

A Routledge FREEBOOK

The Economics of Health and Health Care

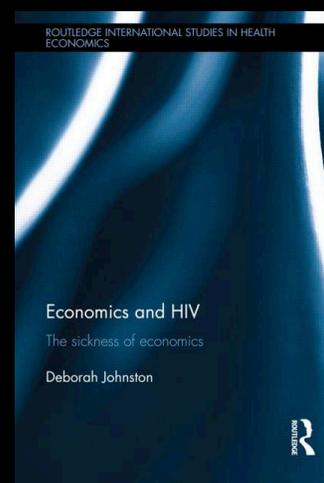
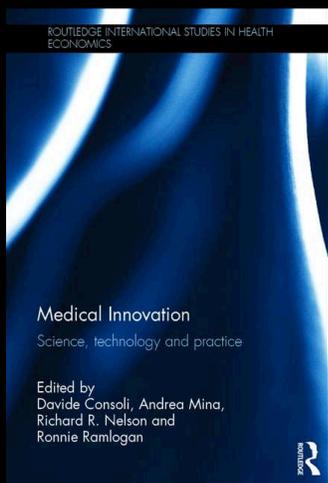
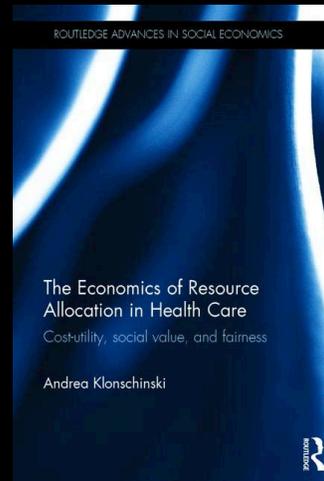
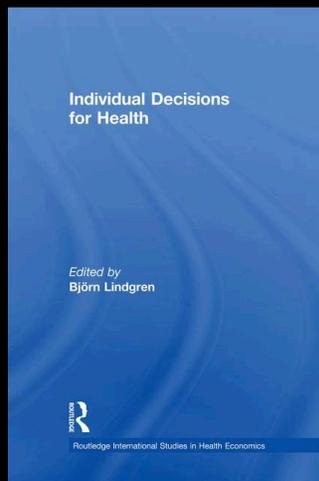
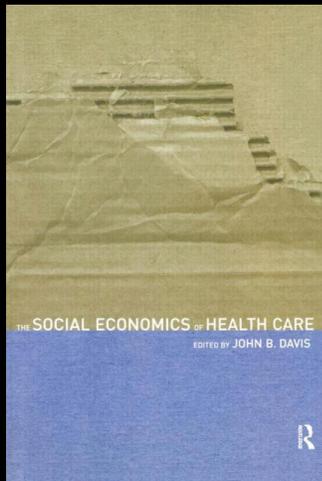
A Research Compilation



TABLE OF CONTENTS

-  Introduction
-  1 • Accounting for Fairness and Efficiency in Health Economics
-  2 • Health, Genuine Uncertainty, and Information
-  3 • On the Normative Status of Empirically Elicited Prioritization Preferences
-  4 • Coronary Artery Disease
-  5 • Kenya's Health Innovation Capacity
-  6 • A Political Economy Perspective on the Economic Impact of AIDS

READ THE LATEST ON THE ECONOMICS OF HEALTH WITH THESE KEY TITLES



VISIT WWW.ROUTLEDGE.COM/ECONOMICS
TO BROWSE THE FULL RANGE OF OUR ECONOMICS TITLES

REQUEST AN INSPECTION COPY OF *THE ECONOMICS OF HEALTH AND
HEALTH CARE* HERE



Introduction

We put together this FreeBook to complement the publication of the 8th edition of *The Economics of Health and Health Care*, the market-leading textbook covering all aspects of health economics.

The six chapters included here focus on different elements of the field of health and health care economics, and range from the theoretical to the more practical. Our intention in selecting these chapters was to provide you with an overview of the work that's been done in this subject area over the past decade, and highlight the breadth and depth of this field of study.

Our hope is that you will use this FreeBook in your classroom alongside *The Economics of Health and Health Care*. The variety of content included here means that there is something for everyone, whether you are looking for case studies or additional research that your students can use as a jumping-off point for their own study.

These six chapters have been reproduced in their entirety, and all formatting and in-text references have been preserved from the original work. This means that you may notice some inconsistencies in things like citation format from chapter to chapter, or that you may come across a reference to previous content in the original work (and not this FreeBook). We don't anticipate that these details will have any impact on the value of this collection or impede its use as a learning or reference tool.

To learn more about any of the books featured in this FreeBook, please feel free to use any of the links on page 3 of this book, or go directly to the chapter page and navigate to our website from there. To request an inspection copy of *The Economics of Health and Health Care* and consider the textbook for use in your course, [click here](#).

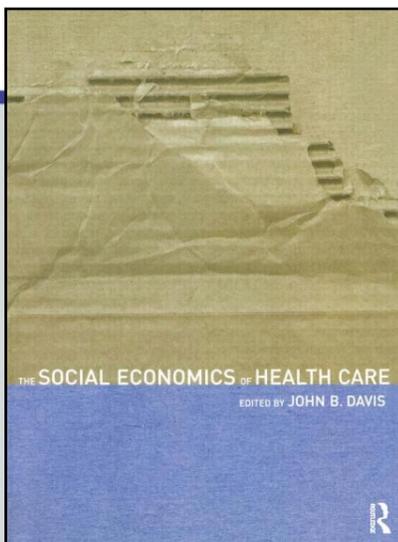


CHAPTER

1

ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

BY JOSHUA COHEN AND PETER UBEL



This chapter is excerpted from
The Social Economics of Health Care
Edited by John B. Davis

© 2001 Taylor & Francis Group. All rights reserved.



[Learn more](#)



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

Introduction

The use of economic reasoning to analyze health care is a comparatively recent development. However, the seeds for this development were sown long ago. Highly influential neoclassical economists like Paul Samuelson, emulating the success of the natural sciences in employing calculus, derived behavioral rules from mathematically tractable first-order principles. Postwar neoclassicals wanted to reduce the ambiguity of certain economic concepts and remove value ladenness from economics. It was thought that by increasing the hardness of economics, they could do both. Formulating economic theory mathematically undoubtedly increased hardness, which in turn made certain concepts less ambiguous. Theoretical claims that did not lend themselves to mathematical translation were even cast aside by some neoclassicals: '[A]ny sector of economic theory which cannot be cast into the mold of such a [mathematical] system [of equations] must be regarded with suspicion as suffering from haziness' (Samuelson 1947: 9). Samuelson admitted that the degree of hardness in economics could only be considered 'intermediate,' but that 'when one descends [in degrees of hardness] lower... say to certain areas of sociology... [they] are almost completely without substantive content' (1947: ix).

Post-war neoclassicals were often critical of the methods being used in the 'softer' social sciences. Neoclassicals attempted to impose more rigorous economic methods on these other social sciences, a practice that became known as *economics imperialism*. The neoclassical view that economic theory studies the allocation of scarce means with alternative uses (Robbins 1932) did not preclude the study of scarce noneconomic means with alternative noneconomic uses. Perhaps the most successful example of economics imperialism is found in Gary Becker's work. In the 1960s and 1970s, Becker revolutionized the study of sociological phenomena such as family, race, and class, using a neoclassical apparatus of analytical tools (Becker 1976).

Health economics is a relative newcomer to the economics profession. It became a subdiscipline (with its own peer-reviewed journals and professional associations) in the 1980s. Health economics fits neatly under the rubric of economics imperialism as health economists generally analyze health care using a neoclassical model of behavior. In the process of subjecting health care to economic analysis, health terms are translated into economics jargon; health becomes a product, patients consumers, doctors health care providers, the hospital



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

a product delivery system, and care managed.

As the field of economics has become increasingly quantitative, it has become less likely for health economists to focus attention on ethical values such as *fairness* that are presumably less quantifiable. This said, the need to emphasize the ethical dimension associated with economic decisions is perhaps more evident in health care than in other sectors of the economy. Certainly, the ethical issues raised by possible tradeoffs between fairness and efficiency in health care illustrate the risks of relying too much on quantitative economic approaches, such as Paretian welfare analysis, that ignore ethical considerations.

Health economists may see themselves solely as engineers, offering technical solutions to technical problems. However, they are at least indirectly concerned with ethical issues. When a health economist is asked to evaluate the *kind* of health output to be produced—longer life expectancy, increased quality of life, fewer sick days—answers to this question presuppose a certain ethical view about what is best for society. Furthermore, when health economists deal with questions of distribution such as 'To whom should Medicaid funds be allocated?' they are concerned with fairness. Even if supposedly value-neutral Pareto distribution rules are used in the name of improved efficiency, a certain fairness norm is implied. The Pareto definition of efficiency *implies definite value judgments* about fairness.

This chapter describes how ethics and economics interact in the area of health care rationing. This section has explored the advantages and limitations of using Paretian welfare economic analysis to evaluate the fairness-efficiency implications of health care rationing. The second section explains how the possibility of a fairness-efficiency tradeoff needs to be assessed both within the dollars domain and between the dollars and rights domains. The third section evaluates a practical case of explicit health care rationing, the Oregon health care initiative. This initiative is assessed in terms of how it has promoted fairness and efficiency.

Conventional wisdom on the fairness-efficiency tradeoff

'The market needs a place, and the market needs to be kept in place' (Okun 1975).

Rights and dollars domains Okun (1975) distinguishes *rights* and *dollars* domains in society. The chief distinguishing characteristics of the rights and dollars domains, according to Okun, are *equality*¹ and *efficiency*, respectively. In the rights domain, political and judicial institutions provide universally distributed rights and



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

privileges that proclaim the equality of all citizens. However, in the dollars domain, society's economic institutions rely on market-determined incomes that produce substantial disparities among citizens in material well-being. Hence, some degree of inequality, unequal spread of wealth, is guaranteed by virtue of the competitive aspects of the market process. There are winners and losers in the marketplace where motives to maximize profits and consumer utility reign. Under an ideal set of assumptions, utility- and profit-maximizing behavior lead to the most efficient outcomes.

It seems that the two domains can be further distinguished by the *inexchangeability of rights* for other rights, as opposed to the *exchangeability of dollars* for commodities with dollar value and vice versa. Rights may not be bought and sold, whereas dollars can (either for commodities with dollar value or for other convertible currencies). For example, every citizen has a right to vote, and each citizen's vote is counted equally. No one has to pay for this right, nor is the wealthier citizen's vote weighed more than the poor citizen's. Furthermore, no one is supposed to sell their vote.²

Policymakers can intervene in both domains. In the rights domain, policymakers help to establish and fortify citizens' rights. In the dollars domain, they can mitigate for the possibly negative side effects of an unregulated market distribution of dollars.³ The unregulated market evidently promotes economic growth, but does this unevenly across the population.

Policymakers appear to be faced with two types of *tradeoffs*. On the one hand, within the dollars domain, efficiency can be traded off for a more equal distribution of dollars, or vice versa. On the other hand, efficiency can be traded off for an enlargement of the scope of rights, or vice versa.

Consider how this general discussion of the rights and dollars domains applies to health care as it is organized in the United States. In the United States, health care occupies both domains. On the dollars side, health care is a major industry accounting for more than 15 percent of the nation's Gross Domestic Product. Health care has all the attributes that we associate with big business—corporations, stockholders, mergers, bankruptcies, etc. However, a certain circumscribed area of health care is protected from the dollars domain. This area occupies the rights domain, where, for instance, a number of *constitutionally* grounded health care rights are *universally* distributed to citizens such as the right to refuse treatment, and the right to emergency treatment.⁴ Additionally, a number



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

of *legal* rights to health care are specifically targeted at certain subpopulations such as Veterans, the elderly, and the very poor. Programs such as Medicare and Medicaid establish inalienable legal rights to the provision of health care services to those who qualify for them. These programs serve in part as a buffer against the vagaries of the market, making sure that inability to pay for health care services does not prevent eligible citizens from access to health care.

Equality and efficiency

Equality as a proxy for fairness appears to be the guiding principle behind decisions made in the rights domain, whereas efficiency (mediated by utility and profit-maximizing behavior) appears to guide the decision making process in the dollars domain. Equality in this context implies that it is our duty to treat each other *as if* we are equal, disregarding certain characteristics that make us different such as race, gender, and socioeconomic status. This idea hinges on the Aristotelian principle that 'like cases be treated like.' The principle 'treat like cases like' seems to underlie many, if not all, notions of fairness. But, it begs the question 'Which characteristics do we consider morally relevant likenesses?' The various notions of fairness differ not in whether like cases should be treated like, but instead with respect to *what* are considered morally relevant likenesses. For example, when we have life saving treatments that could either save the life of someone who could be returned to perfect health, or the life of someone who will have a disability, are both types of patients alike because they are human beings, and therefore deserve life saving therapy? Or, are they unlike each other in that one will have a disability and the other will not (see Nord et al. 1999)?

It would be presumptive to suggest definitive answers to the latter questions. However, we can say that unless everyone is equal with respect to the likenesses that we may consider to be morally relevant, which of course is not the case, providing everyone with the same amount of health care would not make much sense.

Efficiency in this context comprises a *technical* element: Pareto-efficient and production and consumption; and a *social* element: the Pareto principle. Technically, a number of marginal conditions must apply. The first condition is that the marginal rates of substitution between any pair of distributable commodities, resources, or outcomes (for example, health care outcomes) must be the same for all individuals who consume both commodities or are beneficiaries of both



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

resources or outcomes. The second condition is that the quantity of each commodity, resource, or outcome produced conforms with consumer (or beneficiary) preferences. That is, the marginal rate of transformation in production must equal the marginal rate of substitution in consumption for every pair of commodities, resources, or benefits. If these conditions do not hold, a shift in the patterns of production and/or consumption is possible which would benefit some without injuring others. These technical conditions imply the social element associated with Pareto efficiency, namely, the widely cited Pareto principle which says that a change (for instance, a health care policy change) is desirable if it makes some individual(s) better off without making any others worse off.

It should be noted here that the Paretian form of distributive justice merely accounts for a *preference-satisfaction* view of human well-being. It is far from clear whether such a view suffices as an adequate account of wellbeing—especially in the context of health. The preference-satisfaction view would appear in certain instances to assume too tight a connection between choices and preferences and, in turn, too tight a connection between preference-satisfaction and well-being. Visible acts of choice are supposed to accurately reflect preferences. But, in fact, important choices in health care, choices made by patients and health care providers alike, may be more reflective of certain moral constraints than they are of preferences. As for preference-satisfaction, it may fail to improve well-being if the preferences being satisfied are irrational, poorly cultivated, or simply based on incomplete or false information.

Equality-efficiency tradeoff

As noted above, equality (as a proxy for fairness) and efficiency are presumed to relate inversely. The inverse relationship between equality and efficiency can be depicted using the indifference curve tool. In Figure 4.1 below, following Rawls (1971: 37-39), the indifference curves (I and I') represent possible combinations of equality and efficiency that are considered by society to be equally just.⁵ Society can attach weights to equality and efficiency given its respective notions of distributive justice, whether these are egalitarian, Rawlsian, utilitarian, or libertarian. The slope of an indifference curve at any particular point expresses the relative weight allotted equality and efficiency. A more vertical curve (I) gives relatively more weight to equality, while a more horizontal curve (I'), allots more weight to efficiency. The more important and/or the more poorly satisfied a value



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

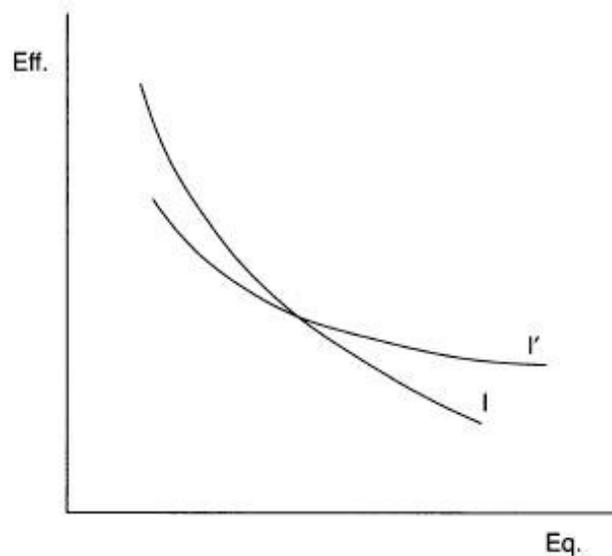


Figure 4.1 Two hypothetical indifference curves representing possible combinations of equality and efficiency

such as equality or efficiency is, the greater the increase in another value required for compensating a loss. In this context, equality can be broadly measured by evaluating the variance in, for instance, wealth or health outcomes across the population. Efficiency can be broadly measured by assessing the total production of wealth or health outcomes divided by the inputs needed to produce them.

Now imagine a three-person world, one impartial distributor and two beneficiaries, X and Y. The impartial distributor distributes health care resources necessary⁶ to produce certain health outcomes. These outcomes can be measured in terms of decreases in morbidity and mortality—life expectancy discounted for the quality of life expected, or quality-adjusted life-years (QALYs). Suppose that there is a line, MN, that represents the locus of possible Pareto-efficient points (see Figure 4.2 below). At each point on the MN curve a Pareto-efficient allocation of health outcomes (for example, QALYs) exists. A Pareto-efficient allocation of health outcomes is one in which no individual can be made healthier—for instance, in terms of QALYs—without making some other individual less healthy. In other words, there is no redistribution possible that makes either individual healthier without making the other less healthy.



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

The Pareto distribution rule appears to ignore the relative (in)justice of the status quo. From a Paretian perspective, for instance, a distributive policy that increases health outcomes of a few already healthy individuals, even if this increase is only marginal, while leaving many unhealthy individuals just as unhealthy, would be considered a Pareto improvement. A state can be Pareto optimal with some people in extreme misery and others rolling in luxury, so long as the miserable cannot be made better off without cutting into the luxury of the rich' (Sen 1987: 32).

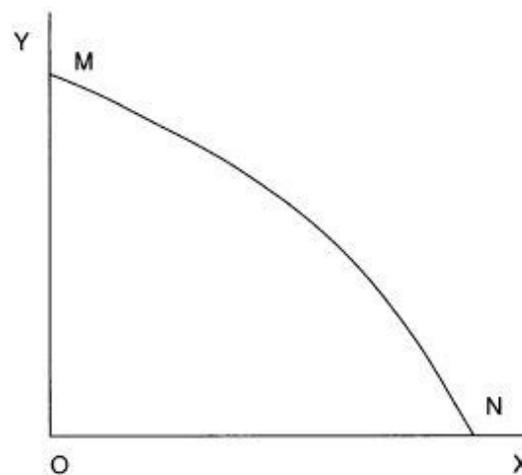


Figure 4.2 Pareto-efficient locus

From the Paretian perspective, the locus of points, MN in Figure 4.2, is *not ranked in terms of equality or fairness, but only according to efficiency*. Giving M to Y is considered *just as efficient* as giving N to X, but we do not know whether it is fairer to give N to X, or M to Y. The only thing we do know is that if we give N to X, and nothing to Y, Y's well-being would stay the same, while X's would improve—a Pareto improvement. Needless to say, it could be that Y has an intense need for M, while X's need for N is only slight. In this case, Y stands more to gain from getting M than X stands to lose from not getting N. But, since the Paretian viewpoint precludes the possibility of interpersonal comparisons of need intensity, we cannot compare X and Y's needs. This is an enormous limitation. In fact, the only thing we can definitively say is that all points below MN are inferior distributions in terms of Pareto efficiency.

If we add the 45-degree line representing a simple equal division of health



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

resources, and we assume simplistically that resources and outcomes are proportionate, we *can* rank points on MN in terms of how much or little equality they denote (see Figure 4.3 below). For example, compare point C to point D. If we place weight on an equal division of health outcomes, D would be considered a 'better' distribution than C, despite the fact that they are both *technically* just as Pareto-efficient.⁷

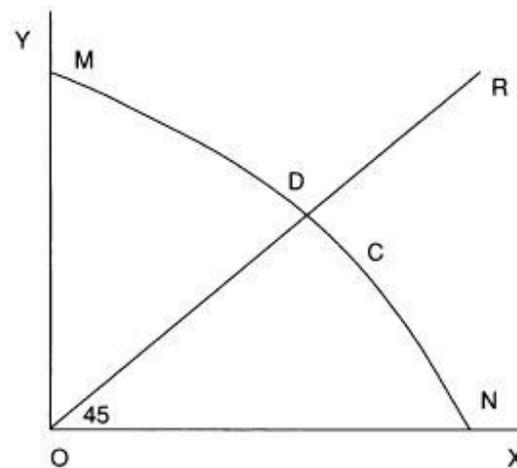


Figure 4.3 Adding 45-degree line

Tradeoffs inevitable?

Notice how in the case of the C-D comparison in Figure 4.3, the presumption that there is *necessarily* an equality-efficiency tradeoff is demonstrated to be false. Efficiency and equality *do not always have to be balanced against each other*. Moving from C to D improves equality while we remain on the same Pareto-efficient locus.⁸

There are in fact a number of other rules of distribution that do not have tradeoff implications—are not zero-sum games. Positive- as well as negative-sum games are possible between equality and efficiency. One such rule—*resource equalization*—is described in Culyer (1990). On the basis of a hypothetical case of global-budget health care rationing, Culyer illustrates the *tradeoff* implications of



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

three different distributive rules. Culyer shows the effects the different rules have on the degree of efficiency, measured in terms of population health outcomes per resource unit expended, and the degree of equality, measured in terms of variance in population health outcomes. The first rule Culyer considers, the *radically egalitarian* distribution rule, equalizes health outcomes across the population. Given differing individual health care needs, such a distributive rule implies allocating the medically neediest a disproportionate share of resources. It may even imply allocating all resources to the medically neediest depending on the initial dispersion of health care needs across the population. A completely even spread of health outcomes raises the health outcome levels for some (the initially sickest) while possibly lowering the outcomes for others (the initially healthiest) as this rule implies actively lowering the latter's health to the lowest common denominator. The population's overall health outcome level may drop as the initially healthiest may stand more to lose than the neediest stand to gain. Culyer states, therefore, that radical equality comes at a price: an inefficient allocation of resources.

Culyer next considers the *utilitarian* distribution rule which aims to maximize aggregate health outcomes. This rule is maximally efficient, given Culyer's definition of efficiency, as the outcomes per unit of input are maximized. However, the gain in efficiency is offset by a loss of equality; the variance in health outcomes across the population increases, leaving the initially sickest even sicker, while the initially healthiest either remain healthy or get healthier. Culyer states therefore that utilitarian efficiency comes at a price; an unequal allocation of both resources and outcomes.

The third and most interesting case Culyer considers is a compromise between the utilitarian and egalitarian rules. He calls this compromise *resource equalization*. This egalitarian rule is less radically egalitarian than outcomes equalization. It is inspired by Dworkin's equality-of-resource-endowment rule (see Dworkin 1985). The point of this rule is to give everyone an equal 'minimally adequate' share of scarce health care resources. People can do what they want with the resources at their disposal, in line with their health-related preferences. As a result, societal health care resource utilization will reflect differing preference and initial health endowment patterns among people. If people start out with equal resources, a health care market allocation will result in some people benefitting more, and others less from their initial resource endowment, depending on their initial health status and their preferences. Access to health care *resources* is spread evenly across



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

the population. But health outcomes are unevenly distributed, both because individuals' needs and preferences are different to begin with, and because each individual's capacity to benefit from resources is different. However, the unevenness of distribution is less than it would have been in the utilitarian case. Additionally, resource equalization is more efficient than a radically egalitarian distribution. This compromise case of resource equalization does not necessarily imply a tradeoff between equality and efficiency. Compared to initial degrees of equality and efficiency, resource equalization may improve both.⁹

The Oregon initiative

The first and second sections above describe in abstract terms the relationship between fairness and efficiency, against the backdrop of Paretian welfare analysis. These two sections also theoretically describe how the fairness-efficiency tradeoff plays itself out both within the dollars domain and between the dollars and rights domains. In this section we evaluate a practical case of explicit health care rationing, the Oregon health care initiative. This initiative is assessed in terms of how it has promoted fairness and efficiency both within the dollars domain and across the dollars and rights domains. The inevitability of a fairness-efficiency tradeoff is called into question.

In 1983, the President's Commission for the study of ethical problems in medicine and biomedical and behavioral research asserted health care's special status among 'goods' that 'society has a role in distributing as equitably as possible.' Evidently, there is something special about health care. In a way different from other kinds of services, health care services allow us to pursue important life goals that could not be pursued without them (Daniels 1985). Because of health care's special place in society, it is often argued that health care should not be distributed solely according to either ability, or willingness, to pay. Given health care's unique status, the Commission suggested that 'all Americans should have adequate access to health care resources,' access being defined by the Commission as the ability each citizen has to secure necessary health care without being barred for reasons of social status, income, ability to pay, place of residence, or other factors extraneous to the appropriate delivery of health care services.¹⁰ For all intents and purposes, however, the federal government has since abandoned fairness as a goal of health care policy (Caplan 1997: 148). If there is a common purpose to government health care policy today, it appears to be *cost containment*



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

which, if viewed charitably, is supposed to be synonymous with efficiency. The growth in health expenditures has been contained over the last 8 years following increased enrollment in managed care. Managed care has managed to control costs somewhat by way of *capitation*, or prospective payments of lump sums to health care providers per patient, rather than retrospective payment per medical intervention as under fee-for-service.

The exponential rise in health care costs over the last three decades, from 7 percent of GDP in 1968 to over 15 percent of GDP in 1998, raises the specter of a health care system out of cost control. However, this cost problem should not mask the problems related to lack of fairness in the health care system. Despite more money being pumped into health care during the last three decades, the numbers of uninsured (more than 44 million in 1998, or 17 percent of the American population) and underinsured have risen dramatically while population-wide health-outcome measures have not improved significantly during this time period.¹¹

Rising costs related to the Medicaid program, together with federal Medicaid budget cuts in the 1980s, led many state governments to arbitrarily limit Medicaid eligibility by lowering the income threshold level by as much as 35 percent. On the face of it, this seems to be grossly arbitrary. Limiting eligibility on income grounds rather than medical necessity does not even seem rational, given that one of the original goals of the Medicaid program was to protect the poor from being in financial dire straits due to the costs of health care. In fact, this way of rationing seems to institutionalize rationing according to ability to pay rather than medical necessity.

In 1989, John Kitzhaber, a physician, community leader, and then state senate chairman in Oregon,¹² launched an initiative designed to reduce the arbitrariness associated with Medicaid rationing. Kitzhaber was fundamentally opposed to rationing according to ability to pay. Instead, he wanted to construct a 'rational,' publicly accountable rationing system. Kitzhaber gathered the support of many local politicians, community activists, physicians, patient representatives, and state Medicaid officials to push the initiative through the state legislature. The initiative's main purpose was to improve the poor's access to Medicaid by adding as many uninsured people to the Medicaid rolls as possible. All residents with incomes below the federal poverty level would become eligible for Medicaid, not just the very poor. As a result, the option of arbitrarily rationing people out of the



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

health care system no longer existed. Rather than rationing by excluding people, the initiative proposed rationing by way of exclusion of health care services.

The Oregon initiative constitutes an effort to develop an integrated social policy approach, not strictly a health care policy approach. Dollars spent on health care imply an opportunity cost in terms of forgone alternative spending outlets. Resource allocations for health care are balanced with allocations in related areas which also affect health, such as education. The Oregon initiative proposes a system of resource allocation that recognizes scarcity and opportunity cost, and hence the need to allocate funds and resources among many competing social programs including health care, education and infrastructure.

It may be helpful to consider the Oregon initiative as analogous to Culyer's resource equalization distribution rule. This is because, with the Oregon proposal, access to resources is equalized according to medical need across a clearly defined subpopulation (all residents with an income below the federal poverty level).

