent repeat suicide behaviour (Links, Bergmans, and Cook 2004) are available, many teens forgo formal interventions (Kalafat and Elias 1995). Adolescence is a period when risk and protective factors have a great impact on developmental trajectories (Cichetti and Toth 1996; Rutter 1990), and mental health problems, such as major depression, have been linked to adolescent learning difficulties, low self-esteem, and impaired peer relationships (Kovacs and Golston 1991; Prinstein et al. 2005; Rudolph et al. 2001). Moreover, adolescent-onset mental health problems can affect psychosocial functioning well into adulthood and increase risk for subsequent, more severe mental health problems (Lewinsohn et al. 1999; Os et al. 1997; Rutter, Taylor, and Hersov 1994; Wolkind and Rutter 1985). Yet, among all age groups, adolescents and young adults demonstrate the highest drop-out rates in mental health services (Edlund et al. 2002), so many who are at risk for detrimental psychosocial outcomes may not receive the full benefit of services.

This paper will begin with a discussion of the effectiveness of psychotherapies largely provided by CAMHS before turning to the primary focus of potential contributors to access and drop-out in CAMHS. Although the majority of this review will focus on the service barriers experienced by caregivers of children and adolescents, the limited discourse on adolescent barriers is reviewed. Much of the process literature on mental health service barriers and engagement has focused on adults entering individual psychotherapy. Although some of the concepts related to adult barriers and engagement apply to adolescents,

it is beyond the scope of this review to include this literature. This review will provide a framework for an intervention which is currently being evaluated to improve CAMHS attendance and develop an understanding of adolescent-defined service barriers.

The Effectiveness of Child and Adolescent Psychotherapies

Before discussing the importance of CAMHS engagement, one key supposition to establish is whether adolescents who receive psychotherapy are better off than those who do not (Kazdin 2000). An accumulating body of research has demonstrated the beneficial effects of child and adolescent psychotherapies (e.g., Chorpita et al. 2002; Hoagwood 2005; Kazdin et al. 1990; Weisz et al. 1995). Within the last 50 years, over 1,500 controlled outcome studies of child and adolescent psychotherapies were completed (Kazdin 2000) and meta-analyses of this research (Casey and Berman 1985; Kazdin et al. 1990; Weisz et al. 1995) have demonstrated that psychotherapy reduces the symptoms of mental health problems when compared to no treatment, with effect sizes of approximately .70 for a range of psychotherapies (Kazdin 2000). However, the vast majority of this literature describes results from efficacy studies in which psychotherapy was delivered within research clinics rather than community mental health centres that are more representative of real world scenarios.

More recently, investigations of community-based CAMHS have been shown to effectively improve functioning (i.e., functioning in school, home, and the community) and to decrease symptoms associated with mental health problems (Angold et al. 2000; Daleiden et al. 2006). Moreover, investigators are starting to disseminate evidence-based practices to community settings and witness positive results. For instance, interventions such as social skills training for externalizing disorders (i.e., attention deficit disorder, conduct disorder, and oppositional defiant disorder) (Tynan, Schuman, and Lambert 1999), interpersonal therapy for depression (Mufson et al. 2004; Santor and Kusumakar 2001), and cognitive behavioural therapy for depression and post-traumatic stress disorder (Asarnow et al. 2005; Stein et al. 2003) have all been administered in community settings and have demonstrated reductions in acute symptom severity.

On a larger scale, mental health services are beginning to evaluate their interventions and integrate evidence-based practices to ensure the best quality of care is offered to children and adolescents. For example, in 2002, the Child and Adolescent Mental Health Division of Hawaii's Department of Health began integrating evidence-based practices by

providing training and supervision to front-line staff and offering improved care coordination to ensure that services were delivered seamlessly (Chorpita et al. 2002). Evaluations were conducted annually using measures such as the Child and Adolescent Functional Assessment Scale (CAFAS) (Hodges, Wong, and Latessa 1998). Outcome results indicated that, over a five-year period, children who were receiving services were progressively experiencing better functioning, as measured by the CAFAS, with the median rate of change significantly improving across quarters t14 = *j*9.65, p G .001, R^2 = 0.87. Additionally, over a four-year period, the length of service episodes was reduced by approximately 40%. The authors suggested that even though children were progressively spending less time in services, they were improving at a faster rate because the quality of services was enhanced and wait-list times decreased (Daleiden et al. 2006).

Although the research suggests that community-based CAMHS can be effective, the quality of psychotherapy offered at CAMHS varies widely between centres and, in some cases, may have no effect (Weersing and Weisz 2002). Greater measures are being taken to ensure the services children and adolescents receive are effective. In Ontario, the Ministry of Children and Youth Services has begun requiring agencies to complete standardized mental health assessments to screen adolescents and assess outcomes (Children's Mental Health Ontario 1999). While the outcomes of child mental health services in Ontario have not been reported, the assessments may increase accountability and identify areas for improvement and funding allocation.

While child and adolescent psychotherapies may reduce symptoms and improve functioning at service termination, the long-term effectiveness of these psychotherapies is unknown. Most of the psychotherapy literature presents results from efficacy studies in which procedures for psychotherapy are highly controlled and child and adolescent participants typically experience one particular mental health problem. Moreover, efficacy research does not entirely reflect the conditions of community-based CAMHS, where the duration of psychotherapy is more limited and clients may present with multiple social and mental health problems.

Although further research is needed to improve CAMHS quality and delivery, the effectiveness research largely suggests that evidence-based psychotherapies offered in CAMHS can be beneficial. However, underuse of CAMHS is problematic, especially when evidence suggests that the children and adolescents most likely to terminate following intake are

from vulnerable, marginalized populations and have more serious presenting issues (Kazdin 1993; Tuma 1989). In an effort to reach a broader spectrum of children and adolescents, it is important to consider the barriers and enablers of treatment engagement in order to alleviate this public health and service delivery problem.

Barriers to CAMHS

High drop-out rates and poor mental health treatment attendance are very common problems in Ontario, the only province in which it has been rigorously evaluated, and the United States (Edlund et al. 2002; Kessler et al. 1997; Romney 1988; Wang et al. 2005). Estimated rates of early mental health treatment termination range from 17 to 45% with the highest rates being demonstrated by adolescents and young adults (Edlund et al. 2002; Romney 1988). Within CAMHS settings, even greater drop-out rates between 45 and 70% have been reported (Armbruster and Kazdin 1994; Kazdin et al. 1997; Kazdin and Mazurick 1994). Clearly, efforts toward understanding the needs of families, children, and adolescents and how to engage them in services are warranted.