For the purpose of rationally excluding a number of health care services, the Medicaid benefit package had to be overhauled and reduced. The method Oregon chose to accomplish this was the *prioritized list*. In the prioritization process, two key questions need to be answered. First, given the Medicaid budget allocated Oregon, the number of people covered has to be established. Second, the health care services for which people are covered has to be determined. So-called *condition treatment pairs* are the building blocks of the prioritized list. Each medical condition (for example, appendicitis) is linked specifically with a treatment (appendectomy).¹³ Rather than eliminating types of services (such as prescription drugs), Oregon's prioritized list would eliminate specific relatively ineffective treatments for specific conditions. A prioritized list would guarantee that health care benefit reductions would eliminate only the least effective treatments. In 1991, Oregon ranked more than 700 diagnoses and treatments in order of importance. The Oregon Health Services Commission decided to draw a line below which everything would be covered, and above which nothing would be covered.¹⁴ Treatments that prevent death and lead to full recovery ranked first in priority. Treatments that prevent death without full recovery are ranked next. Treatments that result in minimal or no improvement in the quality of life ranked last.

The Oregon initiative targets both the rights and dollars domains. In the rights domain, it aims at establishing for all Oregonians a legal right to a basic, though comprehensive, health care package.¹⁵ The Oregon plan also intervenes in the



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

dollars domain by enforcing community rating of health insurance premiums, thus limiting the ability of health insurance companies to risk-rate premiums. It is in this sense comparable to the resource equalization rule. Those below the federal poverty level are guaranteed access to the Oregon Medicaid program, those above are guaranteed access to an employment-based policy that provides at least as beneficial a package as the Medicaid program. What Oregonians do with this access benefit is up to them. How they spend their health care resources depends first and foremost on their initial health status, but also on their health-related preferences and those of their health care providers.

One of the initiative's objectives is to improve efficiency by moving the Medicaid population out of fee-for-service and into the prepaid system of *managed care*. By commissioning the services of cost-conscious, privately run managed care companies, coverage for the poor and uninsured could be made more affordable to the state than if traditional indemnity plans were commissioned. Furthermore, Oregon legislated mandatory community-rating of health insurance premiums in order to prevent health coverage exclusion by managed care companies owing to preexisting conditions. Since managed care companies contain costs by limiting available services and restricting patient choice, the state came up against a federally imposed hurdle that normally applies to state Medicaid programs: federal regulations prohibit rationing of services to those eligible for Medicaid. Oregon requested an exemption from these regulations. Specifically, Oregon requested permission to reduce the number of medical services covered to accommodate for a greater number of enrollees. Oregon obtained federal waivers and eventually passed its initiative through the state legislature, implementing it in 1994.

A task for health economists is to evaluate the efficiency implications of the Oregon initiative. From an apparently middle-of-the-road perspective on efficiency, the Pareto perspective, the Oregon initiative is *socially* inefficient. This is because some will lose benefits as a result of the initiative's implementation. Efficiency appears to be sacrificed for the sake of improved fairness in terms of increased access to and use of the Medicaid system.¹⁶ The Paretian perspective does not, however, consider the possibility of *net* social gains or losses in terms of *both* efficiency and fairness. The Paretian perspective narrowly focuses on the comparatively few Medicaid recipients who lose coverage of unprioritized treatments as a result of health service rationing, and not on the accruing net gains of the entire population of Medicaid recipients. If the social gains in terms of *both* fairness and efficiency outweigh the losses, then a positive-sum game is



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

feasible.

To judge whether there is a positive-sum game, economists would have to examine both efficiency and fairness indicators. On the efficiency side, health economists must evaluate whether covering the uninsured poor is effective without substantially raising the costs of the state's Medicaid program. Empirically, the benefit side of efficiency comprises at least three separate indicators: *access* to health care services, *use* of state health care services, and *health outcomes* related to use. Of these three indicators, data unequivocally confirm improved access. Oregon has added more than 100,000 people to the Medicaid program (Bodenheimer 1992, Blumstein 1997). All persons under the poverty level are now eligible for Medicaid. Before the plan was instituted, only 57 percent of these people were eligible. Furthermore, the initiative mandates health care access for *all* Oregonians to a basic (though comprehensive) health care package. Data also suggest limited improvements in both the use of health care services and health outcomes among the Medicaid population (Blumstein 1997, Oregon State Health Commission Report 1997). On the cost side, the crucial parameter to measure is the difference between actual costs of the currently covered Medicaid population and projected Medicaid costs if the poor uninsured population were not covered. Actual costs have remained within the state's Medicaid budget. However, a better indicator of efficiency cost, opportunity cost, and the cost of alternative ways of improving access, has not been calculated. This makes it difficult to state unambiguously whether the Oregon proposal is or is not cost-effective.

The Oregon initiative appears to promote a *fairer* state Medicaid health care system on two counts. First, if we judge fairness in terms of 'like cases being treated like,' with the only morally relevant likeness being medical need, the Oregon initiative appears fair because, across the entire poor population, equal resources are being expended for equivalent medical needs, regardless of ability to pay. Second, the Oregon initiative enforces community rating of health insurance premiums. As a result, it protects those with preexisting conditions from having to pay exorbitantly high insurance premiums.¹⁷ The Oregon initiative does, however, limit explicit rationing to Medicaid recipients. For this reason, it has been labeled by some as unfair. Critics attack the plan for making the poor bear the burden of providing universal access. Despite it being true that rationing falls on the shoulders of the poor,¹⁸ the plan has not led to a widening of the gap between the health outcomes of the rich and poor (see the Oregon State Health Commission Report 1997). It has improved the poor's *relative* health status compared to the



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

wealthy, even though some current Medicaid recipients are worse off than they were before. As Daniels (1991) predicted, 'the loss of less important services by [some] recipients is more than counterbalanced by the gains of the uninsured. As a result, the plan reduces overall inequality between the poor and the rest of society' (2234).

In brief, Oregon, in theory, shows how fairness and efficiency can work together. In addition, Oregon illustrates a unique opportunity for the private sector, managed care companies, to cooperate with the state's regulatory bodies in mandating access to health care services. This said, in practice, physicians and policy analysts have to learn to accept that increasing access to as many as possible comes at the price of withholding services from some. We therefore need measures of fairness that take account in a sophisticated way of issues such as access, so tradeoffs between fairness and efficiency can be identified. We also need to be able to identify situations where tradeoffs between fairness and efficiency do not occur—situations which have generally been ignored in the economics literature.

Notes

1. In this context, equality is used as a proxy for fairness. Needless to say, equality and fairness are not the same thing. Under certain circumstances, being fair may imply actually treating individuals unequally.
2. Okun distinguishes between the rights and dollars domains. However, the line between the two domains is semi-permeable. Notably, property rights are part and parcel of the dollars domain.
3. Of course, in some instances, policy makers may actually deregulate sectors within the dollars domain to promote efficiency, as has been done throughout the 1980s and 1990s.
4. Most health care rights are termed negative rights. Other than a right to emergency treatment, United States legal statutes do not guarantee a citizen's positive right to health care services.
5. Rawls uses Pareto efficiency as a measure of efficiency, while his concept of equality indicates an equal division of 'primary goods' across the population. Since Rawlsian primary goods include rights as well as commodities, the equality-efficiency tradeoff suggested by Rawls occurs across the rights and dollars domains.
6. Health care resources are necessary, though not sufficient, conditions for producing health benefits. Other resources such as education also contribute to better health outcomes.
7. Note that any move from point to point along the Pareto-efficient locus MN would



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

- be socially suboptimal in terms of the Pareto principle. That is, moves from point to point along MN imply a zero-sum game where one person wins while another loses. Since Pareto only allows for ordinal comparisons between persons, the balance of losses and gains cannot be assessed.
8. Of course, the move from C to D implies that one beneficiary gains while the other loses; a Pareto-inefficient move. But both C and D are technically Pareto-efficient points.
 9. Whether resource equalization is fair or not depends on the initial dispersion of medical needs across the population. If needs are similar across the population, then resource equalization leads to a fair distribution. If needs are dissimilar, it will have to be judged whether the resulting variance in health outcomes is fair or not. To make it fair, the resource-equalization rule may have to be supplemented by the establishment of a 'necessary minimum' level of resources, below which no subpopulation is allowed to fall.
 10. The 1983 Commission was careful not to suggest that the government should establish health care as a constitutional or legal right.
 11. Empirical evidence strongly suggests that lack of (adequate) insurance decreases health care access leading to significantly worse health outcomes. A person's condition upon hospital admission, his use of resources during hospitalization, and likelihood of death all vary according to the person's health insurance status (see Hadley et al. 1991, Lefkowitz and Monheit 1991).
 12. Kitzhaber is now Governor of Oregon.
 13. Of course, for certain conditions, there may be more than one therapy involved.
 14. The line was drawn at 587 on the original list of 700 condition-treatment pairs.
 15. It must be noted that in Oregon, before the initiative, one third of the uninsured were Medicaid eligible, while twice that many were the working poor and their dependants—people not eligible for public assistance, yet who could not afford their own health insurance.
 16. Technically, however, the overhauled Oregon Medicaid system may be just as efficient as before. To illustrate this, see Figure 4.3. Suppose X represents that part of the population who loses some benefits as a result of the overhaul. Y represents that part of the poor population who gains benefits. Assuming that the Pareto-efficient locus does not shift in- or outwards, we could depict the gain accruing to Y and the loss to X as a move along the MN curve from point C to point D.
 17. The Oregon Medicaid program is also a popular program. In a survey carried out by the Oregon State Health Commission, 88 percent of Medicaid patients 107 said they were satisfied with the health care services provided to them (Oregon State Health Commission Report 1997).
 18. This situation is not new. Before implementation of the initiative, health care to



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

the poor was rationed by way of ability to pay.

Bibliography

- Aday, L., Andersen, R., and Fleming, G. (1980) *Health Care in the US: Equitable for whom?* Beverly Hills CA: Sage Publications.
- Arrow, K. (1951) *Social Choice and Individual Values*, Oxford: Oxford University Press.
- Beauchamp, D. (1989) *The Health of the Republic*, Philadelphia PA: Temple University Press.
- Becker, G. (1976) *The Economic Approach to Human Behavior*, Chicago: University of Chicago Press.
- Blumstein, J. (1997) The Oregon experiment: the role of cost-benefit analysis in the allocation of Medicaid funds, *Social Science and Medicine* 45: 545-555.
- Bodenheimer, T. (1992) The Oregon Health Plan—lessons for the nation—part one, *The New England Journal of Medicine* 337: 651-655.
- Caplan, A. (1997) *Am I My Brother's Keeper*, Bloomington IN: Indiana University Press.
- Churchill, L. (1994) *Self-Interest and Universal Health Care*, Cambridge MA: Harvard University Press.
- Cohen, J. (1996) QALYs, needs, and preferences, *Journal of Medical Ethics*: 267-272.
- Culyer, A. (1990) Commodities, characteristics of commodities, characteristics of people, utilities and the quality of life, in S. Baldwin, C. Propper, and C. Godfrey (eds) *Quality of Life Perspectives*, London: Routledge.
- Daniels, N. (1985) *Just Health Care*, Cambridge: Cambridge University Press.
- (1991) Is the Oregon rationing plan fair? *The Journal of the American Medical Association* 265: 2232-2236.
- Dworkin, R. (1985) *A Matter of Principle*, Cambridge MA: Harvard University Press.
- Hadley, J., Steinberg, E., and Feder, J. (1991) Comparison of uninsured and privately insured patients: condition on admission, resource use, and outcome, *The Journal of the American Medical Association* 265: 374-379.
- Kitzhaber, J. (1990) Speech before the Oregon Senate.
- Klappholz, K. (1972) Equality of opportunity, fairness, and efficiency, in M. Peston



ACCOUNTING FOR FAIRNESS AND EFFICIENCY IN HEALTH ECONOMICS

JOSHUA COHEN AND PETER UBEL

Excerpted from *The Social Economics of Health Care*

- and B. Corry, *Essays in Honor of Lord Robbins*, London: Weidenfeld and Nicolson.
- Lefkowitz, D. and Monheit, A. (1991) Health insurance, use of health services, and health care expenditures, *National Expenditure Survey Research Findings 12*, AHCPR Pub. No. 92-0017, Rockville, Maryland.
- Mitchell, L. (1998) *Stacked Deck*, Philadelphia PA: Temple University Press.
- Nord, E., Pinto, J., Richardson, J., Menzel, P., and Ubel, P. (1999) The value of DALY life: problems with ethics and validity of disability adjusted life years, *British Medical Journal* 319 (27 November): 1423-1425.
- Office of Technology Assessment Study (1992) *Evaluation of the Oregon Medicaid Proposal*, US Government Printing Office.
- Okun, A. (1975) *Equality and Efficiency: The Big Trade-Off*, Washington DC: The Brookings Institution.
- Oregon State Health Commission Report (1997).
- President's Commission for the study of ethical problems in medicine and biomedical and behavioral research (1983) *Securing Access to Health Care*, US Government Printing Office.
- Rawls, J. (1971) *A Theory of Justice*, Cambridge MA: Harvard University Press.
- Robbins, L. (1932) *An Essay on the Nature and Significance of Economic Science*, London: Macmillan.
- Rowley, C. and Peacock, A. (1973) *Welfare Economics: A Liberal Restatement*, London: Martin Robertson.
- Samuelson, P. (1947) *Foundations of Economic Analysis*, Cambridge MA: Harvard University Press.
- Sen, A. (1987) *On Ethics and Economics*, Oxford: Oxford University Press.
- Ubel, P., Arnold, R., and Caplan, A. (1993) Rationing failure: the ethical lessons of the retransplantation of scarce vital organs, *The Journal of the American Medical Association* 270: 2469-2475.
- Ubel, P., DeKay, M., Baron, J., and Asch, D. (1996) Cost-effectiveness analysis in a setting of budget constraints: is it equitable? *The New England Journal of Medicine* 334: 1174-1178.



CHAPTER

2

HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL
HAMPUS LYTTKENS

Individual Decisions
for Health

Edited by
Björn Lindgren


Routledge International Studies in Health Economics

This chapter is excerpted from
Individual Decisions for Health
Edited by Björn Lindgren.

© 2003 Taylor & Francis Group. All rights reserved.

 [Learn more](#)



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

Introduction

Ever since the seminal article by Arrow (1963), it has been a commonplace to note that the presence of uncertainty is a particularly salient feature in decisions on health and health care. A number of studies have appeared that introduce probabilistic elements in the individual's demand for health, and there is a substantial literature that focuses on the value of changes in health risks. For example, in the so-called demand-for-health tradition – emanating from Grossman's (1972) well-known work – there are several probabilistic versions,¹ and the implications of risk in the health production function have been explored in a valuation context (Johansson, 1994).² There remains to introduce, however, the distinction between risk – i.e. uncertainty with known probabilistic properties – and genuine uncertainty – i.e. uncertainty with *unknown* probabilistic properties – in formal models of health-related behaviour.

In this chapter, we introduce genuine uncertainty as well as uncertainty aversion into a simple model of individual decision-making about health-related activities. This is potentially important, since, arguably, the conditions for many of the decisions an individual makes concerning his own health more closely resemble conditions of genuine uncertainty than pure risk. The individual is likely to be highly uncertain if not completely ignorant about the probabilities involved. For example, he may be vaguely aware that there is a serious disease called leukaemia while having no idea of whether he is likely to get it and no idea about his possibilities to affect the likelihood of getting it.

Within this format, we investigate the consequences for health-related behaviour of individual attitudes towards health and information as well as of exogenous changes in the individual's decision environment. In the section on Individual trade-offs and personality traits we show how health-related decisions are affected by the individual's degree of pessimism with respect to health states with unknown probabilities; we also consider the effects of the individual being confident or diffident about the accuracy of the information he possesses.

Furthermore, we analyse in the section on Individual trade-offs and exogenous changes in information, prevention, prices, and income, how different forms of health policy and other exogenous factors are likely to affect individual behaviour, in particular the individual's propensity to gather information and engage in own prevention. Specifically, we explore how the individual reacts when provided with information without cost (exogenous information), when there is a reduction in the



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

health risks he faces (exogenous prevention), and finally when he learns something new about the welfare implications of a particular health state. We also make certain observations about the effects of changes in prices and income. In the section on Health policy and socio-economic differences in health and behaviour we discuss the implications of our results for differential behaviour across socio-economic groups, and for the effects of health policy on these cross-sectional clusters.

The model

We now turn to constructing a model where an individual faces two sorts of health states – those to which he can assign probabilities and those which he is genuinely uncertain about. He can influence the relative importance of genuine uncertainty by collecting information.

General specification

Individual utility depends on the anticipated health status, h , and on a vector, a , of activities. Health statuses are ranked in terms of *healthy days*, h^s , and they may equivalently be thought of as being characterized by a number of quality-adjusted healthy days, so that the ranking also takes account of quality-of-life aspects; the key requirement is that the individual is able to make an unambiguous ranking.

The utility function is

$$U(a) = \gamma(a) \cdot \sum_{s=1}^S \pi_s(a) u(h^s, a) + (1 - \gamma(a)) \cdot u_0(a). \quad (3.1)$$

It is a convex combination – weighted by $0 \leq \gamma(a) \leq 1$ – of two components, the first being an ordinary expected utility, and the second one being (a reduced form for) a generalized expected utility. We describe these in turn.

1. The first component is an expected utility over S states of health whose outcomes are ranked from better to worse in terms of healthy days; i.e. $h^1 > \dots > h^S$. (Note that we employ the convention that a state with a higher number is worse.) The individual assigns probabilities, $\pi_1(a), \dots, \pi_S(a)$, to these health states, and u is a von Neumann–Morgenstern utility function which is positive and increasing in h .



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

2. The second component is a reduced form, sharing a number of properties with a generalized expected utility exhibiting 'uncertainty aversion'.³ To construct it, we assume the existence of a set of states $S + 1, \dots, S + S'$ in addition to the states described under 1. These states are distinguished by the individual not being able to assign firm probabilities to them. In the literature on genuine uncertainty and uncertainty aversion,⁴ uncertainty is captured by the assumption that the individual assigns *generalized probabilities* to these in a fashion that is 'pessimistic'. The theory works in a way such that states are assigned generalized probabilities that need not sum to one (and that are not required to satisfy standard additivity properties). When expectations are computed these generalized probabilities are transformed into weights summing to 1. The result of this evaluation is a generalized expected utility $u_0(a) = \tilde{E} \{u(h; a) \mid S+1, \dots, S+S'\}$ where the tilde signifies that the evaluation is by the generalized probabilities. The generalized expectation is taken over the same von Neumann–Morgenstern utility function.⁵ We will work directly with $u_0(a)$ which is a reduced form in that it captures *the effect* of both the genuine uncertainty and the uncertainty aversion (or pessimism) that the individual exhibits in evaluating the uncertainty.

Some of the activities, a , are costly in the individual's budget. The budget should be thought of as, somewhat roughly, capturing the objective costs of various activities – e.g. costs in terms of money and time – while the subjective costs would be captured by the utility function depending directly on a .

The probabilities, $\pi_1(a), \dots, \pi_S(a)$, depend on a , and the cumulative distribution is denoted $F(h; a)$ (i.e. F is the probability that the number of healthy days is less than or equal to h given a).

We will call the states $1, \dots, S$ to which the individual assigns ordinary probabilities the 'firm states', and the remaining states, $S+1, \dots, S+S'$, are called the 'uncertain states'.

Although one could work with a more general specification (with the individual's probabilistic frame of reference being specified completely in terms of generalized probabilities depending on activities), such a specification would be operationally very weak since there is no agreed notion of uncertainty and uncertainty aversion. Our specification has the significant advantage of providing a natural parameterization (through y) of the uncertainty faced by the individual.⁶ Moreover,



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

it is possible to rank the degree of uncertainty aversion (being captured by u_0) and the probabilistic favourability of the firm states (through $\pi_1(a), \dots, \pi_S(a)$) within this framework.

It is important to realize that the loss of generality that follows from the reduced form is only in terms of the flexibility of the utility function (the restriction is qualitatively very similar to imposing, e.g. separability restrictions on a utility function). There seems to be little conceptual loss of generality in this simplification. If for example an exogenous event – like the alarm about the mad cow disease in Britain in 1996 – simultaneously makes people more genuinely uncertain and more dismal regarding the probability of brain disease, this is captured perfectly well by a simultaneous shift in the degree of uncertainty (γ) and in the probabilities. Conversely, it is clear that conventional probability-based models are inherently incapable of capturing uncertainty. One may note that at this stage, the model is algebraically equivalent to a model with only risk present and with u_0 occurring with probability γ , but one should also note that the equivalence is upset as soon as one considers an exogenous change or behavioural implications; for example, preventive behaviour is assumed to have a known effect on the probabilities for the firm states but not on the assessment of the likelihood of uncertain states.

To recapitulate briefly, the individual utility function is given by a convex combination of a standard expected utility over the probability distribution $\pi_1(a), \dots, \pi_S(a)$ over the firm states $1, \dots, S$, (with $\sum \pi_s = 1$), and a reduced form of a generalized expected utility, u_0 , over the uncertain states. He assigns weight γ to the firm states. Hence γ measures the degree of certainty or, as we shall phrase it later, the individual's degree of *confidence* in his information about the firm health states; u_0 , on the other hand, measures the degree of uncertainty aversion – i.e. the utility loss from being uncertain.⁷

Pessimism

The basic tenet of the literature on uncertainty aversion is one of pessimism, and basic pessimism with respect to genuine uncertainty will be the maintained assumption unless otherwise stated. We will take this to mean that the individual assigns higher expected utility to the firm states than to the uncertain states; i.e. that

$$u_0(a) < \sum_1^S \pi_s u(h^s; a) \quad (3.2)$$



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

for each a . In words, the utility of being completely uncertain ($\gamma = 0$) is always lower than the utility of facing no genuine uncertainty ($\gamma = 1$). For many health applications it seems reasonable to assume this to be the main case. It is clear, however, that the model can also capture fundamental optimism, and we will make a comment on that case in the subsection on Pessimism and confidence: basic analysis.

The *degree of pessimism* is arguably an important trait in an individual's personality. There is obviously scope for more or less pessimism, and this corresponds to the individual exhibiting more or less uncertainty aversion; it is modelled by allowing u_0 to vary. In our terminology, a person with a large u_0 is less pessimistic; such a person values relatively highly the health states that are surrounded by genuine uncertainty.

Information and confidence

We will assume that the degree of certainty (γ) is affected by information. It seems intuitively very plausible that by gathering information the individual can reduce the relative importance of uncertainty in his life. Hence we assume that he may take an active decision to increase γ by gathering information and can weigh the pros and cons of such a decision. The decision cannot be contingent on the expected direction of the information; i.e. while the individual may well find *ex post* that π_S has changed in the light of new information, such changes cannot influence the decision to gather information. Assuming this seems natural in our case.⁸ This part of the model is primarily intended to represent the acquisition of general information about health risks. This is in contrast to, for example, the health care setting of undergoing a diagnostic test to determine whether one has contracted a specific disease and submitting to treatment accordingly (individual-specific information); we do not model such a sequence of events – just the individual's propensity to seek information and engage in prevention.

A natural interpretation of the weight γ is that it is a measure of how much *confidence* the individual has regarding the information that he has about the probabilities of the firm health states. Henceforth we will say that a larger γ corresponds to the individual being more confident about the accuracy of his information. Conversely, a person who is diffident about his information (small γ) puts more weight on the uncertain states of the world.⁹

Confidence, notably, is not only a result of information gathering. Rather, just like pessimism, the degree of confidence is an important personality trait in itself, and



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

the implications of this can be explored by allowing y to vary.

Prevention and risk

The final basic feature of the model is *conventional risk* – i.e. probabilities assigned to different health states as described by the distribution $F(h; a)$ (or, equivalently, $\pi_1(a), \dots, \pi_S(a)$). This distribution may change as the result of health-promoting activities. Here we will make use of the fact that probability distributions over the states $1, \dots, S$ are ordered by (first-order) stochastic dominance. One distribution *stochastically dominates* another if for each outcome, h , something smaller is less probable according to the dominating distribution; i.e. a distribution with cumulative distribution function G dominates a distribution F if for each h , $G(h) \leq F(h)$.¹⁰ The activity a_i makes the distribution more favourable if

$$F_{a_i}(h; a) = \frac{\partial F(h; a)}{\partial a_i} \leq 0;$$

i.e. if it induces a shift in the distribution in the direction of stochastic dominance.

We define prevention as an activity which shifts the distribution F in the direction of stochastic dominance (an activity with the opposite effect would be hazardous).¹¹

The interaction between the risk and uncertainty aspects of the model seems to capture an important element in decisions about health-related activities. Even when we are dealing with risk – so that an individual is implicitly thinking in terms of a probability for, say, lung cancer – it seems reasonable to argue that he is often unsure about whether he has in fact the correct probability (π_s) and about the effect of his actions on the probability of ill health ($\partial \pi_s / \partial a$), e.g. the effect of smoking on the probability of lung cancer. The size of y reflects the degree to which the individual is confident about π_s and $\partial \pi_s / \partial a$ (though we cannot separate the two attitudes).

Note that in this model, prevention has no effect on utility under the uncertain health states (u_0). Consequently, the individual will never engage in prevention in order to increase u_0 . The reason for this is the sharp formal distinction in the utility function between the health states with probabilities attached and those where there is genuine uncertainty. (In relation to fully general uncertainty models, this assumption is analogous to assuming that the lower bound of the support – i.e.



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

the worst outcome in terms of healthy days – cannot be affected by prevention.) As soon as the individual believes that it is possible to influence the probability of health outcomes – even if it is a very vague belief or hope – the probabilistic part of the utility function is involved.

A specified set of individual activities

In the following, we will make an unambiguous distinction between information gathering and prevention, which are both distinguished from consumption. We will distinguish between three types of activities that enter the individual's decision problem: information gathering (generically denoted a_1), prevention (a_2), and consumption (a_3). This is not to say that these activities are always separable, but separability is very useful in order to keep the arguments transparent; it is also clear that the effects of an activity entailing both information gathering and prevention are the obvious 'sum' of the effects of the components adjusted for complementarities.

Hence we assume that the informative activity (a_1) affects y but not F , and conversely for the preventive activity (a_2). We will also assume that neither prevention nor information gathering enters the individual's utility function directly (doing so would simply add a cost component – direct negative effects on utility – and this is unlikely to change the results qualitatively); utility is thus a function of health and consumption (a_3). This is in line with the Grossman (1972) model with its basis in household production theory, where healthy days appear in the utility function together with consumption, and where the healthy days are ultimately produced by medical inputs and the like. Consumption is assumed not to influence health. We should note that with this kind of formulation, we cannot analyse strictly hazardous activities such as smoking or rock climbing, which are usually activities that have a positive direct effect on utility but a negative expected net effect on health.

The budget and the maximization problem

We will assume that all our three activities are costly, i.e. have positive prices in the individual's budget. For simplicity, we take the individual's income A to be exogenously given and not dependent on health. In terms of the Grossman (1972) model, one could say that we are investigating a consumption model of health. It seems intuitively clear that if health also had a positive effect on A , this would increase the attractiveness of health-promoting activities in our model.



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

The individual strives to maximize utility. In order to obtain more convenient expressions, we let h^{s+1} denote a fictitious worst state (thus redefining h^{s+1}). With the convention that $u(h^{s+1}; a) = 0$, an equivalent representation of the problem (which follows from a manipulation of sums analogous to integration by parts) is

$$\begin{aligned} \max U(a_1, a_2, a_3) &= \gamma(a_1) \sum_{i=1}^s (u(h^s, a_3) - u(h^{s+1}, a_3))(1 - F(h^{s+1}; a_2)) \\ &\quad + (1 - \gamma(a_1))u_0(a_3), \\ \text{s.t. } \sum_{i=1}^3 p_i a_i &\leq A, \quad a_i \geq 0, \quad i = 1, 2, 3. \end{aligned}$$

We assume that information gathering increases the individual's degree of confidence, $\gamma'(a_i) > 0$, but at a decreasing rate, $\gamma''(a_i) < 0$. This amounts to assuming information acquisition to be a concave problem. We will also assume that choosing the preventive activity is a concave problem; this requires that F is convex in a , i.e. that $F_{aa} > 0$, and that utility is concave in consumption. In addition, we assume that the objective function is jointly quasi-concave in a , and hence that first-order conditions define optimal choices. Finally, we assume that the marginal utility of consumption is independent of the health state. This seems to be a reasonable simplification, since it is equally possible to argue that the marginal utility of consumption is higher for a healthy individual as it is to argue that it is lower. In particular, we assume that the marginal utility of consumption is the same in the firm health states as in the uncertain ones.

Individual trade-offs and personality traits

In this section, we will investigate how the individual's trade-off between activities is affected by his attitudes towards health in terms of pessimism and confidence.

Pessimism and confidence: basic analysis

In order to illustrate our assumption of basic pessimism, we begin by looking at the first-order condition with respect to the information activity a_1 ; for an interior solution, it is

$$\frac{\partial \gamma}{\partial a_1} \left(\sum_{s=1}^s (u(h^s, a) - u(h^{s+1}, a))(1 - F(h^{s+1}; a_2) - u_0(a_3)) \right) = \lambda p_1.$$



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

The derivative of y is positive by assumption, and the right-hand side is positive if the price of activity a_1 is positive. Hence, an interior solution exists only if the expression in the parenthesis is positive, i.e. if the utility from being completely certain is larger than the utility from being completely uncertain; if that is not the case, $a_1 = 0$. This has an interesting interpretation: the individual will invest in gathering information only if he – taking his attitude toward uncertainty into account – from the outset is better off in the firm states; i.e. if he is a basic uncertainty pessimist as we have assumed (cf. expression (3.2)). Conversely, if the individual is instead a basic uncertainty optimist (the opposite of expression (3.2) holding), he may prefer to engage in an activity which makes him more uncertain (e.g. seeking out conflicting information, or gathering information even if he believes it will confuse him).

The last observation has a direct correspondence in the welfare effects of exogenously provided information. Consider an impact, g , which the agent is exposed to exogenously. The natural interpretation of g is information provided by the government, but other interpretations are possible. A complete welfare analysis would consider the trade-off between the positive effects of g and the cost of providing it; we confine ourselves, however, to the simple case where it is costless, because the substantive point to be made stands out more clearly without costs.

Proposition 3.1 *Exogenous information is welfare-improving if the individual is a basic uncertainty pessimist; it has a negative welfare effect if the individual is a basic uncertainty optimist. (See the Appendix for proof.)*

Performing a similar analysis for exogenous prevention – i.e. an activity that improves the individual's health – it is clear from the proof that it is unambiguously welfare-improving. These results are not surprising but nevertheless important since they stress the significance of the assumption of pessimism. The welfare implications of interventions that entail changing prices of the individuals' own efforts are completely clear; lowering prices, and thereby expanding individuals' feasible sets, is always at least weakly beneficial.

It is worth noting that the result highlights a difference between our model and expected-utility models where individuals may gather pieces of information (typically called 'experiments') whose degree of informativeness differ.¹² In such a framework, basic consistency requirements imply that the expected utility *ex ante*



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

be independent of the informativeness of the experiment performed in the absence of a behavioural response to the experiment (the purpose of the experiment being to adjust actions conditional on it). This difference illustrates the additional degree of freedom introduced by allowing for attitudes toward uncertainty. One virtue of allowing for such attitudes is, in our view, that intuitive notions of pessimism and optimism are captured in a more germane fashion.

We should note here that the case of basic uncertainty optimism does not seem altogether unrealistic. One may well imagine that, for example, certain individuals are better off not knowing that they may some day contract osteocarcinoma. Remember, however, that in the analysis to follow, the maintained assumption is that the individual is a basic pessimist with respect to uncertainty.

Given basic pessimism, we now focus on the individual's *degree* of pessimism, i.e. the effects of variations in u_0 . We will employ the traditional approach of differentiating the first-order conditions (henceforth FOC) – which in our case include the FOC for a_1 , a_2 , and a_3 – and the budget constraint implicitly and then solving for the derivatives in question. The FOC are (where we denote derivatives of γ with primes since it only depends on a_1)

$$\begin{aligned} \gamma' \cdot \left[\Sigma(u^s - u^{s+1})(1 - F(h^{s+1}; a)) - u_0(a) \right] - \lambda p_1 &= 0, \\ \gamma \cdot \Sigma(u^s - u^{s+1})(-F_a(h^{s+1}; a)) - \lambda p_2 &= 0, \end{aligned} \quad (3.3)$$

$$\partial_{a_3} u - \lambda p_3 = 0,$$

$$A - p_1 a_1 - p_2 a_2 - p_3 a_3 = 0.$$

The differentiated FOC, letting $\Delta u^s = u^s - u^{s+1}$, take the form

$$\begin{bmatrix} \gamma''(\Sigma \Delta u^s F - u_0) & -\gamma' \Sigma \Delta u^s F_a & 0 & -p_1 \\ -\gamma' \Sigma \Delta u^s F_a & \gamma \Sigma \Delta u^s F_{aa} & 0 & -p_2 \\ 0 & 0 & \partial_{a_3}^2 u & -p_3 \\ -p_1 & -p_2 & -p_3 & 0 \end{bmatrix} \cdot \begin{bmatrix} \partial a_1 / \partial u_0 \\ \partial a_2 / \partial u_0 \\ \partial a_3 / \partial u_0 \\ \partial \lambda / \partial u_0 \end{bmatrix} = \begin{bmatrix} \gamma' \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (3.4)$$



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

where the right-hand side is the negative of the derivative of the first-order conditions with respect to u_0 . The first factor on the left-hand side is the Jacobian matrix for the FOC. Under the assumption that the individual's objective function is (strictly) quasi-concave, we know the FOC define the unique maximum; moreover, we know that the determinant, D , is negative (Simon and Blume, 1994).

The upper left corner of the inverse of the Jacobian matrix is $1/D$ times the following expression

$$\begin{bmatrix} -p_2^2(\partial^2 u/\partial a_3^2) + p_3^2\gamma\Sigma\Delta u^s F_{aa} \\ p_1 p_2(\partial^2 u/\partial a_3^2) - p_3^2\gamma'\Sigma\Delta u^s F_a \\ -p_3(p_1\gamma\Sigma\Delta u^s \partial_a^2 F - p_2\gamma'\Sigma\Delta u^s F_a) \end{bmatrix}.$$

The derivative of a_1 with respect to u_0 is thus

$$\partial a_1/\partial u_0 = \frac{\gamma'}{D} \left(-p_2^2(\partial^2 u/\partial a_3^2) + p_3^2\gamma\Sigma\Delta u^s F_{aa} \right) < 0,$$

and we see that a less pessimistic assessment of the remaining number of healthy days in states that the individual cannot fully describe leads to less resources being spent on information gathering. This is natural since information in this model has the effect of reducing the weight placed on these uncertain states (increasing γ); to gather information is then less attractive, the more highly these uncertain states are valued. Inspection of the corresponding expression for a_2 ,

$$\frac{\partial a_2}{\partial u_0} = \frac{\gamma'}{D} \left(p_1 p_2 \frac{\partial^2 u}{\partial a_3^2} - p_3^2 \gamma' \Sigma \Delta u^s F_a \right),$$

shows that the effects on prevention are ambiguous; more specifically, a_2 decreases along with a_1 if marginal utility of consumption (a_3) is constant, and it is substituted for a_1 if u is sufficiently concave in consumption. The magnitude of changes in the marginal utility of consumption is positively related to the



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

magnitude of *income effects*; this will be discussed in the next subsection. Similar inspection of the derivative of a_3 shows that consumption unambiguously increases.

Proposition 3.2 *Reduced pessimism will divert resources from information gathering, and shift some resources toward consumption; prevention will decrease if income effects are small, and it will increase if income effects are large.*

We now turn to the difference in behaviour between individuals who are inherently confident and individuals who are diffident about the information that they possess. In our model, this corresponds to a shift in the parameter γ . The following statement is a corollary of Proposition 3.2 (and Table 3.2), and it is therefore stated without proof.

Proposition 3.3 *Confidence will increase prevention; information gathering will increase if income effects are small, and decrease if income effects are large.*

This proposition introduces a fundamental *complementarity* which we will encounter repeatedly, and which will be discussed in more details below. This is the complementarity between confidence on the one hand, and a favourable distribution over firm states on the other; in this instance, increasing confidence increases the marginal returns to prevention. The effect on information gathering depends on whether the individual is willing to substitute it for consumption, which in turn depends on income effects to which we now turn.

Income effects

As we have just indicated, several of our results depend on the magnitude of *income effects* of a_1 and a_2 ; i.e. on the extent to which the demand for a_1 and a_2 depends on A . Income effects are small if this dependence is weak; in particular, income effects are zero if a_1 and a_2 are independent of A . In terms of the utility function, income effects are zero if marginal utility of consumption, $\partial u/\partial a_3$, is independent of (a_1, a_2, a_3) ; since we have ruled out $\partial u/\partial a_3$ depending on (a_1, a_2) , only the dependence on a_3 shows up in our expressions.¹³ Throughout the chapter, we will sort out the two obvious cases; the main case will be that they are small, and we will then consider the case where they are large. (Our encountering this problem is not novel. Consumer-surplus analysis and the Coase theorem are significant examples of modes of analysis that depend on income effects being small.)

At a fundamental level, the most convincing reason for income effects being small



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

is that the objects affected directly by an exogenous change constitute a share of the budget; marginal utilities of other goods will then be affected only negligibly. Consequently, the income effect will be negligible. Thus, attention should be concentrated on the case with small income effects in the context of changes that do not affect an individual's total consumption of other goods too much.

This suggests that the nature of the individual's budget is important in this context. Consider first the case when the budget is thought of only in monetary terms. It then seems quite plausible that income effects are small. An individual's outlays on information and prevention is usually only a small part of his budget. This is particularly the case since unhealthy behaviour is implicitly subsidized by (social and private) health insurance when premiums are not differentiated with respect to preventive or hazardous behaviour (Kenkel, 2000, Sec. 3).

The situation is somewhat different in the case when the budget also includes time costs. Gathering information often takes more time than money. Furthermore, physical exercise is a preventive activity that often seems to entail significant time costs; jogging or going to the gym four times a week seems to be good cases in point. Hence, even if leisure is valued less than working time, it does not seem inconceivable that one could encounter cases where income effects are significant.

Individual trade-offs and exogenous changes in information, prevention, prices, and income

This section is devoted to analysing exogenous changes in the individuals' decision environment. The discussion will mostly be phrased in terms of public programmes, but the analysis covers several other possible mechanisms, such as when a person unexpectedly learns that he suffers from diabetes.

We will proceed by invoking the assumption that income effects are small. In formal terms, the first two columns of the inverse of the Jacobian matrix of expression (3.4) are as follows, where the factor $1/|D|$ is omitted, and thus the sign of the determinant taken into account; for the last row, we only state the signs.



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

$$\begin{bmatrix} \underline{p_2^2(\partial^2 u / \partial a_3^2)} - p_3^2 \gamma \Sigma \Delta u^s F_{aa} & \underline{-p_1 p_2 (\partial^2 u / \partial a_3^2)} + p_3^2 \gamma' \Sigma \Delta u^s F_a \\ \underline{-p_1 p_2 (\partial^2 u / \partial a_3^2)} + p_3^2 \gamma' \Sigma \Delta u^s F_a & \underline{p_1^2 (\partial^2 u / \partial a_3^2)} + p_3^2 \gamma'' (\Sigma \Delta u^2 F - u_0) \\ p_3 (p_1 \gamma \Sigma \Delta u^2 \partial_a^2 - p_2 \gamma' \Sigma \Delta u^s F_a) & \underline{-p_3 (p_2 \gamma'' (\Sigma \Delta u^s \cdot F u_0) + p_1 \gamma' \Sigma \Delta u^s F_a)} \\ - & - \end{bmatrix}$$

All terms involving the income effect are underlined, and by income effects being *small*, we mean that these terms are small enough not to determine signs. The sign pattern of the above matrix is then

$$\begin{bmatrix} - & -* \\ -* & - \\ + & + \\ - & - \end{bmatrix}, \quad (3.5)$$

where the asterisks indicate which signs depend on income effects being small.

We explore three sets of exogenous changes directly related to health. First, we look at exogenous information that affects the individual's degree of confidence (γ). Second, we investigate the effects of exogenous prevention, which changes the probability distribution for known health states (F); exogenous information and exogenous prevention may be either a complement or a substitute to one's own corresponding activities. Third, the individual may learn something new about the implications of a specific (firm) health state (h^t changes for some t). We also make certain brief observations about changes in prices and income; these are mostly obvious but also serve as a consistency check on the model.

We will deal with changes in γ , F , and h^t separately. To keep the presentation short, we will phrase the analysis in this section in terms of changes that increase γ , shift F in a favourable direction, and suggest that h^t is better than previous thought. We will discuss unfavourable information and the consequent implications in the section on health policy.

Obviously, exogenous changes may also affect the individual's degree of health-pessimism. For example, u_0 may shift downwards or upwards as news spread about AIDS or about innovations in transplantation technology. This kind of change has already been dealt with analytically in the previous section, and does not require any treatment here.



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

Exogenous information affects the degree of information-confidence (γ)

Consider some exogenous information on health received by the individual which increases γ , e.g. as a result of health campaigns organized by the government or someone else. Since the purpose of the individual's own informative activity is to increase γ , it is useful analytically to treat the new information as a costless activity (g_1) which increases γ and which may or may not be complementary with the individual's own informative efforts. The costless information activity does not enter the utility function directly.

Formally, the degree of certainty is a function $\gamma(a_1, g_1)$, where g_1 is *complementary* with the individual's own information gathering if

$$\frac{\partial^2 \gamma}{\partial a_1 \partial g_1} \geq 0;$$

i.e. if g_1 increases the marginal returns to the individual's own information gathering. If the opposite inequality holds, a_1 and g_1 are *substitutes*.

The negative of the derivative of the FOC with respect to g_1 – i.e. the relevant right-hand side of the expression corresponding to expression (3.4) but with g_1 being the exogenous change – is

$$\left[-\frac{\partial^2 \gamma}{\partial a_1 \partial g_1} \cdot \left[\Sigma \Delta u^s (1 - F) - u_0 \right] \frac{\partial \gamma}{\partial g_1} \cdot \Sigma \Delta u^s \frac{\partial F}{\partial a_2} \quad 0 \quad 0 \right]^T. \quad (3.6)$$

The expression in square brackets in the first element is positive by our assumption of basic pessimism in expression (3.2), and hence the sign pattern is $(-/+ , -, 0, 0)$ in case of complements/substitutes. Evoking the sign pattern in expression (3.5), we thus have for complements

$$\frac{\partial a_1}{\partial g_1} > 0, \quad \frac{\partial a_2}{\partial g_1} > 0, \quad \frac{\partial a_3}{\partial g_1} < 0,$$

whereas, for substitutes, all effects are ambiguous.

It is important to note, however, that for the above sign pattern to be broken, a_1 and g_1 have to be *sufficiently strong* substitutes; for the case of a zero or slightly



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

negative cross derivative, the direct effect on y^* through g_1 is certain to dominate.

Proposition 3.4 *If exogenous information is complementary with own information gathering, it will shift resources from consumption towards information gathering and prevention; if it is a substitute, the effect is ambiguous.*

The first part of Proposition 3.3 is easily derived from the above; a shift in confidence (a shift in y) corresponds to the case where the cross-partial derivative is zero.

Another way of expressing the above is to say that information is *inherently complementary* with prevention; information increases the marginal returns from prevention and vice versa. (Information increases confidence and thus the weight on the set of firm health states; the purpose of prevention is to improve the prospects in these firm states.) In the absence of income effects, the marginal utility of consumption remains constant; exogenous information increases y^* and thereby the marginal returns from prevention, which, in turn, increases the marginal benefit from own information. For this virtuous circle to be broken, exogenous and own information must be substitutes.

While it is fairly obvious that exogenous information can be a substitute to your own informative activity (you get information for free without having to look for it), it may be worth pointing out that the case of complementarity is realistic in many cases. By providing general information on health – in the form of, say, pamphlets – the government can facilitate individuals' collection of specific information since they now know what to look for and where; if you are made aware of the particular high-risk groups to which you belong, this may increase the expected return to your own information gathering.

Some readers, surprised by the complementarity between information and prevention, have objected that prevention is often a response to 'uncertainty' and to a lack of confidence in the precision of one's information. This seems to be true when more information means *less variance*. (In a setting where there is only risk, a reduction in variance seems likely to reduce prevention.)¹⁴ Our results show, however, that when more information means *less genuine uncertainty*, a force working in the opposite direction is present. As will be clear below, this force may be quite powerful in explaining some cross-sectional patterns.

A related observation is that information on specific aspects of health obtained prior to the making of preventive decisions may or may not be complementary



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

with prevention.¹⁵ This stresses the fact that our model captures trade-offs involving the acquisition of general information (see subsection on Information and confidence, p. 44).

Exogenous prevention affects the probability of known health states (F)

Government activity in areas such as workplace safety, traffic safety and air pollution can, from the individual's point of view, be seen as exogenous prevention; i.e. as exogenous activities that shift the probability distribution for known health states (F) in a favourable direction (in the direction of stochastic dominance). While we will use the term 'exogenous prevention' for such exogenous shifts in F , it is obvious that the analysis covers any exogenous occurrence that causes favourable shifts in the probability distribution for known health states. For example, public information on existing health risks may suggest that the probability of ill health is less than the individual previously has reckoned with.

As in the previous case, it is useful to treat exogenous prevention as a costless activity g_2 which may or may not be complementary with the individual's own preventive efforts (and which does not enter the individual's utility function directly). The negative of the derivative of the FOC with respect to such an activity, g_2 , is

$$\left[\gamma' \cdot \Sigma \Delta u^s \frac{\partial F}{\partial g_2} \quad \gamma \Sigma \Delta u^s \frac{\partial^2 F}{\partial a_2 \partial g_2} \quad 0 \quad 0 \right]^T. \quad (3.7)$$

The exogenous activity is complementary with own prevention if the second derivative of F is *negative*, since a better distribution corresponds to a smaller value of F . The sign pattern is $(-, -/+ , 0, 0)$ for complements/substitutes and – analogously to the case with the exogenous information affecting certainty – we have a clean result for complementary activities,

$$\frac{\partial a_1}{\partial g_2} > 0, \quad \frac{\partial a_2}{\partial g_2} > 0, \quad \frac{\partial a_3}{\partial g_2} < 0,$$

whereas for the case where g_2 and a_2 are substitutes, all are ambiguous.



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

Proposition 3.5 *If exogenous prevention is complementary with own prevention, it will shift resources from consumption towards information gathering and prevention; if it is a substitute, the effect is ambiguous.*

Again, prevention and information gathering are inherently complementary, and with small income effects, such a positive feedback circle is broken only if g_2 and a_2 are sufficiently strong substitutes. Analogously to the above, it is not hard to think of exogenous prevention that is strongly complementary with individuals' own efforts. For example, information on proper lifting techniques will help people avoid lower back pain and thereby also enable them to engage in preventive physical activities.

New knowledge about the implications of a firm health state (h^t)

We will now consider the impact of new knowledge about the implications of a firm health state in terms of the number of healthy days, quality-adjusted healthy days, etc. Formally, we will investigate what happens when h^t changes slightly for some state t ; we confine our analysis to changes that are small enough not to change the *ranking* of states.

This covers a great deal of cases of public information on the availability of new medical technologies; for example, the fact that laparoscopic surgery may now be used instead of conventional cholecystectomy, making the operation less dramatic for the patient with consequently shorter convalescence, or that stenoses of the coronary arteries nowadays are often treatable with balloon dilatation (PTCA) rather than with open heart surgery. This may also be the relevant framework for information on the existence of 'hospital infections', which makes inpatient care more hazardous, or new forms of tuberculosis which do not respond to existing therapies.

In order to structure this part of the analysis, we will distinguish two categories of preventive activities. Activity a_2 is said to be *bottom-end* preventive if

$$\frac{\partial F(h; a)}{\partial a_2}$$

is increasing in h ; i.e. if $F(h; a)$ is *more decreasing* for small h . Thus a bottom-end preventive activity has a larger effect on the probabilities of the worst realizations of h (e.g. to stay out of the sun to reduce the risk of malignant melanoma).



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

Correspondingly, a *top-end* preventive activity has the derivative F_a decreasing with h , and has its largest impact on the best realizations of h since it is most negative there (e.g. to avoid standing close to someone with a cold). Note that these are local properties that cannot generally be satisfied globally.¹⁶

Since $F(h; a)$ is piece-wise constant in h and hence independent of h in the interval considered, the negative of the derivative of the first-order condition with respect to h^t is

$$\left[-\gamma' u'(h^t)(F(h^t; a) - F(h^{t+1}; a)) - \gamma' u'(h^t) \left(\frac{\partial F}{\partial a}(h^t; a) - \frac{\partial F}{\partial a}(h^{t+1}; a) \right) \ 0 \ 0 \right]^T \quad (3.8)$$

The sign pattern is $(-, -/+, 0, 0)$ for a bottom-end/top-end preventive activity, and for a bottom-end activity, the derivatives then satisfy

$$\frac{\partial a_1}{\partial h} > 0, \quad \frac{\partial a_2}{\partial h} > 0, \quad \frac{\partial a_3}{\partial h} < 0.$$

For a top-end activity, on the other hand, the result is ambiguous.

Proposition 3.6 *If the preventive activity is bottom-end, a favourable shock will shift resources from consumption towards information gathering and prevention; if the preventive activity is top-end, the effects of a favourable shock are ambiguous.*

These results may be surprising at first sight – information to the individual that something is not as bad as he thought may induce him to increase his own prevention. Comparison with previous results suggests that favourable information is *complementary* with own prevention if prevention is bottom-end. That is actually the case. Consider a state h^t and an activity that is bottom-end preventive in the neighbourhood of h^t . A favourable change in h^t is complementary with own prevention since own prevention reduces the likelihood of the next worse state h^{t+1} (note the ordering conventions) rather than h^t , relatively more than it reduces the likelihood of h^t rather than h^{t-1} . Thus, when h^t shifts upwards (as illustrated in Figure 3.1), the marginal return (in terms of h) to the bottom-end preventive activity increases.

In conclusion, an increase in h^t increases the marginal returns from bottom-end prevention. When several states, h^t , are perturbed, the reasoning applies if prevention is bottom-end over the whole range of outcomes.¹⁷ Along similar lines,



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

a top-end preventive activity can be interpreted as a substitute for favourable information about some states, h^t . The fact that the effect on own information gathering takes the same sign as that of the effect on prevention is to be expected, given the earlier observed complementarity between prevention and information gathering.¹⁸

Changes in prices and income

Let us now turn to the effects of income, A , and prices. The calculations then require the entire top of the inverse of the matrix of derivatives, which is

$$\begin{bmatrix} - & -* & + & -(0) \\ -* & - & + & -(0) \\ + & + & - & - \\ -(0) & -(0) & - & +(0) \end{bmatrix}, \quad (3.9)$$

where the asterisks denote signs that depend on the income effect being small, and where the zeros in parentheses indicate terms that vanish if the income effect vanishes.

The corresponding right-hand sides, i.e. the negative of the derivatives with respect to income and the three prices are

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} \lambda \\ 0 \\ 0 \\ a_1 \end{bmatrix}, \begin{bmatrix} 0 \\ \lambda \\ 0 \\ a_2 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \lambda \\ a_3 \end{bmatrix}.$$



Figure 3.1 An improvement in the state h^t .



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

It is easily seen that the effects of increases in income are positive for both information gathering and prevention; i.e. both a_1 and a_2 are increasing in A . This is true unless the income effect is indeed zero; the very meaning of zero income effects is that all extra consumption is allocated to one good, the consumption good in our case. Further, the own price effects are the expected ones, i.e. $\partial a_i / \partial p_i < 0$. Perhaps less expectedly, the cross price effects are also negative. Once again, this reflects the inherent complementarity between information gathering and prevention; it depends, however, on the income effect being small (this fact is stressed by asterisks in Table 3.1 below). Finally, the effect of an increase in the price of consumption is to shift resources toward prevention as well as information gathering; this is subject to the qualifications that the income effect be small enough, and this is indicated by asterisks in Table 3.1 which gives a summary of income and price effects.

Large income effects

We will now turn to the case where income effects are large in the sense that they dominate the expressions marked with asterisks in the sign table of the two first columns of the inverse of the Jacobian matrix; the table for this case is

$$\begin{bmatrix} - & +* \\ +* & - \\ + & + \\ - & - \end{bmatrix}, \quad (3.10)$$

and we will now provide a set of statements about this case.

Consider first exogenous information, g_1 . The sign pattern of the right-hand side of the matrix equation is given by expression (3.6) and it is $(-/+ , -, 0, 0)$ for exogenous information and own information being complements/substitutes. This implies that for substitutes we have an unambiguous result for prevention and information gathering,

$$\frac{\partial a_1}{\partial g_1} < 0, \quad \frac{\partial a_2}{\partial g_1} > 0.$$

For complements, on the other hand, the only unambiguous implication is that consumption decreases. Consider now exogenous prevention, g_2 . The sign pattern



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

Table 3.1 Effects of income and price changes

	A	p_1	p_2	p_3
a_1	+ (0)	–	–*	+*
a_2	+ (0)	–*	–	+*
a_3	+	?	?	–

Table 3.2 Large income effects

	<i>Substitutes</i>		<i>Complements</i>	
	g_1	g_2	a_1	a_2
a_1	–	+	?	?
a_2	+	–	?	?
a_3	?	?	–	–

is $(-, -/+, 0, 0)$ for a_2 and g_2 being complements/substitutes. For the case of substitutes, the derivatives satisfy

$$\frac{\partial a_1}{\partial g_2} > 0, \quad \frac{\partial a_2}{\partial g_2} < 0.$$

The exogenous activity substitutes the own one, and resources are shifted towards information gathering. For complements, consumption decreases. These results are summarized in Table 3.2.

These results are intuitive. When income effects are large, a small change in consumption has a large effect on the marginal utility of consumption; if the individual increases consumption, the marginal utility of other activities must fall sharply, and vice versa. This amounts to the individual being *reluctant* to change his consumption, and as a consequence he is likely to substitute information gathering for prevention, or vice versa, in the face of exogenous impulses related to health.



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

This pattern is seen most clearly if the impact, g_i , is a substitute for the individual's corresponding own activity, a_i ; then there is an unambiguous case for substitution. If the activities are complementary, there are opposing forces, and there are no clear-cut conclusions for prevention and information gathering. However, since consumption decreases, it is clear that the total spending of resources on a_1 and a_2 combined increases; once again, this reflects the fundamental complementarity lurking in the background.

Now, let us turn to local changes in the health states, i.e. changes in h^t for some t . The sign pattern of the right-hand side derivatives is $(-, -/+, 0, 0)$ for a bottom-end/top-end preventive activity, and we see that for a top-end activity, we have

$$\frac{\partial a_1}{\partial h} > 0, \quad \frac{\partial a_2}{\partial h} < 0;$$

that is, it has similar effects as an exogenous preventive activity that is a substitute for own prevention. This is unsurprising in the light of our arguments, above that a top-end preventive activity is in effect, a substitute for favourable perturbations to the states. Indeed, for a bottom-end preventive activity, the pattern of the second half of Table 3.2 applies.