The majority of child mental health services research on drop-out has investigated demographic variables associated with mental health service use. The collective results of this research demonstrate that, with the exception of socio-economic and ethnoracial minority status, commonly measured variables (i.e., age of child, gender, and type/severity of emotional or behavioural problems) are inconsistent predictors of attrition (Armbruster and Kazdin 1994; Ambruster and Fallon 1994; Cohen and Hesselbart 1993; Dishion and Patterson 1992; Gasquet et al. 1997; Griffen, Cicchetti, and Leaf 1993; Harpaz-Rotem, Leslie, and Rosenheck 2004; Keely and Weins 2008; McKay et al. 2001; Padgett et al. 1993; Schonert-Reichl and Mueller 1996; Verhulst and Van Der Ende 1997; Wu et al. 2001). Furthermore, it is unclear how these commonly studied demographic characteristics interact to effect attrition (McKay and Bannon 2004) or why socio-economic and ethnoracial minority status are associated with drop-out.

Within the last decade, investigations have begun to emerge which explore potential barriers to access and continued participation in CAMHS (Anderson et al. 2006; Chandra and Minkovitz 2006; Kazdin and Wassell 1999; Moli, Berriman, and Evans 2008; Morrissey-Kane and Prinz 1999; Owens et al. 2002; Sheffield, Fiorenza, and Sofronoff 2004; Stevens et al. 2006). On the whole, this research has examined the perspectives of caregivers/parents and service providers to identify barriers to both initial service access and ongoing service participation.

Although, this paper focuses primarily on understanding and addressing barriers to improve ongoing service participation, a brief discussion of the literature on common barriers associated with service access follows.

Barriers to CAMHS Access

Families and adolescents who could otherwise benefit from CAMHS may not access them because of structural or perceptual barriers (Owens et al. 2002). Common structural barriers include factors such as lack of financial resources (Lewit, Terman, and Berhman 1997), inconvenient service locations (Catron and Weiss 1994), long wait-lists (Anderson et al. 2006), and poor service coordination (Owens et al. 2002). Perceptual barriers include factors such as a lack of awareness or misperception of the severity or impact of mental health problems (ibid.), stigma associated with acknowledging mental health problems or seeking mental health services (Chandra and Minkovitz 2006; Copeland 2006; Sheffield, Fiorenza, and Sofronoff 2004), and past negative experiences with mental health providers (Flisher et al. 1997; Owens et al. 2002).

A limited amount of research has explored the impact of structural barriers on service use. For instance, wait-lists are a substantial barrier to service access (Anderson et al. 2006; Greeno et al. 2002) and when agencies offer services within three weeks of request, significantly more families access CAMHS (Greeno et al. 2002). In addition to decreasing wait times for services, providing mental health treatment in community settings such as schools and community centres may alleviate inaccessibility. For instance, in a study of children and families from socio-economically disadvantaged backgrounds, Catron and Weiss (1994) found that only 17% of individuals who completed intake at CAMHS specialty centres subsequently began services, compared to 98% of individuals who began services after completing intake at school-based mental health services. Similarly, in Ontario, the Community Parent Education Program has demonstrated significant improvements in treatment attendance among low-income families from diverse ethnocultural backgrounds when interventions were conducted in community settings rather than mental health specialty centres (Cunningham, Bremner, and Boyle 1995; Cunningham, Bremner, and Secord 1998).

Perceptual barriers such as lack of awareness or misperception of mental health issues may be addressed through public education programs or through increased screening in general medical services. While many parents and adolescents have an awareness of mental health issues, a number of families will avoid services because of the social stigma

attached to attending CAMHS (Chandra and Minkovitz 2006). Chandra and Minkovitz (2006) recommend that school health programs explore teen perceptions in order to address stigma and normalize the experience of psychological distress. In addition, providing CAMHS in community settings rather than mental health specialty centres may decrease the stigma attached to service access.

Past negative experiences with mental health providers are arguably the most significant perceptual barrier to treatment which disproportionately affects economically disadvantaged individuals from ethnoracial minority groups (Belle 1990). Prior to entering CAMHS, parents and children may have encountered considerable discrimination from social service providers or experienced a discrepancy between the services they expected to receive and those that are offered. Many of those families who have experienced these barriers may either refuse to return for services or do so with a degree of mistrust which ultimately inhibits the bond between the family and service provider.

To reduce perceptual barriers, CAMHS need to become aware of the cultural factors leading to the needs and expectations of the populations and communities they serve; however, quality assurance measures alone are often insufficient. In general, policy changes to minimize structural barriers, public education programs to increase awareness and address stigma, and quality improvement initiatives are needed to increase the acceptability and relevance of CAMHS for a diverse range of individuals and families in need.

Barriers to CAMHS Participation

From a clinical perspective, once service is initiated the barriers which contribute to drop-out or consistent participation are the most significant issues to address. A series of articles have demonstrated that when adolescents and their parents enter services both logistical (e.g., insufficient time, lack of transportation) and perceptual barriers contribute to drop-out (Garcia and Weisz 2002; Kazdin, Holland, and Crowley 1997; MacNaughton and Rodrigue 2001; McCabe 2002). However, perceptual barriers (i.e., weak therapeutic alliance, low perceived need for treatment, low perceived relevance of treatment) have been shown to be a significantly greater predictor of treatment discontinuation than logistical barriers (Garcia and Weisz 2002; Kazdin, Holland, and Crowley 1997; Koren et al. 1997; MacNaughton and Rodrigue 2001; Stevens et al. 2006). Moreover, a strong therapeutic alliance, defined as the bond between the client and therapist which develops from the client's trust that mutually agreed on tasks will help

the client meet his goals (Bordin 1979), has been shown to mediate the impact of logistical barriers and improve CAMHS attendance (Kazdin, Holland, and Crowley 1997; Kazdin and Whitley 2003). Since the foundation for a strong therapeutic alliance begins with the first therapeutic contact, effective treatment engagement is critical.

A small number of interventions have been developed to improve attendance rates and enhance engagement among families (McKay, McCadam, and Gonzales 1996; Nock and Ferriter 2005; Szapocznick et al. 1988; Swartz et al. 2007). In general, those engagement interventions focus on building a collaborative relationship with the parent, enhancing the parent's self-efficacy, and providing parents with the space to talk about their challenges. Additionally, the parent's logistical barriers are discussed and, when possible, addressed before treatment begins. All of these strategies have significantly increased attendance at initial clinic appointments. However, other important measures of engagement, such as the child's motivation for services and the degree to which the child follows service recommendations, have not been explored.