For price changes, the only change is that the effects on information gathering and prevention from changes in other prices become ambiguous; the details are set out by asterisks in Table 3.1 above.

Health policy and socio-economic differences in health and behaviour

Aspects of health policy in the model

In our analysis of exogenous shocks to the individual's decision, we have implicitly covered much of the effects of health policy. First, one important aspect of health policy is the attempt to directly influence individual health-related decisions. This is accomplished by (i) price changes of various goods, e.g. taxation of tobacco or subsidies to athletic associations (physical exercise); (ii) regulations, e.g. age-limits for the purchase of alcohol; and (iii) dissemination of information. Effects of price changes have been dealt with, and several aspects of the effects of new



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

information on the individual's informative and preventive activities have also been analysed. Note also that regulations can be seen to a certain extent as price changes (you can always choose not to comply, if you accept the expected penalty).

Second, the model allowed us to analyse the effects on individual behaviour of general preventive measures that can be treated as exogenous from the individual's point of view, such as workplace safety, product safety and environmental regulation.

Third, a common form of health policy is public support of R&D in the health area. The output from this process may change the individual's decision environment, as well as his basic attitudes. Several of the possible effects fall within the categories we have analysed above, but obviously research can also have effects that are difficult to capture in our model, such as the identification of a new disease.

Differential behaviour across socio-economic groups

In this section, we will explore the implications of our results for differences in behaviour across population groups as defined by health-prospects, age, income, and education. In all cases except when we discuss income, we will assume that income effects are small. Several of the differences in behaviour relate to differential reactions to health policy, which is our reason for treating them here.

It is readily apparent that different individuals face different probabilities for ill health; for example, because of genetic factors. In our context, a difference in health prospects across individuals can be modelled as a pure shift in F , i.e. without effects on the marginal effect on individual prevention ($\partial F/\partial a_2$ is unaffected). (It seems equally possible to argue that the level of health risks has a positive or a negative effect on the perceived marginal effect of prevention on health risks.) It is easily seen from expressions (3.5) and (3.7) above that someone with worse health prospects will spend less on prevention (and information gathering).

Ageing is a process whereby an individual's health prospects tend to deteriorate over time. Our results therefore suggest that the older you get, the less you will engage in prevention (and information gathering), unless this has at the same time a sufficiently strong positive impact on the marginal effect of your own preventive efforts. It is interesting to compare this result with the result in the demand-for-health literature, where ageing individuals may choose to increase their spending on medical care in order to counteract the increasingly rapid



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

depreciation of their stocks of health.¹⁹ The mechanism in our model is quite different; it is due to the presence of genuine uncertainty, where a deterioration of health prospects in the probabilistic world reduces the return to information gathering which in turn reduces the marginal return to prevention. Clearly, there are reasons for old people to invest less in prevention from a lifecycle perspective as well; one fundamental reason is that their payoff period is shorter (Cropper, 1977).²⁰

Health policy will also affect the individual's health prospects. We focused above on the case where the exogenous activity g_2 has a favourable effect on F . However, remembering that g_2 may represent *any* exogenous occurrence that shifts F , it is obvious that the opposite may happen. For example, a person may learn through public information that his health prospects are worse (instead of better) than he thought.

Consider the case where there are shifts in F as a consequence of a government action, g_2 , but where F shifts in different directions for different individuals; some receive good news about their health prospects and some receive bad news. One can either imagine that a particular piece of information affects different individuals differently, or that certain information only affects a part of the population. (For example, news that a storage of nuclear waste is moved from one locality to another, or news that those who have acquired artificial suntans in solariums have increased their risk of malignant melanoma.) *Ceteris paribus* – i.e. assuming that the marginal effect of individual preventive activity ($\partial F/\partial a_2$) is unaffected – it is clear from expressions (3.5) and (3.7) that those who receive bad news reduce their prevention (and information gathering), while the opposite is true for those who receive good news.

Proposition 3.7 *An exogenous unfavourable shift in F (worsening health prospects) will lead to a reduction in the individual's preventive activity a_2 , ceteris paribus, while a favourable shift will have the opposite effect. This implies that there is a natural tendency for health differentials in the population to widen, unless those who receive good news about their health tend to be in relatively bad health and vice versa.*

If we allow $\partial^2 F/\partial a_2 \partial g_2 \neq 0$, there are four configurations of signs (combinations of F and $\partial F/\partial a_2$ being increasing or decreasing in g_2), all of which are perfectly plausible since it is possible that the same information is a complement to their own prevention for some individuals while it is a substitute for others, for example because their level of health differs. In three out of four of these cases, health



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

differentials will tend to widen. (The exception is the case where those who receive bad news experience a positive shift in $\partial F/\partial a_2$ and those who receive good news experience a negative shift in $\partial F/\partial a_2$). In a much different model of life cycle investments in health under certainty, Ehrlich and Chuma (1990) also find that health differentials have a natural tendency to widen in the sense that a higher initial endowment of health increases the demand for health investment and attained longevity.

The implication in the second part of the proposition is interesting not least because social differences in health are a major policy concern in many countries;²¹ moreover, the results suggest that the problem is quite fundamental and that government policy might easily contribute to such differences.

This result is also in accordance with the results in the literature on complementarities. It has, for example, been shown that the evolution of modern manufacturing can be understood as driven by exogenous technological change, and that with complementarities, the result will be that modes of organization will exhibit *clustering* in the sense that technologies will evolve – in their model jump – in several dimensions simultaneously (Milgrom and Roberts, 1990).

It is worth making a short digression to consider the likely health status of those who receive good and bad news, even though considering the *ex ante* health status of individuals takes us outside the scope of our model. It is interesting to compare the effects of three different kinds of government action which all serve to shift F : (i) dissemination of information on existing health probabilities; (ii) dissemination of information on the effects of the individual's preventive behaviour; and (iii) preventive activities undertaken by the government. While all three of these government actions can be represented by our g_2 , their properties differ in the specific respect of who gets what kind of news. For concreteness, suppose the individual learns that: (i) 'contrary to previous beliefs, people with blue and brown eyes do not face identical health risks, rather, it has been discovered that all blue-eyed persons face low risk of ill health and all brown-eyed persons face high risks'; or (ii) 'while it was believed that eating carrots and eating broccoli were equally good for your health, we now know that eating carrots is more beneficial than previously thought but eating broccoli does not have any effect on your health'; or (iii) 'new and reduced speed limits for cars will reduce air pollution and this will reduce sickness in densely populated areas'.

In the first of these cases, it is those in good health who receive the good news,



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

and vice versa. Those with blue eyes are on average healthier than those with brown eyes (without previously having known why), and the information tends to increase health differentials. Similarly, assume in the second case a random *ex ante* distribution of the population into carrot and broccoli eaters. The carrot eaters receive good news and have on average better health than the broccoli eaters (because they have happened to have made the correct choice). The information tends to widen the health differentials. In the third case, however, we get the opposite tendency. A preventive action by the government will be beneficial to those who have been exposed to the health risk and who *ipso facto* are likely to be in relatively bad health compared to the unexposed part of the population, *ceteris paribus*. (The *ceteris paribus* condition is important here. It is of course possible that the government strives primarily to reduce those health risks that affect the healthier part of the population.)

The general conclusion is that the effect of a particular government action on social differences in health will depend on which specific form the action takes. In particular, information about the effect of individual preventive behaviour seems likely to widen health differentials whereas preventive activities seem likely to reduce them.

The fact that differences in health are related to differences in economic resources is a major concern of policy makers. There are two sources of such differences in our model. The first one is quite simply the size of the individual's budget. Since health does not affect the budget constraint in our model, we have, as noted above, a consumption model of health, and the effect of an expansion of the budget was to unambiguously increase prevention (unless income effects are zero). Hence wealthier people end up healthier, just as in the consumption part of Grossman's (1972) original demand-for-health model. (In stochastic investment models of health, on the other hand, the effect of an increase in wealth on the demand for health has been found to be ambiguous.²²)

In addition, however, we have seen that the size of income effects are potentially important for behaviour in our model. Moreover income effects are likely to vary across income groups. If we assume that the budget is of a purely monetary nature, we have a case where income effects are likely to be greater among low-income groups (given decreasing marginal utility of consumption and money). Now, we have seen that there is a basic complementarity between information gathering and prevention. We have also seen that as long as income effects are small, this



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

complementarity is only upset if the exogenous impact, g_i , is a sufficiently strong substitute for the corresponding own activity, a_i . This implies that as long as income effects are small, the effect of providing individuals with information or prevention will often be to increase their own efforts; people will 'match the contribution'. This effect is much less likely if income effects are large, the reason being that an individual with a large income effect is much less prone to give up consumption in the face of a more favourable payoff from caring about his health. (Similarly, in the case of unfavourable information, the general tendency is that information gathering and own prevention will fall, but to a lesser extent the larger the income effects.)

Hence if the budget constraint is of a mainly monetary character, this is likely to imply that rich people will take advantage of health policy by 'matching the contribution' to a larger extent than will poor people. Note that this does not mean that poor people do not take advantage of health policy. However, since they are poor, they are less likely to sacrifice consumption and thereby in a sense fail to exploit the complementarity fully. Correspondingly, several of our results seem roughly consistent with the common argument that high-income people are relatively more prone to change their health-related behaviour in response to health information. One may note, for example, that if public information is complementary with the own informative activity, this unambiguously increases prevention among high-income earners (small-income effects), whereas the effect is ambiguous for low-income earners (large-income effects). Moreover, favourable information on a specific health state is certain to increase bottom-end prevention among high-income people and to reduce top-end prevention among those with low incomes.

Education, finally, can of course have many different effects.²³ If, for example, education raises efficiency in the household production of health as in Grossman (1972), this would in our model have the same kind of effect as exogenous prevention; i.e. it would act as a complement to the individual's own prevention. However, in this context it seems particularly relevant to explore the possibility that it affects the individual's confidence in the information that he possesses and perhaps also the effects of new information on his level of confidence. If one believes that education makes an individual more confident about the information that he possesses (larger y), this implies that those with more education will engage more in prevention and information gathering. In general, it is obvious that the effect of g_1 on y is not necessarily positive; the individual's degree of



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

confidence does not necessarily increase as, e.g. new results of medical research are published. Furthermore, one could hypothesize that those with high education will become more confident by receiving public information while those with little education are likely to become less confident. If this happens to be the case, the information will cause increasing information differentials, *ceteris paribus*. (That is, assuming that $\partial y/\partial a_1$ is unaffected; if not, the situation is parallel to the one with health differentials: allowing for $\partial^2 y/\partial a_1 \partial g_1 \neq 0$ gives four cases, and in three of these differentials tend to widen.)

Concluding remarks

We have introduced uncertainty and uncertainty aversion into a simple model of health-related behaviour. We have thereby made a first attempt at filling a gap that has potentially important consequences. Indeed, some of the positive conclusions of our analysis – in particular the fundamental complementarity – indicate that the presence of uncertainty unveils forces that were previously hard to identify.

The model employed seems to capture several interesting features of these health-related decisions. The model allows the individual to experience an element of genuine uncertainty in the way he thinks about his health and also to reduce the importance of this element by information gathering. The individual may also engage in a preventive activity with respect to the health states that he knows something about, and conceptually this may also include consumption of medical care.

We have explored how individuals react to changes in their decision environment. These changes include exogenous information and prevention, such as health education campaigns or safety regulations undertaken by the government, as well as new knowledge about the properties of a particular state of ill health. While of course not being exhaustive, the analysis seems to cover a great deal of the possible effects of new information and health policy. The significance of some personality traits (confidence and pessimism) was also explored.

Perhaps the most interesting analytical result was the fundamental complementarity found between prevention and information gathering, and the implications that this may have for individual behaviour and the effects of health policy. For example, the presence of this mechanism suggests that there may be a



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

natural tendency for health differentials to widen, a tendency which may easily be exacerbated by health policy.

It seems to us that we have far from exhausted the potential for this line of inquiry. For example, we have analysed each activity and each exogenous influence separately, but obviously there will in practice often be interactions between different activities, and public policy will also have mixed effects. Our investigation provides some preliminary indications that analyses of more complex situations can produce interesting results. For instance, we have seen that exogenous prevention such as safety regulation may increase the individual's own preventive efforts. If, however, the public regulatory activity simultaneously makes the individual more diffident about the information he possesses, the result could be reversed.

For empirical applications of the theoretical framework presented above, the most promising avenue – and a reasonable first step – would be to investigate the relationship between the individual's confidence in his probabilistic information and health-related behaviour.²⁴ Results from a qualitative study suggest that such general attitudes towards genuine uncertainty are discernible. The general attitude seems particularly evident when a person, on the subject of health risk information, states that: 'It doesn't matter what you do. . . . The day after you will read that you shouldn't have done it. [. . .] I think that I used to trust things more before. Nowadays they seem to be finding just too much shit.'²⁵ In a qualitative setting, the individual's position in the confidence dimension can be compared with statements about behaviour. Furthermore, based on this kind of qualitative research, it seems possible in the future to insert questions on this confidence attitude in questionnaires on health behaviour and to make the concept amenable to quantitative research.²⁶ A challenge for the future is to find ways of varying the relative importance of genuine uncertainty in an experimental setting, so that the effect on hypothetical choices can be observed, paralleling the efforts to find individual parameter values for risk aversion and time preferences (Barsky *et al.*, 1997). It seems somewhat less obvious that one can elicit the individual's relative (health) pessimism, though a viable tentative hypothesis is that individuals in lower social positions are relatively more likely to feel at home in situations where probabilities are unknown (more favourably disposed to genuine uncertainty), as a corollary to the fact that they are less accustomed to making probabilistic choices.²⁷



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

Appendix

The welfare effects of an intervention, g , that affects utility directly as well as behaviour, is, generally,

$$\frac{\partial U}{\partial g} = \frac{\partial U}{\partial g} + \frac{\partial U}{\partial a_1} \frac{\partial a_1}{\partial g} + \frac{\partial U}{\partial a_2} \frac{\partial a_2}{\partial g} + \frac{\partial U}{\partial a_3} \frac{\partial a_3}{\partial g},$$

but since $\partial U/\partial a_i = \lambda p_i$, and $p_1 a_1 + p_2 a_2 + p_3 a_3$ is constant, the marginal welfare effect is given by the partial derivative $\partial U/\partial g$. It is thus clear that exogenous information – a favourable intervention affecting only γ – will be beneficial only for basically pessimistic individuals

$$\frac{\partial U}{\partial g} = \frac{\partial \gamma}{\partial g} \cdot \left(\sum \Delta u^s (1 - F) - u_0 \right),$$

while prevention – a favourable intervention affecting only F – is always welfare improving,

$$\frac{\partial U}{\partial g} = \left(\gamma \cdot \sum \Delta u^s \frac{-\partial F}{\partial g} \right).$$

Notes

Financial support from the Swedish Medical Research Council (grant no. 27P-10737) and the Bank of Sweden Tercentenary Foundation is gratefully acknowledged. The paper was presented at the 19th Arne Ryde symposium on Individual Decisions for Health in Lund, August 1999, and we are most grateful for insightful comments and suggestions from our discussant, Louis Eeckhoudt. We also appreciate comments from Tomas Philipson and the audience at the seminar. In addition, we have received helpful comments from Dan Anderberg, John Hey,



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

Håkan J. Holm, Bengt Liljas, Björn Lindgren, and two anonymous referees. Any remaining errors are the sole responsibility of the authors.

1. Examples include Chang (1996), Cropper (1977), Dardanoni and Wagstaff (1987, 1990), Liljas (1998), Picone *et al.* (1998), and Selden (1993).
2. Significant contributions with deterministic models are still appearing in the demandfor- health tradition (Grossman, 1998; Ried, 1998), as well as in other areas.
3. In fact, Eichberger and Kelsey (1999) have provided axiomatic foundations for a utility function which is similar to (3.1); i.e. which is a convex combination of an expected utility and a generalized expected utility.
4. See, e.g. Schmeidler (1989). The notion of genuine uncertainty was introduced by Knight (1921); it is sometimes referred to as 'Knightian uncertainty'.
5. The generalized probabilities generate weights, v_s , by means of which the modified expectation is computed: $u_0(a) = \sum v_s u(h^s, a)$.
6. A framework employing a utility function similar to ours is employed by Mukerji (1998) in an exploration of the effect of uncertainty on contractual incompleteness.
7. Note that it may be the case that the outcomes (h) of a firm state and an uncertain state coincide; increasing y then leads to a change in the expectation only due to better knowledge of the probability of this outcome.
8. Clearly, there are cases where information gathering reduces variance rather than uncertainty; we do not consider such information. We will comment on information in the form of 'more informative experiments' in Pessimism and confidence: basic analysis.
9. In the context of understanding the choice between systematic and rule-of-thumb (heuristic) information processing, it has been suggested that the motivation for expending efforts in systematic processing is to attain a certain level of confidence that one's knowledge about risks is sufficient, based on a desire to judge situations correctly. Cf. Griffin *et al.* (1999).
10. See, e.g. Mas-Colell *et al.* (1995).
11. These activities could also be thought of as medical care in the sense that the essence of medical treatment is to improve the probability distribution with respect to future health states. However, health-production



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

possibilities could be dependent on whether the individual is healthy or sick (Picone *et al.*, 1998; Zweifel and Breyer, 1997). Cf. Kenkel (2000), Sec. 2, for an overview of prevention in health economics models of individual health decisions.

12. See, e.g. Hirshleifer and Riley (1992, Sec. 5.2.3).
13. More generally, income effects are small if the second derivatives involving a_3 are small relative to the bordered Hessian of the utility-maximisation problem; see, e.g. Varian (1992, pp. 123–4).
14. Dardanoni and Wagstaff (1990), Liljas (1998), Lyttkens (1992) and Picone *et al.* (1998). In addition, some activities which we tend to think of as ‘preventive’ may in fact be undertaken in order to reduce the relative weight of genuine uncertainty, and not to affect probabilities. Such an activity would in our model by definition be termed an informational activity, just as anything that serves to shift the individual’s F by definition is termed prevention (which may include public information on health risks. Furthermore, if we leave the present framework, there can be other reasons to engage in ‘preventive’ behaviour. For example, the individual may engage in an activity without any known health consequences in order not to experience regret at some future date when such a relationship may have been discovered (cf., e.g. Loomes and Sugden (1982) on regret theory).
15. Louis Eeckhoudt provided an example along the following lines: an early diagnosis makes it possible to tailor the treatment decision to the actual health state. This possibility, however, undermines incentives for *ex ante* prevention. In such a case, acquiring information in the sense of obtaining the diagnosis reduces incentives for *ex ante* prevention. For a model relating to this example and dealing with the effect of information on environmental precaution, see Gollier *et al.* (2000).
16. We are grateful to Louis Eeckhout for calling our attention to the latter fact.
17. Note that the bottom-end property defines a class of preventive activities for which we can make clear predictions about the effect of very specific changes in the individual’s environment.
18. There is an intriguing relationship between the bottom-end and top-end properties and the *Monotone Probability Ratio* order (MPR) introduced by Eeckhoudt and Gollier (1995). This order requires that ratios between values of the distribution function (for two different actions) satisfy a



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

- monotonicity property similar to the one imposed here on the derivative (and thus on differences). Eeckhoudt and Gollier explore the implications of MPR for the demand for a risky asset by a risk-averse individual.
19. Grossman (1972), Muurinen (1982).
 20. Kenkel (1994) provides some empirical support for the notion that investments in prevention decline with age.
 21. Health 21 (1999); Healthy People 2000 (1990); Healthy People 2010 (2000); Hälsa på lika villkor (1999).
 22. Chang (1996), Dardanoni and Wagstaff (1987), Selden (1993).
 23. Cf. Kenkel (2000), pp. 1680–82, for a brief overview of the relationship between schooling and health.
 24. At present, it seems less useful to focus on exogenous shocks for hypothesis testing, both because the direction of influences is often open to question (e.g. information may either increase or decrease confidence) and because exogenous events are likely to have more mixed effects than we have assumed (e.g. a preventive activity may affect the degree of information confidence). Hence testing based on exogenous shocks will have to await more empirical information about the partial relationships (and possibly more theoretical work).
 25. Lindbladh and Lyttkens (2000). Similarly, a tendency to perceive the health-related part of the world in terms closely akin to genuine uncertainty is suggested by the following reaction to posters about the health risks associated with smoking in a waiting room at the primary health care centre: 'I can be knocked down by a car when I cross the street, I can get other diseases if I do not smoke, so I don't think much about such things'. Note that the respondent in question is not arguing that the risk associated with smoking is *small*. He or she rather seems to express a fatalistic attitude, implying that there is a multitude of unknown risks and that therefore no point in trying to influence them.
 26. The relationship between trust in information sources, regulatory agencies, etc. and risk perceptions has been investigated in a number of studies. Cf., e.g. Grobe *et al.* (1999) and Slovic (1999).
 27. Lindbladh and Lyttkens (2000).

References



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

- Arrow, K., 1963. Uncertainty and the welfare economics of medical care, *American Economic Review* 53, 941–73.
- Barsky, R.B. *et al.*, 1997. Preference parameters and behavioural heterogeneity: an experimental approach in the health and retirement study, *Quarterly Journal of Economics* 112, 537–79.
- Chang, F.R., 1996. Uncertainty and investment in health, *Journal of Health Economics* 15, 369–76.
- Cropper, M.L., 1977. Health, investment in health, and occupational choice, *Journal of Political Economy* 86, 1273–94.
- Dardanoni, V. and Wagstaff, A., 1987. Uncertainty, inequalities in health and the demand for health, *Journal of Health Economics* 6, 283–90.
- Dardanoni, V. and Wagstaff, A., 1990. Uncertainty and the demand for medical care. *Journal of Health Economics* 9, 23–38.
- Eeckhoudt, L. and Gollier, C., 1995. Demand for risky assets and the monotone probability ratio order, *Journal of Risk and Uncertainty* 11, 113–22.
- Ehrlich, I. and Chuma, H., 1990. A model of the demand for longevity and the value of life extensions, *Journal of Political Economy* 98, 761–82.
- Eichberger, J. and Kelsey, D., 1999. E-capacities and the Ellsberg paradox, *Theory and Decision* 39, 107–40.
- Gollier, C., Jullien, B. and Treich, N., 2000. Scientific progress and irreversibility: an economic interpretation of the 'Precautionary Principle', *Journal of Public Economics* 75, 229–53.
- Griffin, R.J., Dunwoody, S. and Neuwirth, K., 1999. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviours, *Environmental Research* 80, S230–45.
- Grobe, D., Douthitt, R. and Zepeda, L., 1999. A model of consumers' risk perceptions toward recombinant bovine growth hormone (rbGH): the impact of risk characteristics, *Risk Analysis* 19, 661–73.
- Grossman, M., 1972. On the concept of health capital and the demand for health, *Journal of Political Economy* 80, 223–55.
- Grossman, M., 1998. On optimal length of life, *Journal of Health Economics* 17,



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

499–509.

Hälsa på lika villkor, 1999. *Delbetänkande av Nationella folkhälsokommittén*, SOU 37.

Health 21, 1999, *The health for all policy framework for the WHO European Region*.
WHO, Regional Office for Europe, Copenhagen.

Healthy People 2000, 1990. (US Department of Health and Human Services,
Washington, DC).

Healthy People 2010, 2000. (US Department of Health and Human Services,
Washington, DC).

Hirshleifer, J. and Riley, J., 1992. *The Analytics of Uncertainty and Information*.
Cambridge University Press, Cambridge.

Johansson, P.-O., 1994. Valuing changes in health: a production function approach,
In: R. Pethig, ed., *Valuing the Environment: Methodological and Measurement
Issues*. Kluwer Academic Publishers, Dordrecht, pp. 151–67.

Kenkel, D.S., 1994. The demand for preventive medical care, *Applied Economics* 26,
313–25.

Kenkel, D.S., 2000. Prevention, In: A.J. Culyer and J.P. Newhouse, eds., *Handbook of
Health Economics*. Elsevier, Amsterdam, Vol. 1B, pp. 1675–720.

Knight, F., 1921. *Risk, Uncertainty, and Profit*. Houghton Mifflin, Boston.

Liljas, B., 1998. The demand for health with uncertainty and insurance, *Journal of
Health Economics* 17, 153–70.

Lindbladh, E. and Lyttkens, C.H., 2000. Perceptions of risk information. A perspective
on social differences in health-related behaviour, *Studies in Health Economics*
33, Department of Community Medicine, Lund University.

Loomes, G. and Sugden, R., 1982. Regret theory: an alternative theory of rational
choice under uncertainty, *Economic Journal* 92, 805–24.

Lyttkens, C.H., 1992. A note on the economics of prevention. Individual behaviour
and preventive information, *Actes du Colloque Européen 'De l'analyse économique
aux politiques de santé'* Credes-Ces, Paris.

Mas-Colell, A., Whinston, M. and Green, J., 1995. *Microeconomic Theory*. Oxford
University Press, New York.

Milgrom, P. and Roberts, J., 1990. The economics of modern manufacturing:



HEALTH, GENUINE UNCERTAINTY, AND INFORMATION

BY FREDRIK ANDERSSON AND CARL HAMPUS LYTTKENS

Excerpted from *Individual Decisions for Health*

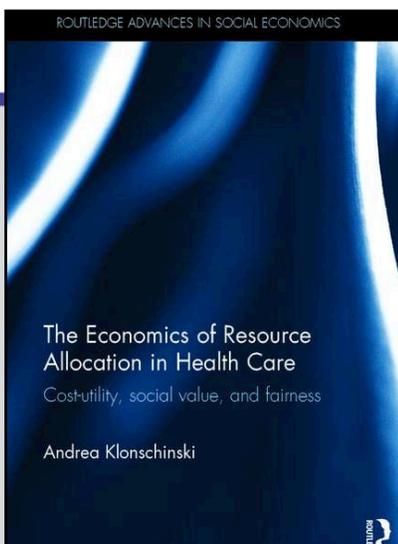
- technology, strategy, and organization, *American Economic Review* 80, 511–28.
- Mukerji, S., 1998. Ambiguity aversion and incompleteness of contractual form, *American Economic Review* 88, 1207–31.
- Muurinen, J.M., 1982. Demand for health: a generalized Grossman model, *Journal of Health Economics* 1, 5–28.
- Picone, G., Uribe, M. and Wilson, M.R., 1998. The effect of uncertainty on the demand for medical care, health capital and wealth, *Journal of Health Economics* 17, 171–85.
- Ried, W., 1998. Comparative dynamics of the full Grossman model, *Journal of Health Economics* 17, 383–425.
- Schmeidler, D., 1989. Subjective probability and expected utility without additivity, *Econometrica* 57, 571–87.
- Selden, T.S., 1993. Uncertainty and health care spending by the poor, *Journal of Health Economics* 12, 109–15.
- Simon, C.P. and Blume, L., 1994. *Mathematics for Economists*. Norton, New York, pp. 529–30.
- Slovic, P., 1999. Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield, *Risk Analysis* 19, 689–701.
- Varian, H.R., 1992. *Microeconomic Analysis*. Norton, New York, pp. 123–4.
- Zweifel, P. and Breyer, F., 1997. *Health Economics*. Oxford University Press, New York, Ch. 3.



CHAPTER

3

ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES



This chapter is excerpted from

*The Economics of Resource Allocation in Health Care:
Cost-Utility, social value, and fairness*

by Andrea Klonschinski

© 2016 Taylor & Francis Group. All rights reserved.



[Learn more](#)



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

7.1 Introduction

Having made a case against the suitability of a value-maximizing framework to deal with fairness issues in health care resource allocation in a satisfying manner, the present chapter challenges the weighting approach from another angle and asks whether empirically elicited prioritization preferences should be assigned any normative-ethical relevance in the academic discourse on medical resource allocation at all. In order to analyze this query, two major lines of arguments in favor of public participation offered in the literature are discussed: first, the idea that by means of correcting for biases in or filling gaps of the theoretical discourse, public involvement in priority setting somehow improves the prioritization debate and, second, the notion that in a liberal and democratic society, respecting individual autonomy requires involving public preferences on a theoretical level. The chapter argues that both threads basically rely on category errors as to the nature of the issue at stake. Consequently, even if the arguments presented up to now were wrong and equity weights indeed accurately mirrored the respondents' concerns for fairness, the weighting approach still would not constitute a promising attempt to solving CUA's fairness problem. Before establishing this claim, however, two preliminary remarks are required.

First, note that the present interest is decidedly not in pragmatic political questions such as how certain policies can be communicated or implemented, how a societal discourse on prioritization can be triggered, how participative methods could "build a habit of active citizenship" (Lengahan 1999: 47), or how the topic of prioritization can be set on the public agenda. In addition, the following considerations certainly are not meant to deny that the public should have a say in the democratic political process when it comes to the actual implementation of health policy. Therefore, it is again important to recognize that the terms participation and deliberation refer to a method of preference elicitation and are not regarded as an element of the political process within deliberative democracy (see Friedrich *et al.* 2012: 415).¹ Instead, the analysis focuses on the question of whether the empirical elicitation of the public's prioritization preferences – i.e., their opinion on what a just distribution of resources consists of – can contribute to the respective normative-ethical academic debate. In so far as the public's views on justice are supposed to be relevant for the theoretical debate on a just resource allocation, the endeavor to integrate public preferences into theory building can be subsumed under the heading of empirical ethics. The second remark concerns the subject matter of this chapter. Although the primary interest is with the weighting



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

approaches, which assign a particularly strong normative role to elicited preferences, this chapter also considers arguments from the broader discourse on public participation in priority setting. The reason for this is twofold: for one thing, the literature on equity weighting is rather moot as to the normative-ethical legitimization of their empirical account; for another, rebutting more modest arguments for consulting the public serves to reinforce the case against equity weighting.

7.2 Two alleged dead-ends of the normative-theoretical discourse

The first line of arguments for involving the public in the discourse on prioritization expresses the hope that this endeavor provides for a way out of the alleged dead-end of the academic prioritization discourse. Thereby, two different kinds of dead-ends can be distinguished, one caused by methodological limitations, the other by a lack of consent when it comes to theoretical arguments. The first impasse is frequently invoked by economists, as was already pointed out above. Due to the discipline's positivistic self-image, economists are usually disinclined to introduce their own value judgments into economic analysis. Dealing only with facts and lacking any scientific means to tackle normative questions, they leave normative-ethical issues to others. This stance is illustrated by the explanation Williams *et al.* (2005: 65) offer for the economists' habitual neglect of concerns for fairness:

The most fundamental [justification] is a denial that economics has any tools to handle such issues, since its current mainstream corpus of knowledge derives from a position in which interpersonal comparisons of welfare are held to be invalid and so are ruled out of consideration. But those willing and able to emancipate themselves from this strict welfarist regime still face severe problems in addressing issues of equity, because equity is an essentially contestable concept in which many rival views flourish.²

The quote's last sentence hints toward the second alleged blind alley of the prioritization discourse. Apparently, scholars of normative disciplines, such as political theory or philosophy, are incapable of reaching agreement as to what a just resource allocation consists of (see Williams/Cookson 2006: 3). In view of the irreversible ethical and moral pluralism in modern societies, there remains no device left to establish the superiority of one account over another. Since priorities have to be established nonetheless, public involvement is supposed to offer a way



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

out of this impasse. This idea seems to underlie Ubel's (2001: 94) following argument:

I am not a proponent of any single theory of distributive justice, which I cannot explain by lack of time or inclination, but can blame only on my indecisive nature or my intellectual weakness. Perhaps because I do not advocate a specific theory I am inclined to take public attitudes seriously. Many allocation dilemmas have no simple solutions, and highly trained, intellectual rigorous philosophers would completely disagree with each other about the best solution, for example, about the extent to which severely ill patients deserve treatment priority. In such situations, the public deserves a role.³

Two aspects are especially noteworthy. First, while Ubel states that he does not endorse a particular theory of distributive justice, it became apparent in the previous chapters that he actually adopts a quite specific account, namely, a consequentialist one (see Ubel *et al.* 1996b: 115, 2000c: 130). His claim to the contrary reinforces the impression that the value-oriented paradigm's normative content is not sufficiently reflected in the health economic literature. Instead, the structure of economic evaluation is viewed as a mere formal vehicle, allowing for the adoption of any substantive theory of distributive justice whatsoever. Second, Ubel indeed seems to attach normative priority to the arguments of "highly trained philosophers" and, concomitantly, ethical theory. When theory runs out, the public should have a say. In this respect, Walker and Siegel (2002: 267) observe: "While proponents of the use of SVPs [social value preferences] in CEA would no doubt decry the creation of philosopher kings, their own arguments may ultimately provide the crowning." On the whole, the quote evokes the impression that consulting the public offers an easy way of solving controversial issues and allows for circumventing the irreducible value pluralism. Eliciting prioritization preferences would thus provide the "key to resolve ethical dilemmas" (Williams *et al.* 2012: 27).⁴

If the elicited public preferences are to fulfill this purpose, they have to be ultimately decisive when it comes to the allocation of health care resources. In other words, the arguments presented imply a form of ethical relativism according to which what is right is defined by what the majority considers to be right (see Hausman 2002: 642f.). And yet, this account is hardly plausible (see Bognar/Hirose



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

2014: 21). In fact, it is generally – implicitly or explicitly – acknowledged even in the empirical literature that not all kinds of prioritization preferences should influence priority setting to an equal extent. To give an example, Menzel et al. (1999: 8f.) argue that population preferences should only be taken into account under the condition “that those preferences are not irrational or ethically objectionable, as when they reflect discriminatory attitudes” and according to Richardson (2002: 632), “there may be a class of population values – hopefully small – where it would be hoped that decision-makers would override population values entirely (racism, sexism etc.)”⁵ Since “preferences sometimes seem ‘dirty’” (Goodin 1986: 76), they have to be laundered before they come to exert any influence on policy making. One way of laundering preferences is by “responding selectively to only certain sorts of citizen preferences,” that is, by excluding some preferences from the survey (Goodin 1986: 96). Note that the problem of dealing with “dirty,” i.e. ethically objectionable preferences, is not merely a hypothetical issue. To give an example, a study by Ubel *et al.* (1999a) revealed that respondents were less willing to allocate organs to intravenous drug users than to smokers or persons consuming high-fat diets. This preference pattern remained stable even if the subjects were informed that the drug users had stopped using and that the drug consumption was not causally responsible for the need of the transplant. The authors draw the conclusion that the respondents “believe that such patients are simply less worthy of scarce transplantable organs” (Ubel *et al.* 1999a: 58).⁶ Apparently, they want to “punish’ individuals for their ‘socially undesirable’ behaviour” (Gaertner/Schokkaert 2012: 162). Likewise, Mortimer (2005: 166) quotes a study on prioritization preferences in which smokers, heavy drinkers, and homosexuals received very low priority and concludes that “it is difficult to think of a motivation other than personal prejudice for some of these preferences.” Such findings reveal that the “answers of the respondents may be influenced by psychological mechanisms with a rather doubtful ethical legitimacy” and, therefore, challenge the normative relevance of questionnaire studies altogether (Gaertner/Schokkaert 2012: 164).⁷

The crucial point is that once it is granted that prioritization preferences can be ethically objectionable, the preferences obviously cannot provide the benchmark for what is to count as ethically objectionable in the first place (see Walker/Siegel 2002: 267). Hence, the question arises as to why and when exactly social preferences should be considered at all. To quote Walker and Siegel (*ibid.*):

If we cannot justify an appeal to SVPs [social value preferences]



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

without first establishing that they are not ethically objectionable, then SVPs will be superfluous to the extent that we can resolve resource allocation issues through the application of ethical principles.

Instead of laundering preferences *ad hoc* on the basis of diffuse fairness intuitions, the reasons for excluding some preferences and counting others need to be spelled out systematically beforehand. The existence of preference laundering in the prioritization studies thus leads to two conclusions. First, phrases such as those quoted above indicate that the normative status of empirically elicited preferences is not sufficiently reflected and remains ambiguous in the literature. Second, preference laundering hints toward the inevitability of normative-theoretical reflections on the issue of which preferences are relevant in the first place (see Powers/Faden 2006: 183f.; Düwell 2009: 202, 208).

In view of the rationales provided for turning to empirical ethics, it could be objected that postulating more elaborate theoretical considerations immediately runs into the second impasse: as different accounts of justice prove irreconcilable with each other, consulting the public is the last resort. In reply to this case, four counterarguments can be put forward. First, if it is really the case that two equally just allocation options are available and consensus cannot be reached as to which one is preferable, the public may indeed function as a “tie-breaker” (Hausman 2002: 643; see Düwell 2009: 206). And yet, such cases do not constitute “ethical dilemmas” because if both options are equally just, there is no moral issue at stake at all and turning to public preferences does not fall into the category of empirical ethics. Hence, this argument “does not justify respecting population values in cases which do raise moral questions, where population values may conflict with what is morally permissible” (Hausman 2002: 643f.). Second, the existence of different, in part diametrically opposed accounts of what a just resource allocation consist of does not provide a *normative* argument for the relevance of the public’s prioritization preferences. At the utmost, it offers a pragmatic rationale. Third, although the quote by Ubel (2001: 94) presented above suggests otherwise, the juxtaposition of a normative-philosophical account to the issue of prioritization on the one hand and a socio-empirical account on the other is erroneous since the turn to prioritization preferences is not normatively neutral either. Assigning normative-ethical relevance to the public’s opinion on rationing decisions requires a substantive theoretical justification itself and so does the decision to exclude some preferences from the survey assessment, as the considerations on preference



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

laundrying pointed out already (see Düwell 2009: 208; Gaertner/Schokkaert 2012: 14).

Against the background of the previous reflections, the very premise of the argument remains to be challenged, for it is doubtful whether the limits of normative-theoretical reflections are so rapidly reached as assumed by the proponents of empirical ethics. Quite to the contrary, the present study demonstrated that the current debate on equity weights and the equity-efficiency trade-off in health care resource allocation is by no means stretched to its theoretical limits, but is rather characterized by too little theoretical and conceptual deliberation. As long as the central concepts and normative assumptions are not clarified on a theoretical level, the unresolved issues will reappear again in the course of the deliberative exercise. This claim will be buttressed by means of an example taken from Daniels (2008: 291ff.). In 2003, the Mexican government passed a law establishing a national health reform, the *Seguro Popular*, in order to offer health insurance to half of Mexico's population. The crucial question was which conditions ought to be covered in the insurance scheme. Therefore, Daniels conducted some workshops on decision-making in health care with key personnel from the central institutions. During the first workshops, it was suggested that four groups should be involved in the decision-making process:

A clinical group would provide important information about the clinical course of the disease and the effectiveness of treatments. An economic group would provide information on the costs and cost-effectiveness of the treatments. Since conformance with ethical norms and social acceptability were *also* criteria mentioned in the law, an ethics working group, [...] and a social acceptability working group would contribute to the deliberative process. (Daniels 2008: 293)⁸

Initially, each of the groups was supposed to rate the different conditions on an ordinal scale from 0 to 5 according to some self-selected criteria. Yet, it turned out that “the two groups assessing values, the ethics group and the social acceptability group” could not quantify their deliberations' results in the required manner (Daniels 2008: 294). Hence, the process was amended to the extent that the clinical and economics group would evaluate the interventions on the basis of criteria such as prevalence, seriousness, and cost-effectiveness, whereas the “ethics



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

group would then base its analysis of the ethical issues on this information” (ibid.). Finally, their results would be assessed by the social acceptability group.

In the context of the present study, it is especially striking that this decision-process clearly distinguishes between economic concerns on the one hand and ethical concerns on the other. While the ethics group is supposed to assess “values,” the economics group’s task is presented as a purely positive analysis. This endeavor tacitly assumes at least two highly questionable premises: first, that cost-effectiveness is of independent normative relevance and, second, that this criterion can be adjusted by considering other concerns for fairness. It becomes obvious that the conceptual problems arising in a theoretical discussion of what a substantial account of justice requires in priority setting are carried over to the procedural account of deliberative decision making. As a consequence, the latter does not provide a solution to the alleged “pervasive disagreements,” but merely reiterates the underlying problems on another level. To make it worse, delegating the allegedly separate concerns of efficiency and fairness to different deliberative groups does not foster the required conceptual inquiry, but is likely to reinforce the unreflected persistence on the respective framework in the different groups, instead. As it turns out, “philosophical guidance of considerable moral weight does not run out quite as readily as some seem to suggest” (Powers/Faden 2006: 181). The exhaustive and troublesome clarification of concepts and the reflection of arguments about the best theory of justice cannot be avoided, especially not by means of empirical surveys.⁹

7.3 The role of empirical data within applied ethics

While the previous arguments in favor of empirical ethics pointed toward a methodological or theoretical impasse of the scientific prioritization debate, a second line of reasoning suggests that ethical theory is biased, incomplete, and too abstract. Therefore, involving the public on the level of theoretical reflections already is required in order to correct for the biases and to enhance the context-sensitivity of ethics by bringing in new ideas and perspectives. To give an example, Schicktanz (2009: 228) argues that due to cognitive, temporal, and spatial constraints, a person reflecting on a subject never comes up with all arguments and aspects relevant for the issue at stake.¹⁰ This shortcoming of individual reasoning cannot be resolved by academic discussions either, because the scholars involved in these debates usually share a common background and a similar living situation, so that they cannot put themselves into the shoes of persons in totally



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

different cultural or socio-economic circumstances. In effect, they become blind to certain arguments and considerations (see Düwell 2009: 206; Schicktanz 2009: 229). The problem is aggravated when it comes to questions of fairness because empirical research shows that individuals' opinions on justice are closely linked to their personal characteristics (see Gaertner/Schokkaert 2012: 10). That is to say, the typical male, white academic earning a mean income will have a very specific view on justice which is unlikely to coincide with the views held by other individuals with different personal traits. To this respect, it is argued that public participation has the potential of correcting biases by bringing in "new ideas and experiences" (Lengahan 1999: 47) and by offering new arguments, rationales, or solutions (see Bruni *et al.* 2008: 15; Gaertner/Schokkaert 2012: 11; Stumpf/Raspe 2012: 419; Frith *et al.* 2014: 18; Stumpf 2014a: 17f.). It is supposed to increase the "intellectual quality" (Schicktanz *et al.* 2012: 137) of ethical decisions and to counteract the creation of an expertocracy, in which a small circle of experts makes decisions on prioritization in a paternalistic fashion (see Stumpf/Raspe 2011: 316).

In a similar vein, some authors argue that ethical theory needs to be supplemented with empirical ethics since it is far too theoretical and out of touch with reality. Ethical theories, Richardson (2002: 637) objects, fail "to bridge the gap between theory and the requirements of practical action" and hence cannot provide concrete advice when it comes to the real-world issue of medical resource allocation.¹¹ Since professional ethics should not be divorced "from the world in which people are living and making ethical choices" (Levitt 2003: 24), investigating public preferences is required for improving the context-sensitivity of armchair-ethics (see Birnbacher 1999: 320f.; Musschenga 2005: 473ff.; Schicktanz 2009: 225; Schicktanz *et al.* 2012). According to this view, ethicists "should not limit themselves to formulating abstract and general principles" (Musschenga 2005: 473). Instead of restricting ethical inquiry to the theoretical investigation of a proposed ethical rule, ethicists have to "consider its practical feasibility, its psychological acceptability and its potential effectiveness in changing attitudes and behaviour in the desired direction" (Birnbacher 1999: 321). In order to make sure that the directives derived from ethical theory "connect with people's own feelings, attitudes and experiences" (Levitt 2003: 24) public preferences have to be accounted for on the level of normative-ethical reflections already.

While these arguments to some extent indeed provide a case for incorporating empirical data into ethical theory, they do not establish the need for empirical



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

ethics as the term is understood here. With the purpose of illustrating this claim, two kinds of empirical data have to be distinguished.¹² Using the example of euthanasia it can be said that any thorough reflection of whether euthanasia is morally acceptable has to rely on empirical data, of course. It may, for instance, be relevant how many and which type of patients are uttering the wish to have their lives ended and for which reasons they are doing so. Furthermore, an in-depth theoretical debate on the subject should acknowledge the possibilities of palliative care and pain therapy. Such information may only be accessible by asking the patients concerned, their relatives, or the physicians dealing with fatally ill patients on a day-to-day basis. Socio-empirical research is thus indispensable in so far as it helps to integrate the specific knowledge and the particular experiences of the persons concerned into ethical theory (see Düwell 2009: 204f.). Beyond that, there are philosophical arguments inherently relying on empirical data, a good example being slippery-slope arguments.¹³ These arguments reject certain practices not on the ground that they are regarded as morally objectionable per se, but because they are likely to trigger other actions with ethically obnoxious consequences. In the case of euthanasia, a slippery-slope argument could point out that although the practice as such is morally acceptable, it should not be allowed since, in the long run, it would erode the general prohibition to kill. The validity of the argument certainly hinges on the probability that the dreaded consequences will actually occur. When it comes to euthanasia, information about the practice and its consequences in countries where it is legally allowed can buttress or invalidate the slippery-slope argument. Again, this information gathering might imply asking the persons concerned. Generally speaking, empirical studies are of the utmost importance for ethics if an ethical evaluation of a certain practice requires knowledge only obtainable by asking the persons involved in that practice (see Düwell 2009: 204).¹⁴ If empirical ethics is interpreted in this sense as empirically informed ethics, the term describes thoroughly conducted applied ethics and does not raise any controversy at all (see Goldenberg 2005: 4; Borry *et al.* 2005: 50).

The situation is entirely different when the empirical data in question are information on the moral judgments of the persons concerned, though. Using the example of euthanasia again, this kind of data would be gathered if the persons were asked whether they regarded euthanasia as morally acceptable or not (see Düwell 2009: 204). As the results of such surveys tell something about what the people think about the legitimacy of euthanasia, they fall into the realm of



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

descriptive ethics, i.e., the branch of ethics aiming at describing the moral attitudes prevailing in a certain society or community. And yet, knowing what the people regard as just does not necessarily contribute to answering the question of what *is* just, that is, to the subject matter of *normative ethics*. Eliciting prioritization preferences, to be sure, is an endeavor in descriptive ethics, whereas allocating resources on the basis of these preferences assigns them a normative role. To quote Powers and Faden (2006: 185), equity weighting constitutes an attempt “to empiricize concerns about justice by turning these concerns into data about the public’s distributive preferences.” Recall, however, that some empirical studies on prioritization revealed discriminatory attitudes toward homosexuals and drug-users. Apparently, whether there are good moral arguments for a certain practice does not depend on what the majority regards as morally right (see Düwell 2009: 205f.). To put it pointedly, ethical questions can in no way be replaced by sociological questions (see Hausman 2000a: 38; Goldenberg 2005: 6).

Against the background of the distinction between these two fundamentally different types of empirical data, it becomes clear that the alleged lack of context-sensitivity of ethical theories does not provide an argument for empirical ethics. In fact, the dichotomy of context-sensitive empirical ethics on the one hand and armchair-ethics in the ivory tower on the other rests on a distorted image of (applied) ethics to begin with. The critics are actually “offering a ‘straw man’ account of applied ethics where absolutely no empirical considerations are included in the deductive process of ethical deliberation” (Goldenberg 2005: 3). To repeat it, empirical information on the subject matter of ethical reflection is a *sine qua non* of applied ethics. Depending on the topic, this information can include knowledge on the living situation and the needs of, say, chronically ill persons with a particular socio-cultural background.

The question of whether applied ethics should include considerations of feasibility on the level of theoretical deliberation already is another issue yet.¹⁵ Although discussing this issue at length is beyond the scope of this study, it should be noted that, for instance, the human rights or the animal ethics movement have enforced moral innovations against the then dominant *Zeitgeist* (see Birnbacher 2002: 104). To this effect, it can be argued with Unger (1996: 10ff.) that the task of moral theory is not to encompass people’s views on justice, but to liberate people from their false convictions or misguided intuitions instead. That being said, as long as prioritization decisions are within the realm of the values of western liberal societies and are not conflicting “too flagrantly with the settled convictions,” it is



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

highly unlikely that they will be rejected by the public anyway (Hausman 2002: 645). As the subsequent section will show, respondents in citizen juries commonly refer to very broad ideals and values such as equality, human dignity, and solidarity. This is not surprising because, as pointed out above, most people probably have never thought about rationing decisions before so that they cannot be expected to have specific and strong opinions or even intuitions as to intricate issues of medical resource allocation. Therefore, the feasibility of a thoroughly developed and critically discussed rationing plan does not present a problem.

Finally, it remains to be emphasized that the term armchair-ethics invokes a seriously misleading image of a scholar sitting alone in his study day in, day out and thinking things through. If ethics, or any science for that matter, was conducted like this, the objection of epistemological skepticism would indeed be sound. And yet, both sciences and humanities are of a genuinely social character (see Goldenberg 2005). By taking part in conferences, engaging in research groups, and publishing results, each scholar participates in the professional and, at least to some extent, the broader public discourse. In doing so, he permanently makes his arguments and intuitions accessible to criticism. This is especially true for the realm of applied ethics, which *per se* requires interdisciplinary work and, thanks to its applied subject, is necessarily embedded in the socio-political discourse anyway (see Düwell 2000: 87; Düwell/Steigleder 2003: 31f.; Schicktanz 2009: 225). Nevertheless, due to the fact that this discourse is basically restricted to experts, the question indeed arises as to whether consulting the public on normative-ethical questions could enrich the academic debate on prioritization by bringing in new aspects, rationales, and perspectives. This query is tackled in the following section.

7.4 On ethical expertise

The hope that public deliberation exercises reveal new arguments, viewpoints, or criteria that did not come up in the academic discourse before is quite optimistic, to say the least. Consider the example of a citizens' conference conducted in Lübeck, a small German town (see Stumpf/Raspe 2011, 2012).¹⁶ In 2010, a group of nineteen citizens met for four weekends to discuss which criteria and principles should guide priority setting in health care.¹⁷ In doing so, they were provided with information material on the subject matter and had the opportunity to consult experts on different aspects of health care resource allocation. In the end, the participants recorded the results of their deliberation process in a citizens' vote



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

(*Bürgervotum*). The aims of the citizens' conference were basically threefold. Beside, first, a political interest in triggering the public debate and demonstrating that ordinary citizens are willing and able to engage in such intricate issues as prioritization, and, second, a scientific interest in the method of the citizen jury as such, Stumpf *et al.* (2014: 54), third, sought to discover new arguments and rationales.¹⁸ Regarding the latter aim, the results are rather disappointing, though. As basic values, the participants enumerate "human dignity," "equality," "solidarity," and "efficiency," and the criteria which should govern resource allocation include "need," "the patient's will," "quality of life," and "cost-efficiency" (*Bürgervotum* 2014: 138, 140f.).¹⁹ While it is certainly reassuring from a political point of view that citizens of a democratic state endorse these values and principles (see Stumpf/Raspe 2012: 423), the results are a far cry from providing new arguments or unexpected lines of reasoning. Hence, it is doubtful whether the conference actually established that citizens can make a "substantial contribution" to the prioritization debate, as Stumpf and Raspe (2011: 317) conclude.

The limited informational content for the academic debate of the *Bürgervotum* in particular and of any public participation exercise in general is not surprising, though, and the reasons for this finding point toward a serious shortcoming of these methods. Indeed, it is not very difficult to settle on such vague umbrella terms as equality or solidarity. The real controversies arise when it comes to the concrete interpretation of these concepts and when the different values or criteria conflict with each other. As to the clarification and operationalization of the notion of equality, Mullen (2008: 398) offers the following example:

For instance it is possible to have agreement on the equal value of each person's life and hence the value of providing life saving [sic] treatment to any who need it, but to also have vehement disagreement on whether this should mean that priority is given to preventing as many deaths as possible, or to giving each person the same chance of receiving treatment if resources mean that not all can be treated.

Such considerations are lacking in the *Bürgervotum*, which is rather characterized by a pretty general style. To give an example, the participants state that each person has an equal claim to medical treatment, but at the same time postulate that resources should be allocated efficiently; yet, efficiency should be attributed less weight when medical need is high (see Stumpf/Raspe 2011: 317). Whether



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

these considerations can be consistently connected with each other remains unreflected. The fact that the *Bürgervotum* does not offer such conceptual reflections is certainly not surprising given the fact that the authors are not trained in challenging conceptual work and thorough ethical reasoning. Considered against this background, the hope that the public might bring new ideas and arguments into the debate seems “to involve considerable optimism” (Mullen 2008: 407).

The previous reflections furthermore suggest that this hope seems to rely on a misapprehension of the special competence of philosophers in general and ethicists in particular (see Lübke 2013: 255).²⁰ Although everyone is of course confronted with moral issues time and again, certain professional skills are essential when it comes to approaching more abstract and less quotidian normative questions. This is where ethical expertise comes in. Above all else, this expertise consists in the skills of analytical and critical thinking, the acquaintance with conceptual issues, and the ability of “the careful, critical evaluation of the soundness of arguments” (Brock 1995: 222).²¹ Beyond that, a professional ethicist is usually familiar with different ethical theories and sensitive to the typical problems arising if one path of argumentation or the other is taken (see Birnbacher 2002: 100f.). Having spent much time on thinking about ethical theories and discussing different philosophical positions thus enables the ethicists to enter an ethical discourse with a map of possible paths the argument will take (see Rippe 2000: 159f.).²² This is especially relevant in view of the fact that the academic discourse frequently precedes the public debate on a certain subject. Hence, when the ethicist moves in the public domain, he is already familiar with the major arguments and counter-arguments put forward in the academic sphere and has had the opportunity to carefully reflect on them (see *ibid.*). Furthermore, ethicists are used to raising disturbing questions and pointing out logical inconsistencies without necessarily being able to answer or solve them immediately. This requires a considerable amount of calmness and robustness, especially in the context of interdisciplinary debates (see Birnbacher 2002: 101). Against the background of these considerations, consulting the public can be seen as primarily serving an educational function, mirroring deliberation exercises in philosophy classes.²³

Proclaiming that such an ethical expertise exists does not imply the assertion that ethicists also have a moral expertise and, hence, a privileged route to true moral knowledge, though.²⁴ In the present context, this means that the ethicist does not



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

as a matter of principle know better than anyone else what a just resource allocation is. However, given his ethical expertise and due to the fact that he has much more time to think these things through than a lay person, he is more likely to do so. At least, although the professional ethicist cannot claim to provide the right arguments for the right conclusions, he is able to restrict the range of possible options by eliminating problematic arguments and conclusions (see Birnbacher 2002: 102; Gesang 2002b: 118). Crucially, the authority of the ethicists' moral advice solely derives from the authority, i.e., plausibility and coherence of their arguments (see Rippe 2000: 160f.; Düwell 2000: 106f.).

Beside the required skills in reasoning, the normative-ethical relevance of empirically elicited preferences is further challenged by the fact that dealing with prioritization decisions also requires a "quite particular empirical understanding" (Mullen 2008: 406).²⁵ Since the issue of health care resource allocation is characterized by considerable complexity and abstraction, well-considered, informative answers from lay persons cannot be expected (see Rippe 2000: 149). All in all, empirical ethics can be described with Hausman (2000a: 43) as a "counsel of despair": "Moral questions concerning health and health policy [...] are terribly difficult. [...] But if health administrators, economists, philosophers and theologians are all baffled, surely members of the target group who take the issue seriously must be baffled, too."