Thus far, all of the engagement interventions focus primarily on engaging parents and addressing the barriers they experience bringing their child to mental health treatment. Although parent engagement is important, it is just as crucial to engage the main recipient of the services (i.e., the child or adolescent) and understand barriers from her point of view.

Adolescent Engagement

Adolescents experience a number of social, physical, and cognitive developmental changes which may increase risk for dysfunction on the one hand and create opportunities for healthy development on the other (Cicchetti and Rogosch 2002; Cicchetti and Toth 1996; Steinberg 2005). Although the majority of adolescents will progress through adolescence with maturing skills and improvements in mental health, some will experience greater levels of psychopathology (Holmbeck et al. 2005; Rutter 1990). Cicchetti and Toth (1996) argue that the instability produced by multiple developmental transitions may create opportunities for adolescents, once engaged, to become more responsive to interventions. Thus, adolescence may be the most opportune time for CAMHS to promote enduring positive change.

Nevertheless, adolescents are considered more difficult to engage in CAMHS than adults and children (Diguiseppe, Linscott, and Jilton 1996). Adolescents rarely seek services on their own accord (Boldero and Fallon 1995; Sheffield, Fiorenza, and Sofronoff 2004) and are often

coerced into attending services by their parents or schools. Unfortunately, individuals who are forced to attend services are rarely invested in changing their behaviour (Kazdin et al. 1990; O'Malley 1990) and are less likely to develop a therapeutic alliance with their therapist or participate fully in treatment (Luborsky 1976). Low investment in change and poor alliance are two key predictors of drop-out and poor outcomes (Coatsworth et al. 2001; Dishion and Patterson 1992; Hawke, Hennen, and Gallione 2005; Hawley and Garland 2008; McKay et al. 2001). Alternatively, individuals who make the choice to attend services may be more likely to use psychotherapy effectively and experience favourable outcomes. For instance, Zuroff and colleagues (2007) found that, among adults with depression, those who felt that it was their own choice to participate in psychotherapy were more likely to achieve remission and had lower symptom severity post-treatment than those who did not feel the choice to participate was made freely. Thus, autonomy seems an important predictor of therapeutic alliance, psychotherapy adherence, and subsequent therapeutic outcome.

Addressing the challenge of adolescent treatment motivation and engagement is a key issue for CAMHS. However, adolescents' perspectives on factors which facilitate or prevent service participation have not been well researched. A few studies have begun to explore adolescents' concerns with respect to CAMHS (Chandra and Minkovitz 2006; Eyrich-Garg 2008; Hawley and Garland 2008; Sheffield, Fiorenza, and Sofronoff 2004). For instance, Chandra and Minkovitz (2006) found that 51% of teens (n=274) who were not receiving CAMHS reported that they did not think receiving counselling would help resolve mental health problems. In addition, of the adolescents who entered CAMHS, those who reported a weak therapeutic alliance more likely to drop out prematurely (Hawley 2003) and were less satisfied with therapy (Hawley and Garland 2008).

Although it is difficult to determine what causes teens to perceive CAMHS as ineffective or irrelevant, two related factors may contribute to these beliefs. First, in my professional experience, many adolescents seem to have the preconceived notion that counselling or therapy is about "telling my feelings to an adult" rather than resolving the issues that are important to them. Discussing feelings with an unknown adult can be incredibly uncomfortable for many adolescents. For instance, adolescent boys who follow traditional masculine ideals and adolescents who have experienced abuse or neglect have more difficulty expressing or feeling comfortable with emotion (Barrett and White 2002; Scannapieco and Connell-Carrick 2005). Second, adolescents may feel

that the counsellor will not take their opinions into account, listen, or respect them (Eyrich-Garg 2008). Developmentally, adolescents experience an increasing need for autonomy (Steinberg 2005) and are particularly concerned about having their therapist's or parents' goals forced on them (Diguiseppe, Linscott, and Jilton 1996; Eyrich-Garg 2008). If an adolescent does make the changes ascribed by his parents or therapist, the changes are frequently temporary. Moreover, engagement is unlikely to occur if the adolescent feels that the therapist does not understand her perspective and inadequately focuses on her concerns.

Without adequate empirical knowledge on adolescent engagement, researchers have drawn on the therapeutic process literature to develop interventions to improve attendance (Oetzel and Scherer 2003). Although, most of the process literature has focused on adult engagement interventions which may be applied to adolescents (Eyrich-Garg 2008; Morris and Nicholson 1998), these interventions are not completely generalizable to adolescent concerns. In an effort to address this gap, I sought to develop an engagement intervention specifically for adolescents entering CAMHS.

Adolescent Engagement Intervention

The "adolescent engagement intervention" (AEI) is a brief, pre-counselling interview which is intended to improve the likelihood that teens will attend services and to enhance the current knowledge of teen-defined barriers to services. It is expected that the intervention will improve service retention by exploring adolescents' potential ambivalence toward attending services and making personal changes. Additionally, a component of the intervention will be used to help teens elaborate on their perceptions of entering services and barriers which potentially impact attendance. The AEI was adapted from an adult engagement intervention developed in partnership with colleagues at the University of Pittsburgh (Zuckoff et al. 2004).

The AEI is partially supported by the concepts and strategies of Motivational Interviewing (MI), a client-centred, directive method for enhancing intrinsic motivation for change by exploring and resolving ambivalence (Miller and Rollnick 2002). In keeping with MI, the engagement intervention uses the Transtheoretical Model of Change (Prochaska and DiClemente 1984) to explain behaviour and guide practice. This model of behaviour change describes six phases through which individuals transition: (1) pre-contemplation (no intention to change behaviour in the foreseeable future exists); (2) contemplation (the person is aware that a problem exists and is thinking about

overcoming it); (3) preparation (the person is planning to take action and may take small steps toward change, but is not fully committed); (4) action (the person modifies behaviour, experiences, and environment in order to overcome the problem); (5) maintenance (the person works hard to prevent relapse and tries to consolidate gains made); and (6) termination (the person has completed the change process and no longer needs to work to prevent relapse). Since the majority of teens will enter CAMHS in either the pre-contemplation or contemplation stage of change (Baer and Peterson 2002), the AEI primarily focuses on intervening with adolescents in either of these two stages.

The AEI consists of four topic areas: the story, perception of problem, barriers to participation, and the summary. The "story" section explores the adolescent's feelings about coming to service and the circumstances that led to a referral. The "perception" section focuses on the adolescent's view of the issue that brought him to services and if that view is consistent with other's perceptions. The "barriers" section explores logistic and perceptual barriers to receiving services. In particular, topics such as stigma, past mental health provider experiences, and potential cultural misunderstandings are reviewed. The "summary" section aims to prepare the teen for the next phase of services and to elicit a commitment, even if provisional, to participate in services.

The Story

The AEI begins by offering the adolescent an opportunity to tell her "story" about her experience of coming to services. Therapists who exhibit a non-judgmental stance and respect the adolescent's perspectives are more effective at engaging when compared to those who use a more neutral stance (Oetzel and Scherer 2003; Rubenstein 1996). Thus, it is important to allow the adolescent to share his views while giving them the opportunity to explain his behaviour without judgment. Many adolescents who are coerced into attending services are angry about talking to a therapist (Baer and Peterson 2002). In these circumstances, giving the adolescent the space to talk about his displeasure or eliciting these feelings may help him feel understood and demonstrate a non-defensive attitude.

In addition to offering the adolescent the opportunity to talk about her experience or displeasure, it is also important to give her the choice to share as much or as little as she wishes. In a time of limited resources, it is not uncommon for agencies to require that extensive background information is collected during the first few therapy contacts. In some cases, this requires therapists to ask clients to discuss past experiences which increase the clients' emotional vulnerability before trust is

established. Not only can this experience cause an adolescent to feel overwhelmed with emotion and unsafe, it can also cause the youth to feel that the therapist wishes to establish power over him (Scannapieco and Connell-Carrick 2005). If possible, providing the adolescent with the opportunity to postpone discussing the details of sensitive information can prevent disengagement. As common protocol, the therapist should also educate the adolescent about confidentiality and limitations to confidentiality. Providing the adolescent with personal choice in regards to what is discussed in therapy may enhance the therapy's relevance leading to greater motivation and engagement (Hanna and Hunt 1999; Oetzel and Scherer 2003; Loar 2001).

The goal in the opening phase of the engagement intervention is to ensure that the teen feels understood by capturing her view of her current situation. This discussion also helps the therapist make an assumption in regards to whether the teen falls in the pre-contemplation or contemplation stage of change. Asking the teen about her perception of entering services (i.e., her thoughts about why she is in services, what coming to services is like for her) will help the therapist understand the teen's stage of change and direct the intervention accordingly. Often, the stage of change will be apparent within the first few minutes of the interview; one key indicator is the teen's willingness to discuss the presenting issue that prompted services. If the teen chooses to focus the discussion on his reluctance to come to services rather than the presenting issue, the therapist will assume that the teen is in the pre-contemplation stage of change. If the teen discusses the presenting issue with minimal prompting, the therapist will assume the teen is in the contemplation stage of change.

The middle sections of the AEI are directed by the teen's assumed stage of change. If the teen presents in the pre-contemplation stage of change, a discussion of barriers to services will take place prior to a discussion of her perception of the presenting issue. If the adolescent seems unwilling to discuss the presenting issue, the therapist may opt to ask the teen to discuss the potential benefits or drawbacks of attending services. If the teen readily discusses the presenting issue, a discussion about the impact of this issue will follow.

Perception

The goal of this section of the AEI is to elicit communication about the presenting issue and to explore reasons for and against change. This section is primarily intended for teens in the contemplation stage of change. In some cases, teens who present in the pre-contemplation stage of change may be willing to explore some aspects of this section once

they have developed a level of trust with the therapist. This section usually begins with a discussion of the difference between the teen's perception of the presenting issue verses the referral source's perception. Adolescents are asked to explore their view of other people's concerns about the issue and if they consider these concerns realistic.

Before exploring the advantages and disadvantages of maintaining the behaviour, it is useful to understand the adolescent's values. In general, adolescents tend to value peer and/or family approval, respect, independence, and empathy (Goff and Goddard 1999; Kahle 1983). Understanding the adolescent's values will help the therapist determine if the presenting issue is in conflict with those values. Discussing the discrepancy between the teen's behaviour and values can help enhance the perceived importance of change (Miller and Rollnick 2002).

The positive aspects of the current behaviour or negative consequences of changing are discussed first. Once the therapist expresses that she hears the adolescent's reasons for not making a particular change, the teen may spontaneously discuss reasons why they may change. Some teens may be less willing to discuss the negative aspects of the behaviour or the advantages of changing, and the therapist should gauge the degree to which this can be introduced and discussed. A variety of MI techniques may be used to increase the likelihood that the adolescent will discuss reasons for changing behaviour, including reflecting the adolescent's statements or reflecting in a manner that exaggerates the rationale for not changing (Miller and Rollnick 2002). Once the positive and negative aspects of the behaviour have been explored, the therapist may summarize how the behaviour may not be entirely consistent with the adolescent's values. Ultimately, the goal is to encourage the teen to articulate reasons to change and alternatives to his current behaviour.

If the adolescent readily discusses the reasons for change, the therapist may gather more information about how the teen has attempted to modify or cope with the behaviour in the past. As the adolescent identifies her past coping efforts, an attempt should be made to understand the adolescent's level of self-efficacy. Self-efficacy is one's belief in one's ability to successfully make a change (Bandura 1986) and is an important mediator in an individual's decision to change (Caprana et al. 1998). Focusing on the strengths of these approaches and offering hope may create a self-fulfilling prophecy enabling the adolescent to achieve positive change (Strupp 1993).

This discussion segues into an exploration of the possibilities and challenges counselling may provide. Specifically, the teen's thoughts about counselling should be elicited and the therapist should provide

common information about the purpose and structure of counselling. At this point, the interview turns to a discussion of the potential obstacles and barriers that may interfere with treatment participation and attendance, if not previously discussed.

Exploration of Barriers

This phase of the intervention is designed to uncover potential barriers before they interfere with service participation. Practical barriers are usually explored first because they are less personal than the perceptual barriers. Common practical barriers may include issues related to time. For instance, many adolescents have other after school obligations, such as work or activities. As the practical barriers are resolved, the adolescent may talk about other concerns he has about services. However, if these barriers are not offered, the therapist should probe about perceptual barriers. For instance, the therapist should ask the adolescent about her attitudes or beliefs about mental health problems or services. A discussion of the adolescent's views of mental health problems may be especially helpful if the adolescent is from a different cultural background than the therapist. The therapist may provide psychoeducation about the prevalence of mental health problems to help normalize the adolescent's experience (Oetzel and Scherer 2003).