This emphasis of a special ethical expertise may aggravate the concern that decisions on priority setting are made by small and closed circles of experts or "philosopher kings." Yet, this fear is unsubstantiated. As pointed out already, ethicists cannot claim a special moral expertise *per se* so that the authority of their moral suggestions completely hinges on the quality of their arguments. Moreover, they certainly have no political authority to actually make the decisive decisions on priority setting. Rather, their task consists in the thorough reflection on the normative-ethical issues at stake, including in particular a detailed conceptual inquiry (see Lübke 2015: 66). In order to avoid an undue influence of one particular person or theory, politicians and the public should be made aware of the differences in opinion that remain even after thorough discussion (see Gesang 2002b: 133). On this level, it is the ethicists' task to present their respective arguments in an intelligible way to the public so that the people gain the opportunity of evaluating them themselves.²⁶ After the proposals have passed the democratic decision process – during which deliberative exercises can certainly help informing the public, triggering a discourse etc. – the ultimate decisions on



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

health policy are made by the sovereign, that is to say, the citizens anyway.²⁷ If, by contrast, normative-ethical reflections on priority setting supposed to inform the political decision making process relied on what the public think already, political consulting would be superfluous in the first place.

7.5 Respecting individual autonomy

Up to now, the arguments presented for involving the public in the debate on resource allocation focused on improving the prioritization decisions in one way or another. A different thread claims that in a liberal and democratic society, respecting individual autonomy requires public participation in order to legitimize prioritization decisions. Thereby, legitimization can be understood in two ways. First, it can be interpreted strategically, as Menzel's (1999: 262f.) following argument illustrates:

the defense of a policy decision is more difficult if those who bear its primary effects have not had that notable input. The implications for the use of CEA are not hard to see [...] CEA's moral and political future will be brighter if it can note precisely where in its process influential values have been contributed by representatives of those who are affected by the final decision.

This rationale offered for the participation of stakeholders, i.e., the persons potentially affected by prioritization decisions (see Friedrich *et al.* 2012: 412), apparently refers to the political feasibility of CUA (see Walker/Siegel 2002: 268). Arguments along this line are common in the literature; public participation is endorsed in order to enhance both the public's "acceptance of priority-setting decisions," improve the people's "trust and confidence in the health care system" (Bruni *et al.* 2008: 15), and to "strengthen the relationship between citizens and decision makers" (Shah 2009: 82).²⁸ As stated at the outset of this chapter, such strategic functions of public participation are not of interest in the present context, though. Instead, the current concern is with the second meaning of legitimization, which refers to a profound moral legitimization of certain policy measures within a democracy (see Powers/Faden 2006: 179).

Such arguments for public participation rest on the assumption that, due to the basic value of autonomy, only the consent of the persons concerned can legitimize priority setting in health care: "If it is maintained that people should have control over the governance of their own lives and interests it seems plausible to suggest



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

that they should have a say in deciding ethical frameworks” (Mullen 2008: 401).²⁹ This case meshes neatly with the economic emphasis on consumer sovereignty and anti-paternalism. Accordingly, it is reinforced by the claim that since the citizens finance the health care system, they should be the ones to decide about the distribution of its resources: because “health care is a public good [...] it makes sense to ask the public how it would choose to distribute the good” (Ubel/Loewenstein 1996: 234).³⁰ Taking these two arguments together leads to the claim that those “whose health is at stake and who are paying for the health care should be the ones to decide” (Hausman 2002: 643).

And yet, both the reference to individual autonomy and the alleged parallel between prioritization decisions and democratic decision making is mistaken. To begin with, autonomy, just as consumer sovereignty, concerns private decisions of the individual, the prime example being the choice of consumption goods (see Lübke 2013: 254, 2015: 65f.). If individual liberty and autonomy are accepted as basic values, there is indeed “a strong moral argument to be made for letting competent adults make decisions about their own well-being” as long as the rights of others remain unharmed (Walker/Siegel 2002: 267). Issues of justice, by contrast, inherently involve considerations about the rights and the well-being of other persons. Referring to the individuals’ autonomy when it comes to rationing decisions thus amounts to a category error. By the same token, the argument that those financing and potentially benefitting from the health care system should decide on the distribution of resources is erroneous, as Hausman (2002: 643) points out:

If health and its protection raised no moral questions, there would be no question that health policies should depend on the wishes of those they serve. [...]. Funding cosmetic surgery rather than inoculations is not just a matter of social preferences. Age-weighting, discounting, and weighting of different health states all matter ethically, and no reason has yet been given why the evident interests of the target population should imply that their values govern.

The analogy between rationing decisions and democratic decisions, in turn, fails for two reasons. First and foremost, democracy obviously is not tantamount to the majority rule (see Price 2000: 276; Powers/Faden 2006: 180; Schicktanz 2009: 230). Instead, democratic processes are embedded in a constitutional framework



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

guaranteeing fundamental rights and liberties, especially for minorities (see Sagoff 1986: 308; Hausman 2012: 97). To quote Dworkin (1978: 133f.), “[d]e-mocracy is justified because it enforces the right of each person to respect and concern as individual.” Therefore, arrangements are required which make sure that democratic procedures do not undermine their own justification and this is the very rationale for a system of rights (see Dworkin 1978: 134).³¹ Once rights and liberties are guaranteed, it is up to the individual to choose her preferred way of life as she sees fits, i.e., to lead an autonomous life according to her personal preferences (see Walker/Siegel 2002: 267). The debate on prioritization, however, is a dispute concerning the constitutional framework, so to speak, and not on the individual style of living. It is about the establishment of criteria for a just system of allocating medical resources, which needs to make sure that the rights to equal concern and respect are guaranteed.

The second discrepancy between democratic decision making and public participation in the rationing discourse consists of the kind of legitimization at stake. Involving certain samples from the public into the prioritization discourse can never virtually legitimize a collective decision the way a democratic vote does, of course. This is because in the former case, not everyone had the opportunity to participate in the respective survey or the deliberative group. Yet, it is precisely this “fundamental notion of equality” that provides democracy for its normative justification (Walker/Siegel 2002: 271). To this extent, even a statistically representative sample cannot represent the whole population in a morally relevant sense, for even if the elicited preference pattern was indeed identical with the preference pattern of the whole population, “the democratic procedural justification is lost” (ibid.).³² As a result, “mimicking majorities” by means of empirical elicitation of prioritization preferences ultimately fails (Powers/Faden 2006: 184). To sum it up, quoting Hausman (2002: 644f.):

Although moral questions concerning what is right and wrong are very different from most scientific questions, there is no stronger case for deciding moral decisions democratically than there is for deciding scientific questions democratically. [...] Those who support empirical ethics might complain that this disenfranchises the population. Of course it does. It disenfranchises everyone, because it maintains that the answers to evaluative questions, like the answers to factual questions, cannot be found in taking a vote. With respect to political decisions, on the other hand, there is, of course,



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

no question of disenfranchising anyone.

Before concluding this chapter, a remark on a third possible line of argument for the empirical elicitation of prioritization preferences is required. It may be argued that the whole critique offered here is beside the point, since consulting the public is not connected with any normative intention at all, but merely amounts to descriptive, scientifically motivated research instead (see Friedrich *et al.* 2012: 415). A first reply to this argument was already given above. The endeavor of positively describing prioritization preferences apparently presumes the existence of something that can be described in the first place. As argued in the previous chapter, however, it is doubtful that subjects have certain prioritization preferences, just waiting to be elicited. By contrast, it is much more likely that they form their preferences during the elicitation task so that the elicitation methods gain pivotal normative importance. As shown by means of different deliberation exercises, respondents are also likely to change their normative framework during the task. Hence, the elicitation of prioritization preferences is a far cry from merely mirroring preference structures prevalent in the public. Beyond that, it should be noted that the discourse on resource allocation is not of mere academic interest, but is politically highly contested. Once published, the empirical data can, and most probably will be used to support or to reject certain policy measures (see Murray 1996: 2f.).³³ Also, they may enfold “the normative power of the factual,” as Borry *et al.* (2005: 48) put it: “the publication of research data gives rise to standpoints and influences decisions which, in turn, can alter the normative structure of the reality of action.” As to the current concern, this could mean that the more often empirical surveys show that the majority of subjects consider individual responsibility as the most important allocation criterion, the more this result becomes a fact itself. In the end, it seems all too easy to slide from mere descriptive data to normative conclusions (see Borry *et al.* 2005: 42f., Schicktanz 2009: 225ff.). Hence, claiming that the data elicited only serve a descriptive purpose is irresponsible.

In sum, the literature reviewed here does not offer any convincing arguments for the normative-ethical relevance of the public’s prioritization preferences. The rationales focusing on the improvement of the prioritization discourse neglect the difference between empirically informed ethics on the one hand and empirical ethics on the other. In addition, it misconceives the competence of professional philosophers. The second group of arguments, in turn, basically relies on a consumer choice framework and disregards the difference between genuinely private choices on the one hand and moral judgments, concerning the rights and



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

claims of other persons on the other. The parallel drawn between prioritization and democratic decision making is additionally flawed because the latter are characterized by provisions to protect individual rights and because including only a sample of the population in the prioritization discourse cannot confer democratic legitimacy on the results. Hence, even if the arguments put forward in this study were wrong, the weighting approach still would not offer an appropriate account of dealing with fairness problems. In conclusion, two important qualifications deserve to be repeated. First, none of this implies that ethical reasoning should be independent of empirical data. Quite to the contrary, as pointed out above, thoroughly conducted applied ethics certainly has to be empirically informed. Second, the previous considerations are not a plea for an expertocracy. When it comes to the implementation of health policy, the citizens have the last word anyway.

Notes

1. In view of this focus of interest, the questions of how or whom to ask can be neglected and the notion of “the public” will be used broadly. See on this concept Williams *et al.* (2012: 29) and Schicktanz *et al.* (2012: 134f.).
2. See also Marckmann/Siebert (2002b: 187), Gaertner/Schokkaert (2012: 14), and for a critique Lübbe (2015: 64).
3. See also Ubel *et al.* (2000b: 899).
4. In the same vein, Lengahan (1999: 47) states that “public involvement could resolve some of the dilemmas inherent in rationing decisions and lead to better decisions.” See Richardson (2002: 627), Richardson/McKie (2005: 271), and Gaertner/Schokkaert (2012: 14).
5. See also Lomas (1997: 106) and Olsen *et al.* (2003).
6. See also Diederich/Schreier (2013: 280).
7. Olsen *et al.* (2003: 1171) observe a lack of “research to distinguish ethically based considerations from prejudices.”
8. Italics added.
9. The same case is made by Hausman (2002: 644, 652), Walker/Siegel (2002: 267), Goldenberg (2005: 7), Powers/Faden (2006: 184, 190), Düwell (2009: 205), and Lübbe (2013: 254f, 2015: 67f.).
10. See also Schicktanz *et al.* (2012: 134).
11. See also Richardson/McKie (2005: 271).
12. For similar distinctions see Borry *et al.* (2004: 44ff.) and Düwell (2009: 204ff.).



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

13. The example of slippery-slope arguments is provided by Birnbacher (1999: 329), Musschenga (2005: 476), Düwell (2009: 205), and Schicktanz *et al.* (2012: 133).
14. The necessity of including empirical data into applied ethics is highlighted by Birnbacher (1999), Düwell (2000: 83), and Düwell/Steigleder (2003: 30ff.).
15. Note that there is no consensus about general questions as to the definition and, hence, the character and purpose of applied ethics. See Karminsky (2002: 37ff.).
16. See also the contributions in Stumpf/Raspe (2014).
17. On the concrete proceeding see Stumpf (2014b).
18. See also Stumpf/Raspe (2012: 419) and Stumpf (2014a: 17f.).
19. My translation.
20. For the sake of readability, I will refer to philosophers and ethicists in the following, but the same reasoning applies to anyone working in normative disciplines and having the required argumentative skills, such as law scholars (see *ibid.*). Note that both the nature and the very existence of a special ethical expertise are contested in the literature. See Düwell (2000), Rippe (2000), Birnbacher (2002), Gesang (2002b), Düwell/Steigleder (2003: 29), and Schicktanz *et al.* (2012).
21. See also Rippe (2000: 159), Düwell (2000: 107), Runtenberg/Ach (2002: 18f.), Kaminsky (2002: 36f.), Birnbacher (2002: 101), Goldenberg (2005: 5), Mullen (2008: 406), Schicktanz *et al.* (2012: 131), and Lübbe (2013: 255).
22. The image of a moral map is also invoked by Runtenberg/Ach (2002: 22f.) and Gesang (2002b: 133).
23. Daniels (1998: 41) states: “Anyone who has taught students or quizzed audiences at lectures on these types of cases knows that initial responses are modified once cases are given that probe the reasons for them.”
24. The juxtaposition of ethical and moral expertise stems from Birnbacher (2002: 101). See Rippe (2000: 159), Düwell (2000: 106f.), and Gesang (2002b: 117ff.).
25. See also Goldenberg (2005: 5).
26. See Düwell (2000: 107) and Gesang (2002b: 132f.). This may constitute the pivotal difference between ethical experts and experts from other disciplines: when we seek technical advice, for instance, we usually trust the experts’ statement and do not ask the technician how he came to his



ON THE NORMATIVE STATUS OF EMPIRICALLY ELICITED PRIORITIZATION PREFERENCES

Excerpted from *The Economics of Resource Allocation in Health Care*

conclusion. The case is totally different when it comes to moral questions, though. Here, it is crucial that the person seeking the advice understands and accepts the reasoning behind a certain conclusion. Developing and explicating this reasoning is thus the pivotal task of the ethical expert. See Düwell (2000: 107).

27. See also Price (2000: 276), Walker/Siegel (2002: 272), Bognar/Hirose (2014: 153), and Lübke (2015: 66).

28. See also Friedrich *et al.* (2012: 412f.), Williams *et al.* (2012: 40ff.), Gaertner/Schokkaert (2012: 8f.), and Williamson (2014: 6).

29. See also Cookson/Dolan (1999: 64) and Schicktanz *et al.* (2012: 134).

30. See also Nord (1999a: 2), Bruni *et al.* (2008: 15), and Diederich/Schreier (2013: 265).

31. Hausman (2002: 644) states:

Fundamental rights and liberties and a regime that secures the self-respect of citizens cannot be abrogated by majority vote.

Permitting democracy to extend to such questions would lessen security and would place justice and equal moral status of adult citizen at risk.

32. See also Walker/Siegel (2002: 269) and Powers/Faden (2006: 187f.).

33. Murray *et al.* (2002a: 15) state:

In examining the properties of various summary measures, it is important to bear in mind the ultimate goal of influencing the policy process [...]. Because of their potential influence on international and national resource allocation decisions, summary measures must be considered as normative measures.

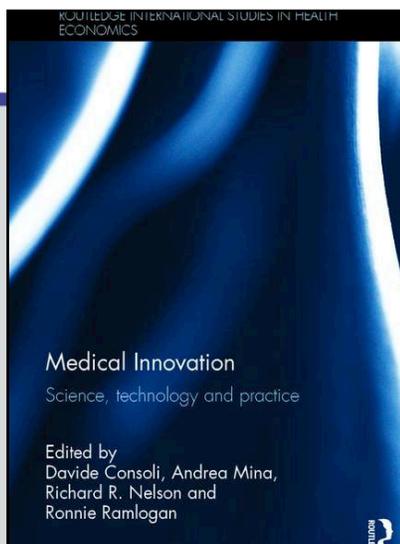


CHAPTER

4

CORONARY ARTERY DISEASE

BY ANDREA MINA, RONNIE RAMLOGAN,
STAN METCALFE, AND GINDO
TAMPUBOLON



This chapter is excerpted from

Medical Innovation: Science, technology and practice

Edited by Davide Consoli, Andrea Mina, Richard R. Nelson, and Ronnie Ramlogan

© 2016 Taylor & Francis Group. All rights reserved.



[Learn more](#)



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

Introduction

Cardiovascular diseases are a leading global cause of death (WHO 2011). Coronary artery disease (CAD) is the most common type of cardiovascular disease: it is the end result of a process called atherosclerosis, by which an atheroma – a plaque of fatty deposits – forms on the inner layer of the coronary artery and impedes the flow of blood to the heart. In the early stages, the build-up of these deposits is silent (symptom-less) but, as the disease progresses, it can induce a variety of conditions ranging from mild chest pains (angina) and shortness of breath to heart failure and sudden death.

In this chapter we investigate what has been regarded by many as one of the most important medical breakthroughs of the last decades of the twentieth century.^{1,2} Percutaneous Coronary Intervention (PCI) or, as it is more familiarly known, coronary angioplasty³, emerged in the late 1970s to become a technique that is now more commonly used than the principal surgical alternative, coronary artery bypass surgery (CABG), in the treatment of advanced coronary artery disease. With the benefit of experience, PCI has proved to be an effective mode of treatment that has also brought with it a major transformation in the division of labour in cardiology, with the field of Interventional Cardiology being recognised as a separate and distinctive sub-speciality of the broader field of practice. But it has also proved to be a black box of uncertainty in which progress is hesitant in the face of challenging emergent medical problems. Mina *et al.* (2007) and Ramlogan *et al.* (2007) present large-scale bibliometric analyses of the medical literature on PCA and of patent records of relevant inventions. In this chapter we revisit the way in which this innovation emerged as a set of problem-solving sequences and identify the mechanisms of operations of a 30-year long co-evolutionary process of knowledge, technology and institutional change.

Treatments for CAD prior to the 1980s

For sufferers of the disease chest pains, angina, can be induced by physical exertion and/or emotional stress. These increase the heart's requirement for oxygen-enriched blood but, with narrowed arteries restricting blood supply, the sufferer experiences severe discomfort. The pain usually becomes evident when the vessel is only able to deliver at 30 per cent of its capacity. At this level of closure, the oxygen-enriched blood that the heart receives is only adequate when



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

the body is at rest. A heart attack may occur when a coronary artery becomes completely blocked and insufficient oxygen supply to the heart, called ischemia, results in the death of the heart muscle.

As recently as the 1960s, treatment options for angina or acute myocardial infarction (heart attack) consisted of few medications (mainly nitroglycerin), rest and hope.⁴ Between the 1960s and 1970s beta blockers and calcium channel blockers were added to the cardiologist's arsenal for dealing with angina. Unfortunately, while these started to provide effective relief by reducing the frequency and force of the heartbeat, they were not a cure for the underlying problem. Surgical treatment options also improved from the 1960s with the development of coronary artery bypass surgery.⁵ This is a major invasive and complicated surgical procedure taking between three and six hours to perform. It requires general anaesthesia, the use of a heart-lung machine to substitute for heart-lung functioning during surgery and a lengthy post-surgery recuperation period. Put simply, the idea behind the procedure is to improve the blood flow to the heart muscle by bypassing the blockages with blood vessels harvested from the leg (saphenous vein) or chest wall (internal mammary artery).

At the time of its introduction bypass surgery was regarded as a revolutionary procedure. The idea of stopping a heart, restoring its blood supply then restarting it, bordered on the miraculous. The technique spread rapidly although figures for the volume of procedures undertaken in the early period are patchy. In 1973 around 25,000 operations were performed in the US and this increased to 70,000 by 1977 (OTA 1978). Elsewhere, the absolute number of procedures was small in comparison. In the UK, for example, 2,297 operations were carried out in 1977 and this increased to 4,057 by 1980 (British Heart Foundation 2004).

The diffusion of this procedure was not without controversy and this was primarily related to the evidence base on which bypass surgery was being promoted. A debate raged throughout the 1970s about the quality of evidence that was being assembled about the efficacy of bypass surgery relative to medical therapy.⁶ So much so that the US Office for Technology Assessment (OTA 1978) was decidedly lukewarm about the procedure arguing that, in terms of survival, the VA randomised trial and a number of others showed that surgery did not appear to bring any appreciable longevity benefit when compared with medically treated patients. The OTA did concede the point that bypass surgery gave 'excellent symptomatic relief from angina pectoris' (p. 43) although it was careful to caution



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

about placebo effects associated with surgery.⁷

The advent of a new treatment modality

It is against this background of uncertainty about the efficacy of coronary bypass surgery that Percutaneous Coronary Intervention (PCI) emerged. Its development was truly original but it did nevertheless build on prior knowledge and existing techniques, as is often the case in the history of medical technologies. The two techniques that laid the foundations for PCI were cardiac catheterisation and transluminal angioplasty.

Cardiac catheterisation is a diagnostic procedure in which a catheter (a thin flexible tube) is inserted into the right or left side of the heart. This procedure could then be used to produce angiograms (x-ray images) of the coronary arteries and the left ventricle, the heart's main pumping chamber, and/or used to measure pressures in the pulmonary artery and to monitor heart function. Werner Forssmann is credited with being the first to introduce a (urological) catheter into the right atrium (his own!) in 1929. Branded as 'crazy' by his contemporaries, his immediate reward for this achievement was dismissal from his German hospital (he went on to win the Nobel Prize in 1956). By the 1950s, however, following the work of Cournand, Seldinger and others, diagnostic catheterisation had become established as the main technique for investigating cardiac function.

The American Charles Dotter coined the term Percutaneous Transluminal Angioplasty (PTA) (Rosch *et al.* 2003) for a treatment that he developed in the 1960s. During an angiographic examination he inadvertently reopened an occluded right iliac artery (in the lower abdomen) by pushing through it with an angiography catheter. Realising the therapeutic potential of dilating narrowed arteries, Dotter went on to refine the technique but, as was the case with Forssmann, this was met with scepticism by the American medical fraternity. However, European radiologists took a more positive view and soon institutionalised the term 'dottering' (of patients) to refer to improving the patency of arteries by the introduction of a series of coaxial catheters. One of Dotter's European followers, Eberhardt Zeitler, used the technique on a large number of patients and this helped the diffusion of the technique among practitioners.⁸ Among them was the German radiologist Andreas Gruentzig, who is named in the annals of medical history as the pioneer of PCI.



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

Gruentzig was exposed to the Dotter method when he was based at the Ratchow Clinic in Darmstadt (Germany) in the mid-1960s. At Ratchow he was a Research Fellow in Epidemiology studying coronary artery disease but later, wanting to become a cardiologist, he moved to the University of Zürich (in 1974). Here he began to think about whether the Dotter method could be applied to the heart recognising that 'any application of the dilatation procedure to other areas of the body would require technical changes' (King 1996, p. 1624).

Encouraged by his colleague and Joint Head of Cardiology, Wilhelm Ruttishauser, Gruentzig proceeded to develop by himself new prototypes of catheters with a soft balloon at the tip that could be inflated to break the plaque inside the vessel, the foundation of what he called Percutaneous Transluminal Coronary Angioplasty (PTCA). Two crucial developments followed. First, in 1972, he introduced a balloon made of PVC, a tough, less compliant material than latex, with which he had experimented earlier. Second, in 1975, he developed a new single and then, more importantly, a double lumen catheter. This was a single catheter with two tubular channels, one for inflating the balloon and the other for injecting contrast media and monitoring intravascular pressure. Some early results were presented to the American Heart Association meeting in 1976 to a largely sceptical audience but, by 1977, he had succeeded in performing the first PTCA on a human patient in Zürich. The technique diffused relatively quickly thereafter, particularly in the US.

The development of PCI: phase 1

The development of PCI can be broken down into several overlapping phases of activity involving both clinical practice and technology. The first phase was connected with the development of PTCA or balloon angioplasty, Gruentzig's original breakthrough. This was a phase of intense exploration of the search space opened up by this new technology. Practitioners were concerned with identifying the medical conditions under which this new technique would deliver benefits to patients under tolerable margins of risk. A study by Cowley *et al.* (1985), for example, produced under the sponsorship of the National Heart Lung and Blood Institute (NHLBI), provided early evidence of the effectiveness of balloon angioplasty procedures in medical centres across the US and in other countries. This was one of a number of foundation papers that provided early evidence from a Registry set up by NHLBI in 1979 to collect, analyse and disseminate the results from using the balloon angioplasty procedure in medical centres across the US and



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

in other countries (Mullins *et al.* 1984). By the year 2000 NHLBI had set up several registries to understand the long-term efficacy of PTCA and the alternatives approaches that were emerging. The first registry followed 3,079 patients who received PTCA between 1977 and 1982. A second registry followed the outcomes of 1,500 patients from the first registry for a minimum of five years plus an additional 2,000 newly entered patients who received PTCA in 1985 and 1986, while a third registry, the New Approaches to Coronary Intervention (NACI), followed approximately 4,424 patients between November 1990 and February 1997.⁹

Two major unexpected medical problems, acute occlusion and restenosis, emerged to trouble the early diffusion of PTCA.

Acute occlusion

The PTCA procedure involves making small incisions in the groin or arm under local anaesthetic where a catheter can be fed through to the obstructed coronary vessel. The balloon attached to the end of the catheter is used to unblock the vessel and consequently restore the blood flow. In a small but significant number of cases, the procedure resulted in weakening and collapse of the internal structure of the artery which would subsequently require emergency bypass surgery. Several studies in the early 1990s comparing the outcomes of CABG with PTCA showed higher re-intervention rates with PTCA (Reul 2005). The incidence of acute vessel closure was 3–5 per cent, and would occur within the first 24 hours of the procedure due to vessel dissection or acute thrombus formation (Hamid and Coltart 2007). Thus, in the early days, the practise of PTCA was contingent on having emergency coronary artery by-pass operating facilities available. Second, tissue trauma at the site of the procedure sometimes triggered blood clotting which, depending on severity, would require major invasive treatment. Over time, however, the occurrence of this would be countered with anti-thrombolytic drugs.

Restenosis

The second major problem was restenosis – the appearance of a new constriction in the artery (Pepine *et al.* 1990). It had a high incidence rate, between 25–50 per cent in patients having undergone balloon angioplasty. This constriction was not atherosclerotic in nature but resulted from the outgrowth of ‘endothelial’ cells that normally line blood vessels and has been likened to ‘over-exuberant’ tissue healing and regeneration similar to scar formation. This tended to occur during the first three to six months after the procedure and effectively nullified the intervention so that the patient would require further revascularisation.



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

Various candidate devices were tested to solve the problems but these were of limited success in dealing with restenosis. For example, in contrast to dilating and compressing the plaque John Simpson developed an atherectomy device to remove it with barotrauma (change in pressure) (King 1998). Such attempts eventually led to the development of the rotational atherectomy device, a diamondcoated burr rotating at around 150,000 revolutions per minute to remove plaque through abrasion. While this device met with limited success, high complication rates, cost and the paucity of data from randomised trials, limited widespread acceptance and this technique has been largely confined to patients with hard, calcified plaque (Whitlow 1997). Another method, laser angioplasty, used the energy created by an argon laser focused on the end of a catheter to ablate tissue on contact with the catheter tip, but this did not compare favourably with balloon angioplasty (King 1998).

Phase 2: the introduction of the stent

The second phase in the development of the new procedure witnessed the introduction and development of the stent in response to the problems that were limiting the potential of the new technique. The term 'stent' originated in dentistry and dates back to the nineteenth century, but it appeared in the non-dentistry medical literature when used by Dotter in 1983 in a study of percutaneously implanted vascular endoprosthesis in canine experiments (Balcon *et al.* 1997). In the late 1980s stent developed into a metal meshwork structure (Ruygrok and Serruys 1996) which could be inserted with the balloon catheter, opened up and left in the vessel to support the walls of the artery and prevent acute vessel closure and restenosis.

Ulrich Sigwart performed the first coronary stent procedure in 1984 in Lausanne (Towers and Davies 2000) and several papers published between 1985 and 1988 described the progress with the balloon-mounted coronary stent developed by Julio Palmaz and Richard Schatz (and produced by Johnson & Johnson) (Kent 2010), in both canines and humans, and the self-expanding Wall stent used in humans.

The improved technique was met favourably by practitioners and rapidly diffused. By 1999, coronary stenting was performed in 84.2 per cent of PCI procedures (Serruys *et al.* 2006) but, as with PTCA before it, the diffusion of stents into wider practise presented the cardiology community with unanticipated new challenges.