If the adolescent has had past experiences with mental health providers, a discussion of the helpful or unhelpful aspects of those encounters can aid the therapist in understanding the adolescent's preferences for subsequent encounters. In addition, it is useful to ask the adolescent how he might let the therapist know if the latter misses an important point or offers suggestions that are unreasonable. Some adolescents feel that it is disrespectful to question their therapist's impressions; however, if the therapist welcomes feedback, the teen is more likely to address misinterpretations and ineffective suggestions before they impair the therapeutic relationship. A discussion of cultural barriers should begin at the first session and continue throughout the course of therapy. A conversation regarding how the adolescent feels about seeing a therapist who is of a different racial or ethnic background serves as an entry point into further discussions of cultural barriers. Although it is beyond the scope of this paper to comprehensively review cultural factors which may impede engagement, therapists are encouraged to consult an authoritative source to guide practice. One such source is the text *Counseling the Culturally Different: Theory and Practice* (Sue and Sue 2003), however, an effort should be made to avoid overgeneralizing as each teen will have different levels of acculturation. At the end of this

section, if the adolescent is ready, she is encouraged to discuss her goals and how the therapist might help her reach those goals.

Eliciting Commitment

The goal in this final phase of the engagement intervention is to elicit a provisional commitment to attend services. The therapist may emphasize that the adolescent does have a choice over what to discuss in sessions and that it is acceptable to tell the therapist if something is not helpful. The session concludes with a review of the adolescent's strengths and the instillation of hope to build self-efficacy towards change.

Future Directions

This paper summarizes an approach which explicitly attends to the process elements and content areas to explore in order to elucidate the barriers which can enable therapists to enhance therapeutic engagement. Not only can therapeutic engagement strategies potentially improve outcomes and reduce early drop-outs by addressing the gap between service needs and service delivery, they can also be used inform the development of CAMHS.

Providing timely CAMHS that adolescents and their families find useful and pertinent to their social environments may be an effective avenue towards treatment engagement. To create mental health services that are more relevant to adolescents, eliciting their views on what is most likely to effect change is important and implementing those suggestions may improve service acceptability and outcome. Unfortunately, such information is rarely used to develop services or practices (Worrall-Davies and Marino-Francis 2008). By using the AEI, I hope to gain an understanding of the hopes, fears, perceptions, barriers, and enablers related to adolescent mental health service engagement. This research will be used in part to improve the way services are provided and build more effective adolescent engagement strategies and context relevant interventions.

Acknowledgements

The author would like to thank the Lupina Foundation for its support and Drs. Holly Swartz and Allan Zuckoff for their guidance in their development of the engagement session for adults from which the adolescent version is adapted. The author would also like to express her gratitude to Dr. Paula Ravitz and Ms. Regine King for their comments on early drafts of this manuscript.

References

Anderson, C.M., et al. 2006. Why Lower Income Mothers Do Not Engage with the Formal Mental Health Care System: Perceived Barriers to Care. *Qualitative Health Research* 16 (7): 926–43.

Angold, A., et al. 2000. Effectiveness of Non-residential Specialty Mental Health Services for Children and Adolescents in the "Real World." *Journal of the American Academy of Child and Adolescent Psychiatry* 39 (2): 154–60.

Armbruster, P., and T. Fallon. 1994. Clinical, Sociodemographic, and Systems Risk Factors for Attrition in a Children's Mental Health Clinic. *American Journal of Orthopsychiatry* 64 (4): 577–85.

Armbruster, P., and A. Kazdin. 1994. Attrition in Child Psychotherapy. In *Advances in Clinical Child Psychology*, edited by T.H. Ollendick and R. J. Prinz, 81–108. New York: Plenum Press.

Asarnow, J.R., et al. 2005. Effectiveness of a Quality Improvement Intervention for Adolescent Depression in Primary Care Clinics: A Randomized Controlled Trial. *Journal of the American Medical Association* 293 (3): 311–19.

Aseltine, R., and M. DeMartino. 2004. An Outcome Evaluation of the SOS Prevention Program. *American Journal of Public Health* 94 (3): 446–51.

Baer, J.S., and P.L. Peterson. 2002. Motivational Interviewing with Adolescents and Young Adults. In *Motivational Interviewing: Preparing People for Change*, edited by W.R. Miller and S. Rollnick, 320–32. Guilford Press: New York.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory.* Edgewood Cliffs: Prentice Hall.

Barrett, A.E., and H.R. White. 2002. Trajectories of Gender Role Orientation in Adolescence and Early Adulthood: A Prospective Study of the Mental Health Effects of Masculinity and Femininity. *Journal of Health and Social Behavior* 43 (4): 451–68.

Belle, D. 1990. Poverty and Women's Mental Health. *American Psychologist* 45 (3): 385–89.

Boldero, J., and B. Fallon. 1995. Adolescent Help-seeking: What Do They Get and from Whom? *Journal of Adolescence* 18 (2): 193–209.

Bordin, E.S. 1979. The Generalizability of the Psychoanalytic Concept of the Working Alliance. *Psychotherapy: Theory, Research & Practice* 16 (3): 252–60.

Caparana, G.V., et al. 1998. Impact of Adolescents' Perceived Self-regulatory Efficacy on Familial Communication and Antisocial Conduct. *European Psychologist* 3 (2): 125–32.

Casey, R.J., and J.S. Berman. 1985. The Outcome of Psychotherapy with Children. *Psychological Bulletin* 98 (2): 388–400.

Catron, T., and B. Weiss 1994. The Vanderbilt School-based Counseling Program: An Interagency, Primary-care Model of Mental Health Services. *Journal of Emotional and Behavioral Disorders. Special Series: Center for Mental Health Services Research Projects* 2 (4): 247–53.

Chandra, A., and C.S. Minkovitz. 2006. Stigma Starts Early: Gender Differences in Teen Willingness to Use Mental Health Services. *Journal of Adolescent Health* 38 (6): 754.e1–754.e8.

Children's Mental Health Ontario. 1999. *An Urgent Priority for Ontario: Summary and Recommendations*. Toronto: Children's Mental Health Ontario.

Chorpita, B.F., et al. 2002. Toward Large-scale Implementation of Empirically Supported Treatments for Children: A Review and Observations by the Hawaii Empirical Basis to Services Task Force. *Clinical Psychology: Science and Practice* 9 (2): 165–90.

Cicchetti, D., and F.A. Rogosch. 2002. A Developmental Psychopathology Perspective on Adolescence. *Journal of Consulting and Clinical Psychology* 70 (1): 6–20.

Cicchetti, D., and S.L. Toth 1996. *Adolescence Opportunities and Challenges*. Rochester: University of Rochester Press.

Coatsworth, J.D., et al. 2001. Brief Strategic Family Therapy versus Community Control: Engagement, Retention, and an Exploration of the Moderating Role of Adolescent Symptom Severity. *Family Process* 40 (3): 313–32.

Cohen, P., and C.S. Hesselbart. 1993. Demographic Factors in the Use of Children's Mental Health Services. *American Journal of Public Health* 83 (1): 49–52.

Copeland, V.C. 2006. Disparities in Mental Health Service Utilization among Low-income African American Adolescents: Closing the Gap by Enhancing Practitioner's Competence. *Child and Adolescent Social Work Journal* 23 (4): 407–31.

Cunningham, C.E., R.B. Bremner, and M. Boyle. 1995. Large Group Community-based Parenting Programs for Families of Preschoolers at Risk for Disruptive Disorders: Utilization, Cost effectiveness, and Outcome. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 36 (7): 1141–59.

Cunningham, C.E., R.B. Bremner, and M. Secord. 1998. *COPE: The Community Parent Education Program: A School-based Family Systems Oriented Workshop for Parents and Children with Disruptive Disorders*. Hamilton: COPE Works.

Daleiden, E., et al. 2006. Getting Better at Getting Them Better: Health Outcomes and Evidence-based Practice within a System of Care. *Journal of the American Academy of Child and Adolescent Psychiatry* 45 (6): 749–56.

Diguiseppe, R., J. Linscott, and R. Jilton. 1996. Developing the Therapeutic Alliance in Child-adolescent Psychotherapy. *Applied and Preventative Psychology* 5 (2): 85–100.

Dishion, T.J., and G.R. Patterson. 1992. Age Affects in Parent Training Outcomes. *Behavior Therapy* 23 (4): 719–29.

Edlund, M.J., et al. 2002. Dropping Out of Mental Health Treatment: Patterns and Predictors among Epidemiological Survey Respondents in the United States and Ontario. *American Journal of Psychiatry* 159 (5): 845–51.

Eyrich-Garg, K.M. 2008. Strategies for Engaging Adolescent Girls at an Emergency Shelter in a Therapeutic Relationship: Recommendations from the Girls Themselves. *Journal of Social Work Practice* 22 (3): 375–88.

Fergusson D.M., L.J. Horwood, and E.M. Ridder 2007. Conduct and Attentional Problems in Childhood and Adolescence and Later Substance Use, Abuse and Dependence: Results of a 25-year Longitudinal Study. *Drug and Alcohol Dependence* 88 (Suppl. 1): S14–S26.

Fergusson, D.M., and L.J. Woodward 2002. Mental Health, Educational and Social Role Outcomes of Adolescents with Depression. *Archives of General Psychiatry* 59: 225–31.

Flisher, A.J., et al. 1997. Correlates of Unmet Need for Mental Health Services by Children and Adolescents. *Psychological Medicine* 27 (5): 1145–54.

Garcia, J.A., and J.R. Weisz. 2002. When Youth Mental Health Care Stops: Therapeutic Relationship Problems and Other Reasons for Ending Youth Outpatient Treatment. *Journal of Consulting and Clinical Psychology* 70 (2): 439–43.

Gasquet, I., et al. 1997. Psychosocial Factors Associated with Help-seeking Behavior among Depressive Adolescents. *European Journal of Child and Adolescent Psychiatry* 6 (3): 151–59.

Goff, B.G., and H.W. Goddard. 1999. Terminal Core Values Associated with Adolescent Problem Behaviors. *Adolescence* 34 (133): 47–60.

Greeno, C., et al. 2002. Return to Treatment after Assessment in a Community Children's Mental Health Clinic. *Psychiatric Services* 52 (5): 624–26.

Griffin, J.A., D. Cicchetti, and P.J. Leaf. 1993. Characteristics of Youths Identified from a Psychiatric Case Register as First-time Users of Services. *Hospital and Community Psychiatry* 44 (1): 62–65.

Hanna, F.J., and W.P. Hunt. 1999. Techniques for Psychotherapy with Defiant, Aggressive Adolescents. *Psychotherapy: Theory, Research, Practice and Training* 36 (1): 56–68.

Harpaz-Rotem, I., D. Leslie, and R.A. Rosenheck. 2004. Treatment Retention among Children Entering a New Episode of Mental Health Care. *Psychiatric Services* 55 (9): 1022–28.

Hawke, J.M., J. Hennen, and P. Gallione. 2005. Correlates of Therapeutic Involvement among Adolescents in Residential Drug Treatment. *American Journal of Drug and Alcohol Abuse* 31 (1): 163–77.

Hawley, K.M. 2003. The Therapeutic Working Alliance in Child and Adolescent Psychotherapy. Dissertation Abstracts International: Section B, the Sciences and Engineering, 63(7-B), 3474. US: University Microfilms International.

Hawley, K.M., and A.F. Garland. 2008. Working Alliance in Adolescent Outpatient Therapy: Youth, Parent and Therapist Outcomes. *Child and Youth Forum* 37 (2): 59–74.

Health Canada. 2002. *Economic Burden of Illness in Canada, 1998.* (Catalogue No. H21-136/1998E) Ottawa: Health Canada.

Hoagwood, K.A. 2005. Family-based Services in Children's Mental Health: A Research Review and Synthesis. *Journal of Child Psychology and Psychiatry* 46 (7): 690–713.

Hodges, K, M. Wong, and M. Latessa 1998. Use of the Child and Adolescent Functional Assessment Scale (CAFAS) as an Outcome Measure in Clinical Settings. *The Journal of Behavioral Health Services and Research* 25 (3): 325–36.

Holmbeck, G.N., et al. 2005. Cognitive Behavioral Therapy with Adolescents: Guides from Developmental Psychology. In *Child and Adolescent Therapy: Cognitive Behavioral Procedures*, edited by P.C. Kendall, 419–63. New York: Guilford Press.

Kahle, L.R. 1983. *Social Values and Social Change: Adaption to Life in America.* New York: Praeger.

Kalafat, J., and M.J. Elias. 1995. Suicide Prevention in an Educational Context: Broad and Narrow Foci. In *Suicide Prevention: Toward the Year 2000*, edited by M.M. Silverman and R.W. Maris, 123–33. New York: Guilford Press.

Kazdin, A.E. 1993. Premature Termination from Treatment among Children Referred for Antisocial Behavior. *Journal of Clinical Child Psychology* 31 (3): 415–25.

———. 2000. Developing a Research Agenda for Child and Adolescent Psychotherapy. *Archives of General Psychiatry* 57 (9): 829–35.

Kazdin A.E., D. Bass, W.A. Ayers, and A. Rodgers. 1990. Empirical and Clinical Focus of Child and Adolescent Psychotherapy Research. *Journal of Consulting Clinical Psychology* 58 (6): 729–40.