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

In many cases the use of a stent resulted in a sub-acute thrombotic coronary artery. This is an occlusion (blood clot) at the stented site that formed anywhere between 24 hours and up to 30 days after the procedure. The possibility of such an event necessitated a complex anti-coagulation treatment and prolonged hospital stays. However a further unexpected complication arose, that of in-stent restenosis. Under normal circumstances new tissue grows inside the stent covering the metal mesh when this is placed inside the artery. This allows the smooth flow of blood over the stented area. Subsequently, however, scar tissue may develop underneath the new healthy lining and this may be so thick that it obstructs the blood flow. This new kind of restenosis usually occurs between 3 to 6 months after the procedure and affects about 25 per cent of patients (Dangas and Kuepper 2002).

The enthusiasm for stents waned considerable in the early 1990s, particularly as the evidence being gathered showed that stenting did not give a better success (or lower complication) rates than those of routine balloon angioplasty (Balcon *et al.* 1997). However, the results from three important trials marked a turning point for the medical community and favoured increasing acceptance and use of stenting technology as an integral part of PCI. Studies by Serruys *et al.* and Fischman *et al.* in 1994 followed by Colombo *et al.* (1995) provided persuasive evidence of the advantages of stenting compared with 'simple' balloon angioplasty and to the finding that the success rate of the procedure depended heavily on the placement of the stent, focusing on full expansion, adequate deployment of the stent using intravascular ultrasound and by the use of simplified and more effective anticoagulation protocols.

Subsequently, further exploration focused on improving the use of stents and dealing with the problem of the plaque reforming inside the stent (in-stent restenosis). This very problem also triggered the emergence of parallel trajectories of research, principal among them being drug-eluting stents in which stents coated with drugs are locally delivered to the point of the lesion.

While the introduction of stenting technology reduced the restenosis rate from 40 percent to around 30 per cent (Fischman *et al.* 1994), it created a new problem - 'in-stent restenosis' - and drug-eluting stents (DES) were developed to specifically target this problem. Stents were now coated with a polymer that released a drug to inhibit the cell proliferation causing restenosis over the course of 6 to 9 months. Cordis launched its Cypher® stent in 2003, and this was followed by the Taxus® stent from Boston Scientific in 2004. In subsequent years other manufacturers



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

followed with variations in the design (type of metal used; strut thickness; mechanics of strut interlinkage), drug used and the release characteristics of the polymer.

DES quickly attained a high penetration, accounting for around 50 per cent of the coronary stent market (Li and Kozlek 2012) due to the lower rates of restenosis, although there was an initial lack of long-term safety and efficacy data. Subsequent studies compared individual DES both with their bare metal (BMS) equivalents and with each other (see, for example, Mahmoudi *et al.* 2011). In general the evidence has shown that, compared with BMS, DES significantly reduce the incidence of restenosis to levels of under 10 per cent (Farooq *et al.* 2011) but meta-analyses were beginning to show that DES had a much greater risk of very late stent thrombosis (where a blood clot forms inside the stent more than a year after insertion) compared with BMS (Bavry *et al.* 2006). Thus the consensus that emerged was that DES require a longer period of dual antiplatelet therapy relative to BMS in order to prevent stent thrombosis. On-going studies continue to evaluate newer stent platforms, drugs, polymers, polymer-free stents and bio-absorbable stents and it is becoming clear that there will not be one single stent suitable for all patients and lesions (Wilson and Cruden 2013).

Beyond technology: co-evolutionary dynamics of knowledge and institutions

The development of coronary angioplasty is a story of entrepreneurial individuals as well as the building of a community of practitioners. It unfolds along interrelated epistemic, social and institutional dimensions. In spite of the initial conservative reaction to his work, Gruentzig's pioneering insights and first successes opened up tremendous opportunities for further improvements of the new technique whereby a growing number of practitioners started experimenting with it and a growing range of devices started to be developed after the introduction of the first tested prototypes (Figure 2.1).

In this respect, the nature of this innovation process mirrors that of other significant medical innovations in being an uncertain co-evolutionary process of knowledge and technique (Gelijns and Rosenberg 1994; Gelijns *et al.* 1998; Gelijns *et al.* 2001) in which trajectories often emerge in the form of sequences of innovative ideas (Metcalfe *et al.* 2005). These involve coherent directions of change



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

– or trajectories (Dosi 1982) – and signal the cumulateness of interdependent research activities whose results build on previous knowledge. Furthermore, these imply specific configurations (for example, technical designs) embodying ways of combining knowledge that gradually become formal or informal standards (Utterback 1994) in the unfolding search for solutions to problems.

The growth of PCI practice was accompanied by many improvements in devices and practice including the invention of the steerable balloon catheter by Simpson in the early 1980s (Simpson *et al.* 1982). It is this contribution that enabled clinicians to access the most distal lesions and to achieve greater direction control of the catheter through the coronary system. Most importantly, the structure of these many complementary contributions to the innovation sequence also reflects

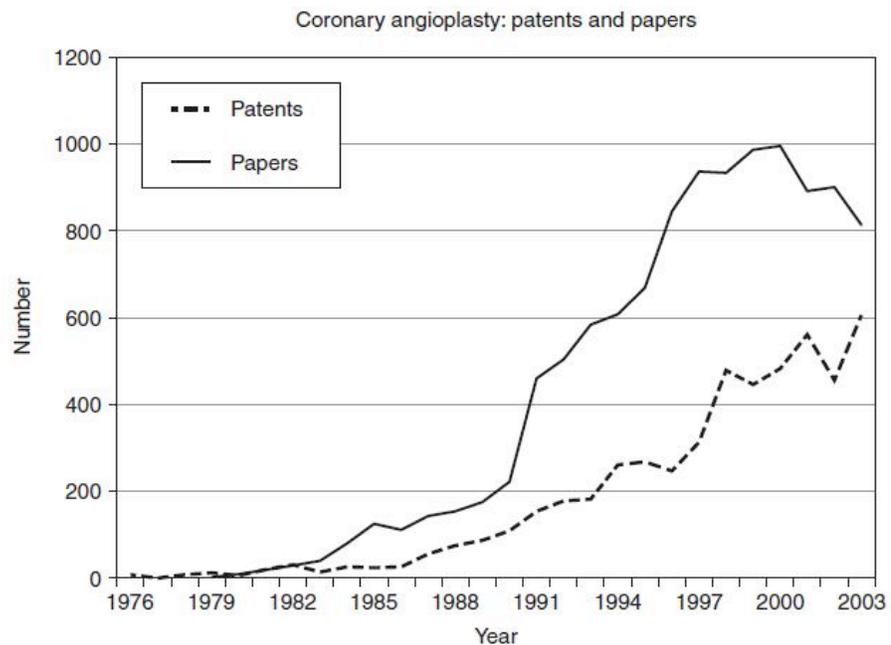


Figure 2.1 The growth of codified knowledge.

Source: Mina *et al.* 2007, Fig. 3, p.797, with permission.

the shift in the nature of the dominant problem over time. The solution to the catheter problem and Greuntzig's balloon device to compress the plaque opened up new territory but it was soon found that restenosis – the reformation of the plaque after the procedure – occurred in a significant number of patients, drastically reducing the efficacy of the treatment and raising its real cost. The solution to this problem was the invention and innovation of the stent (Eeckhurst



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

et al. 1996).

As we have observed above, after the original breakthrough further exploration focused on improving the use of stents and dealing with the problem of the plaque reforming inside the stent (in-stent restenosis). This very problem also triggered the emergence of parallel trajectories of research. Among the various solutions that were explored were stents coated with drugs that are locally delivered to the point of the lesion, pharmacological therapies that precede or follow stenting, and radiation therapies. And the search for yet better solutions is ongoing.

An important aspect of this process was the initial conditions for the early diffusion of PCI. When Gruentzig presented his first results to the American Heart Association conference in 1977 his intervention generated enough interest for him to receive numerous requests from other cardiologists wanting to learn the technique. A close-knit – and practice-based – micro community developed around Gruentzig that constituted the backbone for the diffusion of the new technique.¹⁰ It included, among others, Richard Myler who, together with Simon Sterzer, was the first to perform angioplasty in the United States in 1978 (King 1998). Over time Gruentzig's connections with the US intensified to the point that in 1980 he moved from Zurich to Emory University, where he had fewer constraints on the lab time he was allowed by his employer. At Emory not only could he work on an increasing number of cases but he was also provided good teaching facilities, which he was very keen to have. He was convinced that a critical mass of expert practitioners needed to be trained as well as possible to avoid the formation of imperfect skills that would have jeopardised the long-term success of the new technique.

It is in this community that we find the seeds of science as well as the seeds of the industrial complex that was to grow over the following three decades. Gruentzig was not only scientifically minded but also entrepreneurial. His prototype catheter was developed, of all places, in his kitchen and he later entered into a relationship with Schneider, a Swiss medical needle manufacturing company based in Zurich, for manufacturing a marketable device and he applied for a patent with the US Patent and Trademark Office in 1977. This was granted in 1980 and paved the way for the diffusion of coronary angioplasty in the US.

Gruentzig also exerted a considerable degree of control over the production and the commercialisation of his catheters.¹¹ In the early days in Zurich he exerted tight control of the sale of the first angioplasty catheters that were being produced



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

under his instructions by Schneider, a small Swiss company later to be acquired by Pfizer in 1984 and sold on to Boston Scientific in 1998. It has been reported that he required practitioners who bought catheters to receive training or counselling from him on how best to use the device.¹² This significantly contributed to a reduction in the rate of failure of the procedure due to improper use in a phase of the diffusion process that was most delicate for the future success of angioplasty.

When the procedure started to be experimented with on a larger scale, the need also emerged to evaluate the performance of the technique in a systematic way. John Abele, who went on to found one of today's market leaders (Boston Scientific) and who started collaborating with Gruentzig in the development of double-lumen catheters, recalls that there were informal discussions between Gruentzig and Myler and himself about creating a registry to document what was being learnt in clinical practice wherever the technique was used. They then developed a very simple but comprehensive form where results of every procedure could be recorded. The registry that resulted from the collection and organisation of the forms lasted for no longer than 18 months but was not to disappear. In 1979 the National Heart, Lungs and Blood Institute (NHLBI) – the relevant branch of the National Institute of Health – invited a small working group to discuss how evaluation of PTCA could be managed. It was agreed that a workshop should be held to review the preliminary evidence and that a voluntary registry be set up. The PTCA registry, which we mentioned earlier, was established in March 1979 (Mullins *et al.* 1984) and was really a formalisation of the earlier database started by Gruentzig and others.

Thus, in this process of division of social labour, the monitoring function of the system passed to the NHLBI while the regulation of medical devices shifted to the FDA in the same years. Both these complementary modes of governance proved crucial to the extension of understanding and practice, and to the growth of the market on which investment by commercial firms depended. Market institutions do not usually, if ever, operate independently of wider regulatory norms and these norms – formal and informal – like the market itself, often are created and co-evolve with it. This case is a good illustration of this point.

Prior to the 1980s, as we have mentioned, Gruentzig and Myler personally exerted informal control of the availability, quality and safety of catheters. In a way, they acted as the first regulators of angioplasty devices before the FDA started implementing the reform of Medical Devices Law passed in 1976 and became the



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

ultimate arbiter for regulation of medical technologies. The FDA then became the institution that, in the public interest, exerted functions of evaluation and selection of the variety of drugs and devices stemming from the wealth of R&D activities that followed the sparse experimentation of the early years. Needless to say, its role grew in importance along with, on the one hand, the increasing commercial interests associated with angioplasty, which grew fast especially in the US, where the centre of gravity of the community shifted both scientifically and technologically.

Also, as the technique developed, a number of complementary and competing solutions were found in the community gathered around Gruentzig. From there, these new solutions found their way to the market very often through the activities of a number of physician-entrepreneurs such as John Simpson, inventor of the steerable guidewire, who founded no fewer than ten product-innovation based start-ups and then sold them in rapidly expanding knowledge markets to larger device manufactures (Guidant, Abbott Laboratories and Boston Scientific among others). The early development of stents is a case in point. A patent analysis by Xu *et al.* (2012) over the period 1984–1994 showed that small privately held companies created by physician inventors contributed the most patents as well as the most highly cited patents. The ten most highly cited patents in the field were dominated by two private firms, Expandable Grafts Partnership associated with Julio Palmaz and Richard Schatz (the Palmaz-Schatz stent that was licensed to Johnson & Johnson), and Cook Inc. who had an established relationship with pioneering interventional radiologist Cesare Gianturco (the Gianturco-Roubin Flex-Stent) (Xu *et al.* 2012). Both of these were the first stents to be approved by the FDA.

If we consider the overall evolution of the market for PCI devices, we find again not simply a process of growth but a process of punctuated qualitative change in the nature of the problems that are addressed and in the new ways found to solve them. Figure 2.2 proxies the dynamics by looking at the FDA device-approval records, which give a very good picture of the market output of the broader technological exploration that has been charted through extensive networks analysis in previous work (Mina *et al.* 2007; Ramlogan *et al.* 2007). The figure shows the (percentage) composition of FDA-approved products by type of innovations. It illustrates, first of all, the relative 'closure' of the technical problem solved by coronary balloon catheters, which over time accounts for proportionally less and less inventive efforts in favour of other types of catheters and



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

stent-related products. Secondly, it documents the rise of devices that address either complementary or subsequent challenges that arose through practice along the development of the problems-solutions sequence (or dominant technological trajectory) associated with the surgical treatment of the disease.

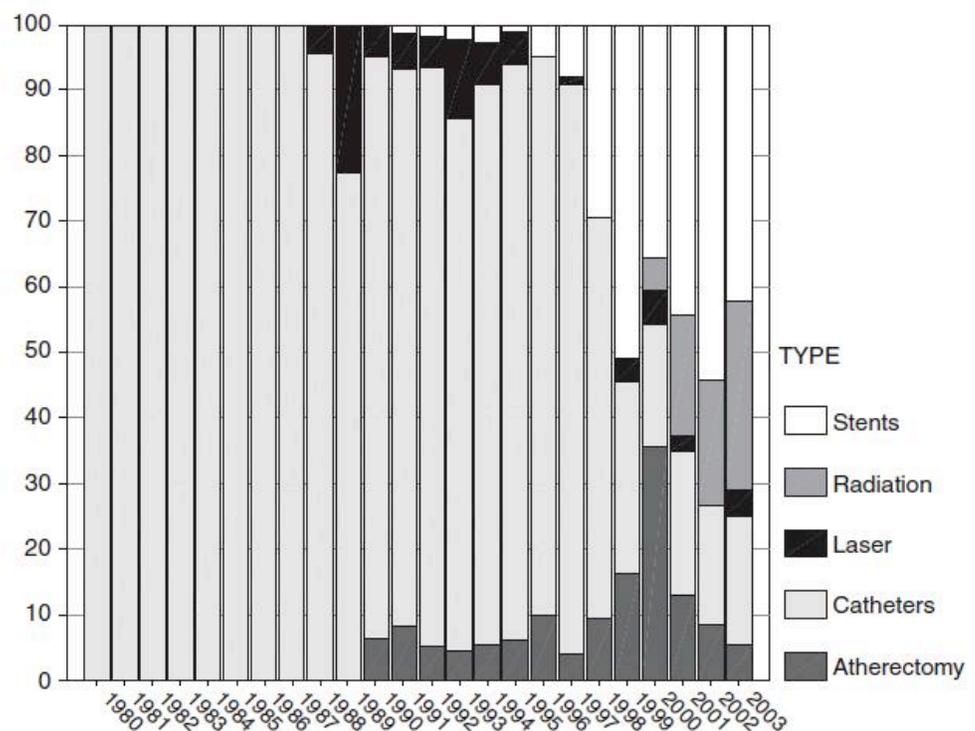


Figure 2.2 Percentage composition of FDA market approvals.

Source: Mina 2009, Fig. 8, p. 459.

Most interestingly, the transformation over time of the scientific and technological knowledge bases is deeply connected with the industrial transformation of the PCI medical device segment. In the earlier stages the industry focussed on the catheter and stent areas, where dominant designs soon emerged. Subsequently, new windows of opportunities opened up along the unfolding technological trajectory: companies from outside the established supply chain with radically different competence bases (for example in the areas of radiation or laser technology) started to challenge the market leaders and began to account for an expanding proportion of PCI devices. The resulting picture is, on the one hand, one of a



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

Schumpeterian process of competition through which incumbents are continuously challenged by innovators and, on the other, a highly distributed innovation system where some firms specialise in complementary market segments while others compete for alternative solutions along the problem sequence.

Conclusion

The development of PCI is a story of how the creative vision of a few lead developers and users became shared by an increasing number of practitioners and spilled over, on the one hand on the need for market and non-market co-ordination mechanisms and, on the other, on the formation of growing profit expectations. Insofar as the advancement of codified knowledge is concerned there is a close correspondence between the development of scientific understanding of the disease and the technological tools used to diagnose and treat the conditions, which is a clear mark of the strong complementarities (but not of linear dependency) between basic and applied medical research. Furthermore, as science and technology co-evolved, so did the set of interdependent problems that emerged along the search for satisfying cures to the disease because, as one problem was solved, others presented themselves, partly in reaction to the newly found solutions, and were opportunities for further change in the nature and composition of the product market.

Certainly a great number of factors contributed to the emergence and diffusion of coronary angioplasty jointly with the exceptional skills of the pioneers, the unpredictable consequences of early successes and the organisation, expansion and renewal of the supply capacity of the sector (Mina *et al.* 2007). At least two other factors were of the utmost importance in enabling the development of minimally invasive therapeutics for coronary artery disease: the concomitant growth of the US venture capital market and the effective institutionalisation of demand for coronary angioplasty. The lion's share of technological breakthroughs was launched since the early 1980s by VC-backed companies and the phenomenon of serial entrepreneurship is a constant pattern of innovation in the field. Overall, it has been estimated that over the past 20 years the contribution of venture capital to the sector amounted to almost 15 billion dollars in more than 2,000 deals for the development of minimally invasive treatments for cardiovascular disease.

With respect to the market need, the order of magnitude of the problem is self-explanatory: heart diseases are the leading cause of death in Western economies and also the costliest among all major preventable chronic diseases. As



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

far as the early introduction of PTCA is concerned, the new technique offered a less invasive and therefore less risky alternative to the more established (but by no means uncontroversial) coronary bypass graft surgery (CABG). However, what initially appeared as a cost-effective measure did not prove so cost-effective once, through practice, it became clear that it was associated with the problem of restenosis. This meant iterations of the procedure, which in turn entailed additional costs and the consequent erosion of the productivity improvement of PTCA. It was only through further improvement through learning and additional technical change, above all the introduction of stents, that the cost-effectiveness of the procedure dramatically improved. This could not possibly have been foreseen at the time of introduction of PTCA. Furthermore, by gradually improving performance, over time incremental technical change radically transformed PTCA from a complement to CABG to a substitute for it (Cutler and Huckman 2003). Again, this could not have been foreseen at the time of early diffusion of PTCA. Needless to say, the consequences of this are far reaching for economic analysis because they question the value of short term cost-benefit analyses and the opportunity costs of neglecting the fundamental uncertainty associated in the early stages with what have become, in the long run, star performers of the health technology sector.

Overall, medical progress in this clinical area takes the form of a series of unfolding trajectories of change that converged and diverged along specific sequences of problems. These can only be appreciated *ex post* as the result of a process of path-dependent emergence through multi-level mechanisms of selection weeding out over time knowledge that is judged as ineffective through clinical use by the relevant communities of practitioners, who typically operate across the boundaries between science, technology and clinical practice, especially in early phases of development of new treatments.¹³ We have also found that the sequence of scientific and technological problems solved along the main trajectory of development, itself an outcome of the dispersed and coordinated interaction of different agents and institutions, bears important implications for the structure and the composition of the market: this not only grew over time, but was also crucially and continuously transformed by new knowledge accumulated over time in the fight against the disease through a co-evolutionary process of epistemic and institutional change.

Acknowledgements



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

Figure 2.1 reprinted from Mina, A., Ramlogan, R., Tampubolon, G. and Metcalfe, J. S. (2007) 'Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge', *Research Policy*, Vol 36, pp. 789–806. Copyright 2007, reproduced with permission from Elsevier.

Figure 2.2 reprinted from Mina, A. (2009) 'The emergence of new knowledge, market evolution and the dynamics of micro-innovation systems', *Economics of Innovation and New Technology* 18(5): 447–466, Figure 8, p. 459. Copyright 2009. Reprinted by permission of Taylor & Francis Ltd, <http://www.tandfonline.com>.

Notes

1. We would like to extend our deep appreciation and thanks to: Dr Luigi Venetucci and Professor Anthony Heagarty (Manchester Medical School); Dr Bernard Clarke (Manchester Royal Infirmary); Professor Spencer B. King III (Fuqua Heart Center, Atlanta); Martin B. Leon MD and Eric A. Rose MD (Columbia Presbyterian Hospital, New York); Professor Annetine Gelijns (and her colleagues at the International Center for Health Outcomes and Innovation Research, Columbia University, New York). They have all been extremely generous with their time, advice and support and have helped to sharpen our understanding of the development of coronary angioplasty.
2. A survey of general internists actively involved in patient care by Victor Fuchs and Harold Sox in 2001 ranked coronary angioplasty 3rd (only behind MRI and CT scanning and ACE inhibitors) of the 30 most important medical innovations over the last 25 years.
3. The earlier literature refers to this technique as Percutaneous Transluminal Coronary Angioplasty (PTCA).
4. The use of nitrates (and nitroglycerin) dated back to the mid- to late-1800s. These provided transient relief by dilating vein and arteries thereby enabling more blood to get to the heart.
5. See, for example, Connolly (2002) for an account of the development of bypass surgery.
6. The Mullins and Lipscomb (1977) review of the literature pointed to incomplete analyses and less than ideally designed studies as few randomised studies were available at that time.
7. One of the interesting features of the spread of bypass surgery is how quickly it was taken up for cohorts of patients for which the medical



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

evidence was not indicated. Anderson and Lomas (1988) expressed concern about the apparent change in clinical policies towards the use of bypass procedures in the elderly without solid evidence on efficacy or cost effectiveness.

8. This account draws in part on King (1996, 1998) and authors' original interviews with him on 10 May, 2005.
9. See <http://clinicaltrials.gov/ct2/show/study/NCT00005677>
10. See also Interview with John Abele (Cohen and Klepper 1996), Part III, available at www.ptca.org/nv/interviewframe.html
11. Interview with Heliane Canepa (Cohen and Klepper 1996), then President of Schneider Worldwide and former collaborator of Gruentzig. Available online at: www.ptca.org/nv/interviewframe.html
12. Interview with John Abele (Cohen and Klepper 1996). Available online at: www.ptca.org/nv/interviewframe.html
13. This does not imply that the outcomes are necessarily optimal or could not have been different!

References

- Anderson, G. M. and Lomas, J. (1988) 'Monitoring the diffusion of a technology: coronary bypass surgery in Ontario'. *American Journal of Public Health*, Vol. 78, pp. 251–254
- Balcon, R., Beyar, R., Chierchia, S., De Scheerder, I., Hugenholtz, P. G., Kiemeneij, F., Meier, B., Meyer, J., Monassier, J. P. and Wijns, W. (1997) 'Recommendations on stent manufacture, implantation and utilization'. *European Heart Journal* 18, pp. 1536–1547
- Bavry, A. A., Kumbhani, D. J., Helton, T. J., Borek, P. P., Mood, G. R. and Bhatt, D. L. (2006) 'Late thrombosis of drug-eluting stents: a meta-analysis of randomized clinical trials'. *Am J Med.* 119, pp. 1056–1061
- British Heart Foundation (2004) 'Coronary Heart Disease Statistics'. Available online at <http://www.heartstats.org/uploads/documents%5C2004pdf.pdf>
- Cohen, W. and Klepper, S. (1996) 'Firm size and the nature of innovation within industries'. *Review of Economics and Statistics* 78(2), pp. 232–243
- Colombo, A., Hall, P., Nakamura, S., Almagor, Y., Maiello, L., Martini, G., Gaglione, A.,



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

- Goldberg, S. L. and Tobis, J. M. (1995) 'Intracoronary stenting without anticoagulation accomplished with intravascular ultrasound guidance'. *Circulation* 91, pp. 1676–88
- Connolly, J. E. (2002) 'The development of coronary artery surgery: personal recollections'. *Texas Heart Institute Journal*, 29(1), pp. 10–14
- Cowley, M. J., Mullin, S. M., Kelsey, S. F., Kent, K. M., Gruentzig, A. R., Detre, K. M. and Passamani, E. R. (1985) 'Sex differences in early and long-term results of coronary angioplasty in the NHLBI PTCA Registry'. *Circulation*, 71, pp. 90–97
- Cutler, D. M. and Huckman, R. S. (2003) 'Technological development and medical productivity: the diffusion of angioplasty in New York state'. *Journal of Health Economics*, 22(2), pp. 187–217
- Dangas, G. and Kuepper, F. (2002) 'Restenosis: repeat narrowing of a coronary artery: prevention and treatment'. *Circulation*, 105, pp. 2586–2587
- Dosi, G. (1982) 'Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change'. *Research Policy*, Vol. 11, pp. 147–162
- Eckhurst, E., Kappenberger, L. and Goy, J. L. (1996) 'Stents for intracoronary placement: current stats and future directions'. *JACC*, Vol. 27, pp. 757–765
- Farooq, V., Gogas, B. D. and Serruys, P. W. (2011) 'Restenosis: delineating the numerous causes of drug-eluting stent restenosis'. *Circ Cardiovasc Interv*, 4, pp. 195–205
- Fischman, D. L., Leon, M. B., Baim, D. S., Schatz, R. A., Savage, M. P., Penn, I., Detre, K., Veltri, L., Ricci, D., Nobuyoshi, M., Cleman, M., Heuser, R., Almond, D., Teirstein, P. S., Fish, R. D., Colombo, A., Brinker, J., Moses, J., Shalmon, A., Hirshfeld, J., Bailey, S., Ellis, S., Rake, R. and Goldberg, S. for the Stent Restenosis Study investigators (1994) 'A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease'. *New England Journal of Medicine*, 331, pp. 496–501
- Fuchs, V. R. and Sox Jr. H. C. (2001) 'Physicians' views of the relative importance of thirty medical innovations'. *Health Affairs*, Vol. 20, No 5
- Gelijns, A. and Rosenberg, N. (1994) 'The dynamics of technological change in medicine'. *Health Affairs*, Summer, pp. 28–46



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

- Gelijns, A. C., Rosenberg, N. and Moskowitz, A. J. (1998) 'Capturing the unexpected benefits of medical research'. *New England Journal of Medicine*, Vol. 339, pp. 693–698
- Gelijns, A., Sivin, J. G. and Nelson, R. R. (2001) 'Uncertainty and technological change in medicine'. *Journal of Health Politics, Policy and Law*, Vol. 26, pp. 913–924
- Hamid, H. and Coltart, J. (2007) "Miracle stents": a future without restenosis'. *Mcgill J Med*. 10(2), pp. 105–111
- Kent, K. (2010) 'The history of interventional cardiology' in S. Redwood, N. Curzen, and M. Thomas (Eds) *Oxford Textbook of Interventional Cardiology*. New York: Oxford University Press, p. 22
- King, S. B. III (1996) 'Angioplasty from bench to bedside to bench'. *Circulation*, Vol. 93, pp. 1621–1629
- King, S.B. III (1998) 'The development of interventional cardiology'. *Journal of the American College of Cardiology*, Vol. 31, Supp. B, pp. 64B–88B
- Li, J. and Kozlek (2012) 'The future of the drug eluting stent market: rebound will be driven by new, ground-breaking technologies and clinical confidence'. *Scientia Advisors Blog*. Available online at: www.scientiaadv.com/blog/2012/08/07/the-futureof-the-drug-eluting-stent-market
- Mahmoudi, M., Delhay, C. and Waksman, R. (2011) 'Safety and efficacy of drug-eluting stents and bare metal stents in acute coronary syndrome'. *Cardiovascular Revascularization Medicine* 12, pp. 385–390
- Metcalfe, J. S., James, A. and Mina, A. (2005) 'Emergent innovation systems and the delivery of clinical services: the case of intra-ocular lenses'. *Research Policy* 34(9), pp. 1283–1304
- Mina, A., Ramlogan, R., Tampubolon, G. and Metcalfe, J. S. (2007) 'Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge'. *Research Policy* 36(5), pp. 789–806
- Mullins, C. B. and Lipscomb, K. (1977) 'Critique of coronary artery bypass surgery'. *Annual Review of Medicine*, Vol. 28, pp. 271–289
- Mullins, S. M., Passamani, E. R. and Mock, M. D. (1984) 'Historical background of the National Heart, Lung and Blood Institute Registry for Percutaneous Transluminal Coronary Angioplasty'. *The American Journal of Cardiology*, Vol. 15,



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

pp. 3C-6C

Myler, R. (2002) 'Coronary and peripheral angioplasty: historical perspective' Chapter 7 in Topol, E. (Ed.) *Textbook of Cardiovascular Medicine*. London: Lippincott Williams & Wilkins

Office of Technology Assessment (OTA) (1978) 'Assessing the efficiency of safety of medical technologies'. US Government Printing Office

Pepine, C. J., Hirshfeld, J. W., Macdonald, R. G., Henderson, M. A., Bass, T. A., Goldberg, S., Savage, M. P. Vetrovec, G., Cowley, M. and Taussig, A. S. (1990) 'A controlled trial of corticosteroids to prevent restenosis after coronary angioplasty'. *Circulation* 81, pp. 1753–1761

Ramlogan, R., Mina, A., Tampubolon, G. and Metcalfe, J. S. (2007) 'Networks of knowledge: the distributed nature of medical innovation'. *Scientometrics* 70(3), pp. 459–489

Reul, R. M. (2005) 'Will drug-eluting stents replace coronary artery bypass surgery?' *Tex Heart Inst J.* 32(3), pp. 323–330

Rosch, J., Keller, F. S. and Kaufman, J. A. (2003) 'The birth, early years and future of interventional radiology'. *Journal of Vascular and Interventional Radiology*, Vol. 14, pp. 841–853

Ruygrok, P. N. and Serruys, P.W. (1996) 'Intracoronary stenting: from concept to custom'. *Circulation*, 94(5), pp. 882–90

Serruys, P. W., de Jaegere, P., Kiemeneij, F., Macaya, C., Rutsch, W., Heyndrickx, G., Emanuelsson, H., Marco, J., Legrand, V., Materne, P., Belardi, J., Sigwart, U., Colombo, A., Goy, J. J., van den Heuvel, P., Delcan, J. and Morel, M. A. for the Benestent group (1994) 'A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease'. *New England Journal of Medicine*, 331, pp. 489–95

Serruys, P. W., Kutryk, M. J. and Ong, A. T. (2006) 'Coronary-artery stents'. *New England Journal of Medicine*, 354, pp. 483–495

Simpson, J. B., Bains, D. S., Roberts, E. W. and Harrison, D. C. (1982) 'A new catheter system for coronary angioplasty'. *American Journal of Cardiology*, Vol. 49, pp. 1216–1222

Towers, M. and Davies, S. (2000) 'Cardiac catheterisation' in M. E. Silverman, P. R.