Kazdin, A., L. Holland, and M. Crowley. 1997. Family Experience of Barriers to Treatment and Premature Termination from Child Therapy. *Journal of Consulting and Clinical Psychology* 65 (3): 453–63.

Kazdin, A., and J. Mazurick. 1994. Dropping out of Child Psychotherapy: Distinguishing Early and Late Dropouts over the Course of Treatment. *Journal of Consulting and Clinical Psychology* 62 (5): 1069–74.

Kazdin, A.E., and G. Wassell. 1999. Barriers to Treatment Participation and Therapeutic Change among Children Referred for Conduct Disorder. *Journal Clinical and Child Psychology* 28 (2): 160–72.

Kazdin A., and M.K. Whitley. 2003. Treatment of Parental Stress to Enhance Therapeutic Change among Children Referred for Aggressive and Antisocial Behavior. *Journal of Consulting and Clinical Psychology* 71 (3): 504–15.

Keely, M.L., and B.A. Weins. 2008. Family Influences on Treatment Refusal in School-linked Mental Health Services. *Journal of Family Studies* 17 (1): 109–26.

Kessler, R.C., P.A. Berglund, C.L. Foster, et al. 1997. Social Consequences of Psychiatric Disorders, II: Teenage Parenthood. *American Journal of Psychiatry* 154 (10): 1405–11.

Kessler, R.C., C.L. Foster, W.B. Saunders, et al. 1995. Social Consequences of Psychiatric Disorders, I: Educational Attainment. *American Journal of Psychiatry* 152 (7): 1026–32.

Koren, P.E., et al. 1997. Service Coordination in Children's Mental Health: An Empirical Study from the Caregiver's Perspective. *Journal of Emotional and Behavioral Disorders* 5 (3): 162–72.

Kovacs, M., and D. Goldston. 1991. Cognitive and Social Cognitive Development of Depressed Children and Adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* 30 (3): 388–92.

Langlois, S., and P. Morrison. 2002. Suicide Deaths and Suicide Attempts. *Health Reports*, vol. 13, no. 2. Ottawa: Statistics Canada.

Lewinsohn, P.M., P. Rohde, D.N. Klein, and J.R. Seeley. 1999. Natural Course of Adolescent Major Depressive Disorder. I. Continuity into Young Adulthood. *Journal of the American Academy of Child and Adolescent Psychiatry* 38 (1): 56–63.

Lewit, E.M., D.L. Terman, and R.E. Behrman. 1997. Children and Poverty: Analysis and Recommendations. *The Future of Children* 7 (2): 4–24.

Links, P.S., Y. Bergmans, and M. Cook. 2003. Psychotherapeutic Interventions to Prevent Repeated Suicidal Behavior. *Brief Treatment and Crisis Intervention* 3 (4): 445–64.

Loar, L. 2001. Eliciting Cooperation from Teenagers and Their Parents. *Journal of Systemic Therapies* 20 (1): 59–77.

Luborsky, L. 1976. Helping Alliance in Psychotherapy. In *Successful Psychotherapy*, edited by J.L. Cleghorn, 92–116. New York: Bruner/Mazel.

MacNaughton, K.L., and J.R. Rodrigue. 2001. Predicting Adherence to Recommendations by Parents of Clinic Referred Children. *Journal of Consulting and Clinical Psychology* 69 (2): 262–70.

McCabe, K.M. 2002. Factors that Predict Premature Termination among Mexican-American Children in Outpatient Psychotherapy. *Journal of Child and Family Studies* 11 (3): 347–59.

McKay, M.M., and W.M. Bannon. 2004. Engaging Families in Child Mental Health Services. *Child and Adolescent Psychiatric Clinics of North America* 13 (4): 905–21.

McKay, M., K. McCadam, and J. Gonzales. 1996. Addressing Barriers to Mental Health Services for Inner City Children and Their Caretakers. *Community Mental Health Journal* 32 (4): 353–61.

McKay, M.M., J. Pennington, C.J. Lynn, et al. 2001. Understanding Urban Child Mental Health Service Use: Two Studies of Child, Family, and Environmental Correlates. *Journal of Behavioral Health Services and Research* 28 (4): 1–10.

Miller, W.R., and S. Rollnick. 2002. *Motivational Interviewing: Preparing People for Change*, 2nd ed. New York: Guilford Press.

Moli, P., J.A. Berriman, and J. Evans. 2008. Would I Attend Child and Mental Health Services (CAMHS)? Fourteen to Sixteen Year Olds Decide. *Child and Adolescent Mental Health* 13 (1): 19–25.

Morris, R., and J. Nicholson. 1998. The Therapeutic Relationship in Child and Adolescent Psychotherapy: Research Issues and Trends. In *Handbook of Children and Adolescents*, edited by T.R. Kratochwill and R. Morris, 405–25. Needham Heights, Mass.: Allyn and Bacon, Inc.

Morrissey-Kane, E., and R. Prinz. 1999. Engagement in Child and Adolescent Treatment: The Role of Parental Cognitions and Attributions. *Clinical Child and Family Psychology Review* 2 (3): 183–98.

Mufson, L., et al. 2004. The Effectiveness of Interpersonal Psychotherapy for Depressed Adolescents. *Archives of General Psychiatry* 61 (6): 577–84.

Nock, M.K., and C. Ferriter. 2005. Parent Management of Attendance and Adherence in Child and Adolescent Therapy: A Conceptual and Empirical Review. *Clinical Child and Family Psychology* Review 8 (2): 149–66.

Oetzel, K., and D.G. Scherer. 2003. Therapeutic Engagement with Adolescents in Psychotherapy. *Psychotherapy: Theory, Research, Practice and Training* 40 (3): 215–25.

O'Malley, F. 1990. Developing a Therapeutic Alliance in the Hospital Treatment of Disturbed Adolescents. *Bulletin of the Menninger Clinic* 54 (1): 13–24.

Ontario Ministry of Children and Youth Services. 2006. *A Shared Responsibility: Ontario's Policy Framework for Child and Youth Mental Health*. Toronto: Queen's Printer for Ontario.

Os, Jv., et al. 1997. Developmental Precursors of Affective Illness in a General Population Birth Cohort. *Archives of General Psychiatry* 54 (7): 625–31.

Owens, P.L., et al. 2002. Barriers to Children's Mental Health Services. *Journal of the American Academy of Child and Adolescent Psychiatry* 41 (6): 731–38.

Padgett, D.K., et al. 1993. The Effect of Insurance Benefit Changes on Use of Child and Adolescent Outpatient Mental Health Services. *Medical Care* 31 (1): 96–110.

Prinstein, M., et al. 2005. Adolescent Girls' Interpersonal Vulnerability to Depressive Symptoms: A Longitudinal Examination of Reassurance-seeking and Peer Relationships. *Journal of Abnormal Psychology* 114 (4): 676–88.

Prochaska, J.O., and C.C. DiClemente. 1984. *The Transtheoretical Approach: Crossing the Traditional Boundaries of Therapy*. Malabar, Fla.: Krieger.

Romney, D.M. 1988. A Retrospective Study of Dropout Rates from a Community Mental Health Centre and Associated Factors. *Canada's Mental Health* 36 (1): 2–4.

Rubenstein, A.K. 1996. Interventions for Scattered Generation: Treating Adolescents in the Nineties. *Psychotherapy: Theory, Research, Practice and Training* 33 (3): 353–60.

Rudolph, K.D., S.F. Lambert, A.G. Clark, and K.D. Kurlakowsky. 2001. Negotiating the Transition to Middle School: The Role of Self-regulatory Processes. *Child Development* 72 (3): 929–46.

Rutter, M. 1990. Psychosocial Resilience and Protective Mechanisms. In *Risk and Protective Factors in Development and Psychopathology*, edited by J. Rolf, A.S. Mastersen, D. Chichetti, and S. Winetraub, 181–214. New York: Cambridge University Press.

Rutter, M., E. Taylor, and L. Hersov. 1994. *Child and Adolescent Psychiatry: Modern Approaches*. Oxford: Blackwell Scientific Publications.

Santor, D.A., and V. Kusumakar. 2001. Open Trial of Interpersonal Therapy in Adolescents with Moderate to Severe Major Depression: Effectiveness of Novice IPT Therapists. *Journal of the American Academy of Child and Adolescent Psychiatry* 40 (2): 236–40.

Scannapieco, M., and K. Connell-Carrick. 2005. *Understanding Child Maltreatment: An Ecological and Developmental Perspective*. New York: Oxford University Press.

Schonert-Reichl, K.A., and J.R. Mueller. 1996. Correlates of Help Seeking in Adolescence. *Journal of Youth and Adolescence* 25 (6): 705–31.

Sheffield, J.K., E. Fiorenza, and K. Sofronoff. 2004. Adolescents' Willingness to Seek Psychological Help: Promoting and Preventing Factors. *Journal of Youth and Adolescents* 33 (6): 495–507.

Simonoff, E., et al. 2004. Predictors of Antisocial Personality: Continuities from Childhood to Adult Life. *British Journal of Psychiatry* 184 (2): 118–27.

Stein, B.D., et al. 2003. A Mental Health Intervention for Schoolchildren Exposed to Violence: A Randomized Controlled Trial. *Journal of the American Medical Association* 290 (5): 603–11.

Steinberg, L. 2005. *Adolescence*, 7th ed. Boston: McGraw-Hill.

Stevens, J., Kelleher, K., Ward-Estes, J., et al. 2006. Perceived Barriers to Treatment and Psychotherapy Attendance in Child Community Mental Health Centers. *Community Mental Health Journa*l 42 (5): 449–58.

Strupp, H.H. 1993. The Vanderbilt Psychotherapy Studies: Synopsis. *Journal of Consulting and Clinical Psychology* 61 (3): 431–33.

Sue, D.W., and D. Sue 2003. *Counseling the Culturally Different: Theory and Practice*, 3rd ed. Hoboken: John Wiley and Sons Inc.

Swartz, H.A., et al. 2007. Engaging Depressed Patients in Psychotherapy: Integrating Techniques from Motivational Interviewing and Ethnographic Interviewing to Improve Treatment Participation. *Professional Psychology: Research and Practice* 38 (4): 430–39.

Szapocznik, J., et al. 1988. Engaging Adolescent Abusers and Their Families into Treatment: A Strategic Structural Systems Approach. *Journal of Consulting and Clinical Psychology* 56 (4): 552–57.

Tuma, J.M. 1989. Mental Health Services for Children. *American Psychologist* 44 (2): 188–99.

Tynan, W.D., W. Schuman, and N. Lampert. 1999. Concurrent Parent and Child Therapy Groups for Externalising Disorders: From the Laboratory to the World of Managed Care. *Cognitive and Behavioral Practice* 6 (1): 3–9.

Verhulst, F.C., and J. Van Der Ende. 1997. Factors Associated with Child Mental Health Service Use in the Community. *Journal of the American Academy of Child and Adolescent Psychiatry* 36 (7): 901–9.

Waddell, C., et al. 2002. Child Psychiatric Epidemiology and Canadian Public Policy-making: The State of the Science and the Art of the Possible. *Canadian Journal of Psychiatry* 47 (9): 825–32.

———. 2005. A Public Strategy to Improve the Mental Health of Canadian Children. *Canadian Journal of Psychiatry* 50 (4): 226–33.

Wang, P.S., et al. 2005. Twelve-month Use of Mental Health Services in the United States: Results from the National Comorbidity Survey Replication. *Archives of General Psychiatry* 62 (6): 629–40.

Weersing, V.R., and J. Weisz. 2002. Community Clinic Treatment of Depressed Youth: Benchmarking Usual Care against CBT Trials. *Journal of Consulting and Clinical Psychology* 70 (2): 3–29.

Weisz, J.R., B. Weiss, S.S. Han, D.A. Granger, and T. Morton. 1995. Effects of Psychotherapy with Children and Adolescents Revisited: A Meta-analysis of Treatment Outcome Studies. *Psychological Bulletin* 117 (3): 450–68.

Wolkind, S., and M. Rutter. 1985. Separation, Loss and Family Relationships. In *Child and Adolescent Psychiatry: Modern Approaches*, edited by M. Rutter and L. Hersov, 34–57. Oxford: Blackwell Scientific.

Worrall-Davies, A., and F. Marino-Francis. 2008. Eliciting Children's and Young People's View of Child and Adolescent Mental Health Services: A Systemic Review of Best Practice. *Child and Adolescent Mental Health* 13 (1): 9–15.

Wu, P., et al. 2001. Depressive and Disruptive Disorders and Mental Health Service Utilization in Children and Adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* 38 (9): 1081–90.

Zuckoff, A., et al. 2004. Engagement Session: Treatment Manual. Unpublished Manuscript.

Zuroff, D., et al. 2007. Autonomous Motivation for Therapy: A New Common Factor in Brief Treatments for Depression. *Psychotherapy Research* 17 (2): 137–47.