CORONARY ARTERY DISEASE

Excerpted from *Medical Innovation*

Flemming and A. Hollman (Eds) *British Cardiology in the 20th Century*. London: Springer-Verlag, p. 151

Utterback, J. M. (1994) *Mastering the Dynamics of Innovation*. Boston, MA: Harvard Business School Press

Whitlow, P. L. (1997) 'Is rotational atherectomy here to stay?' *Heart* (Supplement 2), 78, pp. 35–36

Wilson, W. M. and Cruden, N. L. M. (2013) 'Advances in coronary stent technology current expectations and new developments'. *Research Reports in Clinical Cardiology*, 4, pp. 85–96

World Health Organization (WHO) (2011) *Global Status Report on Non-communicable Diseases 2010*. Geneva: World Health Organization

Xu, S., Avorn, J. and Kesselheim, A. S. (2012) 'Origins of medical innovation: the case of coronary artery stents'. *Circulation Cardiovascular Quality and Outcomes*, 5: pp. 743–749



CHAPTER

5

KENYA'S HEALTH INNOVATION CAPACITY

Reconfiguring Global
Health Innovation

Padmashree Gehl Sampath

 Routledge International Studies in Health Economics

This chapter is excerpted from
Reconfiguring Global Health Innovation
by Padmashree Gehl Sampath

© 2011 Taylor & Francis Group. All rights reserved.



[Learn more](#)



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

5.1 Introduction

The Kenyan economy registered a GDP growth of 6.9 per cent in 2007 and has been in the process of recovering from one of the country's longest recessions (World Bank Indicators, 2007). Despite its relatively good economic performance over the past few years, the country ranks low on knowledge indicators measured in terms of secondary and tertiary schooling, researchers in R&D per million of population and domestic credit institutions. Kenya has one of the lowest scores on the human development index within sub-Saharan Africa and the country's performance on life expectancy, education levels and decent living standards *have remained the same or worsened over the past 15 years*. The Ministry of Health provides public health care services and controls 52 per cent of all medical facilities available in the country (Export Processing Zones Authority, 2005).

Most biotechnology-led research in the country has focused on agricultural biotechnology, where the initial developments can be traced back to the early 1980s with the application of tissue culture in crops such as citrus fruits (at the Kenyan Agriculture Research Institute, KARI) and Pyrethrum (University of Nairobi). Around the same time, the use of biotechnology in livestock and health research also began with a focus on the generation of disease diagnostic technologies employing hybridoma and DNA molecular techniques. A major breakthrough was the Rinderpest Vaccine that was jointly developed by KARI, Pirbright (UK) and the University of California (Gichuki, 2006).

Despite these successes, the majority of Kenya's research projects in health biotechnology in Kenya are concentrated in the public sector, involving mainly the public research institutes, often with donor funding (Clark *et al.*, 2005; Kireia *et al.*, 2003). Health innovation in general, and health biotechnology in particular, has been slow to take off, despite the fact that Kenya has a relatively high amount of private pharmaceutical enterprises in the East African region.

This chapter provides an analysis of the current status of health and biotechnology innovation in Kenya and evaluates the relevant policy support that needs to be provided in order to enable innovation activities. The data contained in this chapter is based on a sector survey of health innovation that followed the same methodology as the other chapters in Part II. In some places, the information and analysis is substantiated by additional information that was conducted in 2007 in Kenya on biotechnology research across both agricultural and health biotechnology areas.¹ In total, 41 firms, 32 hospitals and 25 public research



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

institutes were surveyed (98 in total) in addition to 45 governmental agencies. In addition to this, 65 people including top level management in firms (both multinational and local), governmental officials, directors and research staff in PRIs and hospitals were interviewed through face-to-face interviews.

5.2 Key actors in health innovation

Kenya's innovation infrastructure comprises of a strong base of public sector institutions dedicated to research and standard setting, such as the Kenya Industrial Research and Development Institute (KIRDI), Kenya Bureau of Standards, Kenya Intellectual Property Institute (KIPI). Several other research institutes exist in order to conduct research on sectoral priorities, including health biotechnology. It has six main universities, of which the Jomo Kenyatta University and the University of Nairobi are the most important, apart from a range of polytechnic institutions that impart technical education in various science intensive disciplines.

As is the case in most latecomer innovation systems, a large part of Kenya's innovation activities in health are concentrated in the public sector, although there is a relatively vibrant private sector in the country performing manufacturing, distribution and retailing activities in pharmaceuticals.

5.2.1 Governmental agencies

The creation of the Kenya Industrial Property Office in 1990 (and its transformation into KIPI with greater decision-making power and authority) following the enactment of the industrial property Act was a major strengthening act as far as IPP in Kenya is concerned. The National Council on Science and technology in Kenya has a broad mandate of focusing on innovation and new technologies of importance to the country, such as biotechnology and ICTs. There are two other health agencies that are expected to promote entrepreneurship and standard setting: the Pharmacy and Poisons Board and the Kenya Bureau of Standards. The Pharmacy and Poisons Board is responsible for the regulation of trade in health products, including those employing health biotechnology. It prescribes the apparatus and also oversees compliance with good manufacturing practices, labelling and other formal requirements. The Kenya Bureau of Standards (KEBS) operates under the Ministry of Trade and Industry, and is responsible for setting standards for weights and measures, purity and identity. KEBS is the



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

national standards body and is established under the Standards Act (cap 496) laws of Kenya. This Act of 1974 seeks to promote and provide for standardization of commodities and a code of practice.

5.2.2 Research institutes and universities

The Kenya Medical Research Institute (KEMRI) is the premier R&D organization with a national mandate to “conduct health sciences research and generate research findings applicable towards improvement of health status”. KEMRI was established in 1979 as one of the five public research institutes under the Science and Technology Act (cap 250) laws of Kenya. The institute has ten research centres located in various parts of the country meant to address various aspects of health research including health biotechnology. These centres have over 800 scientific and technical staff with over 200 biomedical scientists in such fields as microbiologists, clinicians, social scientists, pharmacists, epidemiologists, immunologists, virologists. The technical staff includes laboratory technologists, public health and clinical officers and pharmaceutical technologists (KEMRI website). KEMRI has developed at least two products namely *Hepcell* which is a diagnostic kit for detecting human hepatitis B surface antigen and *particle agglutination*, which is also a diagnostic kit for the same, developed using locally produced reagents, doesn't require electric power and its results can be viewed using the naked eye. The development of both these products has been supported by the Japan International Cooperation Agency, which has also funded the product commercialization facility at KEMRI.

KIRDI is a technology development and management agency of the government, which operates under the Ministry of Trade and Industry. It was established and incorporated under the Science and Technology Act (cap 250) in 1979 with an intention similar to that of KEMRI. The institute has several objectives that are important for all innovation activities including those in health biotechnology, such as enhancing the national industrial technology innovation process as a strategy towards rapid socio-economic development, contributing to the development of sufficient capacity for industrial research and development, contributing to the creation of the national wealth in disembodied technologies that are appropriate and accessible to micro and small enterprises in Kenya, and facilitating access by local enterprises to Business Development Services including cleaner production and industrial information. These two institutes are supported through a range of smaller research institutes and international research institutes



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

operating in Kenya such as African Insect Science for Food and Health (ICIPE) and local universities and firms. KARI has a well-established biotechnology tradition, which also lends strength to health innovation through biotechnology inputs.

5.2.3 Private sector

Kenya's local pharmaceutical sector is made up of firms engaged in production and manufacturing, retailing and distributing of health products in the local market as well as in the COMESA economic region. The range of pharmaceutical products manufactured in Kenya includes antibiotics, anti-malarials, anti-amoebics, analgesics, anti-diarrhoeal, antacids, tranquilizers, anti-spasmodics, vitamins and anti-ulcer treatment, according to the Export Processing Zones Authority (2005). The large Kenyan firms that are involved in production are mostly subsidiaries of large MNCs, such as Bayer, Aventis and Novartis with the exception of some local firms such as Macs Pharmaceuticals and Cosmos Pharmaceuticals.

Box 5.1 Important pharmaceutical firms in Kenya

- Alpha Medical Manufacturers
- Aventis Pasteur SA East
- Bayer East Africa Limited
- Beta Healthcare (Shelys Pharmaceuticals)
- Cosmos Limited
- Dawa Pharmaceuticals
- Didy Pharmaceutical
- Diversey Lever
- Eli-Lilly (Suisse) SA – Nairobi
- Elys Chemical Industries Ltd
- GlaxoSmithKline – Nairobi
- High Chem East Africa Ltd
- Mac's Pharmaceutical Ltd
- Manhar Brothers (Kenya) Ltd
- Novartis Rhone Poulenc Ltd
- Novelty Manufacturers Ltd
- Pfizer Corp (Agency)
- Pharmaceutical Manufacturing Co (K) Ltd
- Pharmaceutical Products Ltd
- Regal Pharmaceutical Ltd
- Universal Pharmaceutical Ltd

Source: author's survey, 2007



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

5.2.4 NGOs/donors and teaching hospitals

Several international NGOs, like the African Medical and Research Foundation and ICIPE, are actively engaged in conducting research on local health problems like Rift Valley Fever and Sleeping Sickness. Kenya has a large number of such hospitals led by Agakhan University Hospital and the Moi University School of Medicine and Referral Hospital.

5.3 Current strengths in health innovation

Health innovation in the country is relatively successful when compared to other countries in East Africa (see Tanzania in Chapter 6). One of the most important successes of the local system of innovation is the presence of private sector enterprise. While several firms only distribute and are engaged in retailing, some local firms are actively engaged in production and R&D of drugs. Of note in this context is Cosmos Pharmaceuticals, a local firm that is engaged in supplying several ARV drugs (for treatment of HIV/AIDS) to many other countries in the region and elsewhere, and anti-TB drugs to international procurement agencies like the Global Fund (PRM Database, Global Fund, July 2008).

While Cosmos has had some problems in accreditation for ARV drugs (they remain, however, one of the earliest producers of the same in sub-Saharan Africa), they are the fifth largest suppliers of anti-TB drugs and an important supplier of anti-Malarial drugs to countries that receive Global Fund (GFATM) grants worldwide (Grace and Gehl Sampath, 2008). Cosmos is a local success story wherein the company has constantly sought to upgrade its ability to reverse engineer and produce drugs of importance to global and local public health. Apart from drugs mentioned here, Cosmos produces a range of local pharmaceutical products for the Kenyan and COMESA market (field interviews). Another company that deserves mention here is Universal Corporation, again a local firm that has invested enormous efforts in upgrading its pharmaceutical manufacturing capabilities and has commenced the WHO prequalification/accreditation process. It has also established an R&D unit for natural products and has hired an expert from India to assist them in this direction.

5.4 Local push factors in the sectoral system for health



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

The local push factors are relatively weaker when compared to India and Bangladesh analysis conducted in the previous two chapters. However, Kenya still remains one of the most entrepreneurial sectors in health innovation in the region demonstrating capacity not only for health innovation but also for biotechnology based work (Gehl Sampath and Oyeyinka, 2009a, b). The push factors that are responsible for capacity are not so much a coherent sectoral policy vision, as much as local demand for medicines in the Kenyan market as well as the COMESA region, both of which are discussed here. A third factor, collaborations and linkages (especially international collaborations), play a major role in shaping capacity in the sector, which is discussed in Section 5.5. However, as this section shows, there are significant obstacles to building capacity locally.

5.4.1 Lack of human skills of relevance to health innovation

The university disciplines that deal with various aspects of health innovation, such as pharmacology, medical sciences and newer technologies that could be integrated into health innovation such as biotechnology, are in need of revision. Our surveys also sought to ascertain the kinds of capacity being created to perform biotechnology research. We found that the limited capacity that is being created is focused more on tangible infrastructure (such as laboratories and equipment) and is not matched by the expansion of human skills to utilize these facilities as part of structured research agendas for both health and agricultural biotechnology. This once again confirms earlier results on Kenyan biotechnology capacity (see for example the survey by Odame *et al.*, 2002), which concluded that over the years capacity building has focused more on hardware components (expanding the physical facilities) and not so much on post-graduate training in MSc and PhD levels. This over-emphasis on hardware components has resulted in an increase on the demand for non-scientific staff to manage the expanded physical facilities, thus explaining the low scientific–non-scientific staff ratios captured in our surveys that covered all five public sector institutes.

Table 5.2 presents the net enrolment from 2000 to 2005, and shows that only half of the amount of students that attend primary school eventually progress to secondary schooling and, alarmingly, only 3 per cent of the total receive tertiary education. The survey also found that the share of staff with PhD degrees in the public research institutes was extremely small, again confirming the findings that most of the “core research” staff required to conduct biotechnology based R&D are missing from the expanding infrastructure endowments (Oyeyinka and Gehl



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

Sampath, 2009b). We found that most of the biotechnology work was concentrated around tissue culture and other basic biotechnologies, rather than seeing pockets of excellence across a broader range of expertise, given the broad range of international collaborations going on in the country.²

Table 5.1 Animal health biotechnologies in Kenya (2006)

Technology	Description	Year	Status
Latex agglutination diagnostic test for Contagious Caprine Preural Pneumonia (CCPP)	A quick test for identifying goats that have been exposed to CCPP organism. The test works with either whole blood or serum and is therefore a good field test. The reagent also keeps well in the absence of refrigeration.	1985	Ready for commercialization, but not yet commercialized.
Latex agglutination diagnostic test for Contagious Bovine Preural Pneumonia (CBPP)	A quick test for identifying cattle that have been exposed to CBPP organism. The test works with either whole blood or serum and is therefore a good field test.	1990	The test was undergoing more evaluation.
Latex agglutination diagnostic test for Lumpy Skin Disease	A quick test for identifying cattle that have been exposed to lumpy skin disease virus. The test works with either whole blood or serum and is therefore a good field test.	2003	
ELISA based diagnostic test for Contagious Caprine Preural Pneumonia	The technique uses a unique carbohydrate as an antigen for coating ELISA plates. It is therefore a very specific test.	2005	It is a new test that needs more validation before being availed for wider use.
RT-PCR diagnostic test for Rift Valley Fever	The technique uses non-structural gene as a template for amplification.	2003	Needs more validation.
Antigen detection test for Nairobi Sheep disease virus	A test for quick detection of Nairobi sheep disease virus without going through tissue culture inoculation.	2004	Under development.
Recombinant capripox/Rift Valley Fever vaccine	A sheep ox virus vaccine which expresses a glycoprotein gene from Rift Valley Fever virus. When inoculated into the sheep, it induces an immune reaction which protects against challenge with Rift Valley Fever virus.	1998	Further evaluation in cattle (calves) and goats in containment.
Inactivated vaccine for Nairobi sheep disease	Nairobi sheep disease virus inactivated with beta-propiolactone or formaldehyde.	2002	Under development.
Vaccinia/Rinderpest Recombinant Vaccine	The vaccines were found effective in confined trials.		The vaccine has undergone confined trials at KARI-Kiboko.

Sources: field interviews; Gichuki (2006).

In recent years, the government has sought to address the issue of adequate human and technical capacity by establishing courses in biotechnology in most of the public universities in Kenya. In fact all the six public universities across the country are offering biotechnology courses at under-graduate and post-graduate levels. For example, Kenyatta University offers both BSc and MSc courses in



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

Biotechnology. The University of Nairobi in 2005 established the Centre for Biotechnology and Bioinformatics as a centre of excellence to facilitate capacity building and generate marketable products by harnessing biotechnology. The centre's mandate is to enhance knowledge and skills in biotechnology and bioinformatics to impact on agricultural and industrial output, health and environmental management.

Table 5.2 School enrolment ratios, Kenya 2000–2005

<i>School enrolments (% net)</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>
Primary enrolment	67	–	63	77	76	–
Secondary enrolment	33	34	35	37	40	–
Tertiary enrolment	3	3	3	–	3	–

Source: World Development Indicators, 2007.

The centre formation underscores the realization that biotechnology is an interdisciplinary subject with wide ranging applications of scientific and engineering principles in different fields such as agriculture, food and feed, medicine, industry and the environment, which are of profound importance to mankind. The centre aims to strengthen national capabilities in the field of basic sciences and technology and in the development of research in biotechnology and bioinformatics, in addition to promote and conduct basic research in the areas of molecular biology, biotechnology and bioinformatics. It also has the mandate of facilitating the application of biotechnology in research and encourages its use for the development of marketable products and to offer training facilities for manpower development in biotechnology and bioinformatics at the national and regional level. The centre is also expected to institutionalize links between universities, scientific research institutions and the private sector in order to promote the focus on product development, and to network with institutions in developed and developing countries as well as the international centres of biotechnology and bioinformatics.

It is expected that the integration of biotechnology courses within the universities' curricula as well the emergence of training centres such as the Centre for Biotechnology will help boost the country's manpower and skills to conduct biotechnology research and innovation in the years to come. The bigger challenge however is to equip these centres with the necessary training materials and



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

facilities, and relevant manpower, which is very different from merely establishing organizations. Meeting this greater challenge will require that Kenya begins to allocate more of its GDP to science and technology activities (which is currently less than 0.5 per cent of its GDP as noted earlier). It also calls for universities and research institutes to seek alternative, innovative funding mechanisms, the incentives for which are presently not in place.

A National Consortium of National Health Research Kenya (see online, available at: <http://cnhrkenya.org/>), seeks to address some of these issues particularly of relevance to health and health biotechnology. Its main function is to create collaborative linkages between all health research institutions and universities' centres of excellence, governmental agencies (including the Ministry of Health) and private sector enterprise in the country. The Consortium's primary focus is on enabling health R&D capacity to be built in an innovative manner.

5.4.2 Lack of knowledge infrastructure in public sector institutions

Although the local research institutes have been involved in several projects that involve more demanding technologies for health innovation both animal and human (as shown in Table 5.1), none of these projects have led to the commercial products within the country and most of the products were undergoing evaluations at the time of the survey in 2007. Organizations like KIRDI and KEMRI, which have the mandate to develop technologies for the use of local entrepreneurs in both traditional and new technology sectors, operate with extreme staffing and funding shortages, and hence are not able to fulfil their mandates. It is noteworthy here that although at a first glance, there seems to be a lot of technical expertise in these organizations – for example, the ten research centres under KEMRI are reported to have around 800 technical staff – these figures are often misleading. The personnel reported as “technical staff” cover a host of functions outside the traditional biotechnology R&D domain, and includes public health services and clinical officers. The surveys also found that often the few researchers who are skilled to perform biotech and health R&D in these centres are diverted into other tasks that are not so research and funding intensive mainly due to lacking capacity of the government to endow these centres appropriately to carry out their original mandates. Most researchers at universities and public research institutes complain regularly of a lack of funds and initiative on part of the government to support and strategically direct locally important research in health and biotechnology. The extraordinary reliance on external donor funding for research, which is at best



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

sporadic and not dependable, means that innovative activities in academic institutions in the country continue at a rate that hardly reflects its true potential. The development of health innovation has been stunted by the absence of relevant knowledge infrastructure. The amount of resources that are allocated to public sector organizations for research is negligible.

In particular, the survey shows that the universities fare much worse than public research institutes on a number of fronts. Most researchers and scientists in universities spend over 90 per cent of their time in teaching and administration and suffer from severe intellectual isolation (as a result of lack of opportunities to interact with research institutes and the private sector) as well as lack of mandate. This lack of mandate is a problem common amongst several African latecomer countries and can be traced back to stagnating academic systems that demonstrate inertia to change and revision that could make it more current. The interviews and the survey data show that the universities have much more difficulty gaining access to international funding. Most university staff interviewed complained that the technical assistance to research by donors almost always get channelled through the government into public research institutes and do not find their way to university centres of excellence. This makes the issue of funding even more of a challenge for the universities. The interviewees also expressed the need for the National Council of Science and Technology to review and understand the university needs in this respect (field interviews).

5.4.3 Lack of institutional support to private sector R&D

A majority of the private sector enterprise in the country is engaged in importing, distributing and retailing of pharmaceutical products while some firms manufacture. While there is no reliable data on how much of the private sector is engaged in manufacturing as opposed to retailing and distributing, Kenya fares relatively well in comparison to other countries in the region (and in sub-Saharan Africa as a whole) On the whole, the picture for private sector health R&D as captured by the survey looked rather bleak.

Despite the intent of the National Biotechnology Policy of 2006, little has been done to promote the emergence of a vibrant private sector for health biotechnology. Discounting the few firms that conduct local R&D, engaging in health innovation activities is a tough task due for the sector. There is little governmental infrastructure for technological incubation, acquisition of technology required to pursue innovation and funding, especially small-firm risk sharing and



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

financing initiatives of the kind its local enterprise would need. Some existing incentives to promote local innovation, such as a governmental policy of promoting local production of pharmaceuticals through a procurement process that gives preference to local producers even if the local price is greater than that of exports up to a threshold of 15 per cent, is hardly implemented in practice thus not serving its purpose.

Table 5.3 Descriptive statistics: innovation and government assistance in Kenya

	<i>Number of firms</i>	<i>%</i>	<i>Number of PRIs</i>	<i>%</i>
<i>New product development</i>				
No	34	82.9	15	60.0
Yes	7	17.1	10	40.0
<i>New process development</i>				
No	29	70.7	21	84.0
Yes	12	29.3	4	16.0
<i>Government sponsored R&D programmes</i>				
No	41	100.0	17	68.0
Yes	0	0.0	8	32.0
<i>Government R&D subsidies</i>				
No	41	100.0	15	60.0
Yes	0	0.0	10	40.0
<i>No. of observations</i>	41		25	

Source: author's survey, 2007.

On a closer look, there appear to be inherent weaknesses in several institutions that are fundamental to the creation of new knowledge and the use of already existing knowledge in innovation activities. Financial support institutions to promote local innovation and entrepreneurship in Kenya have not been performing well and have been constantly on the decline since 2000. Reviewing the data available in the World Development Indicators Database (World Bank Indicators, 2007) for the years from 2001 to 2005, we found that domestic credit to private sector has been on the decline (from 28 per cent of total GDP in 2000 to 27 per cent of total GDP in 2005). Similarly, FDI remains close to zero (fluctuating between 1 per cent of GDP and less between 2000 and 2005). Data on other important indicators such as R&D expenditure as percentage of GDP and



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

researchers in R&D (per million people) are not available.

The innovation patterns of the private sector were captured by survey data, which show that the local public research institutes are more often involved in new product development than local firms while both groups of actors are equally involved in new process development. Table 5.3 contains the patterns of product and process innovations both in firms and PRIs, and the impact of governmental support and assistance. While the PRIs are involved more frequently on average in new product development than firms, both firms and PRIs are involved equally frequently in new process development. Governmental subsidies and R&D programmes, however, are mainly directed towards PRIs and not towards firms. As Table 5.3 shows, the percentage of PRIs that benefit from government sponsored R&D programmes is on average statistically and significantly larger than that of firms that benefit from government sponsored R&D programmes. A similar pattern is shown for government subsidies. Most of the actors interviewed observed that the level of involvement of the private sector in health innovation and product development is still very low and urged that the private sector should be encouraged by enacting appropriate policies and incentives.

Table 5.4 Impact of governmental incentives on firm and organization level innovation

<i>Policy incentive</i>	<i>Firms</i>	<i>PRIs</i>
1 Government innovation incentives	1.63	1.80
2 Skilled manpower	2.05	3.24
3 Local collaboration with universities	1.85	2.84
4 Local collaboration with PRIs	1.90	2.84
5 IPP	1.70	1.72
6 Quality of local innovation infrastructure	1.80	2.24
7 Venture capital	1.92	1.80
8 Government support to SMEs	1.70	–
9 Government–firm technology transfer	1.60	–

Source: Oyeyinka and Gehl Sampath, 2009b.

Table 5.4 contains the survey responses of firms and public research institutes towards various policy incentives and the extent to which they contributed towards health biotechnology innovation. The respondents were asked to rank on a scale of 1 (not helpful at all in innovation activities) to 5 (extremely helpful). Therefore, any ranking above 2.5 indicates a strong impact of the policy incentive in promoting



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

innovative activity.

Three points stand out from the survey responses in Table 5.4. First, public research institutes rate (and are relatively better endowed than firms as the surveys more generally show) on all innovation incentives presently implemented in the system. Second, hardly any of the conventional policy incentives that are needed to support innovation in health biotechnology in public and private sectors seem to play a significant role presently, underscoring the low level of health and biotechnology innovation activity in the system. Third, there is a very clear consensus on the importance of human skills as an innovation incentive for the sector. This, although intuitive in general, helps to capture the most important lacunae in the system as it presently stands.

As in the case of Bangladesh, the subsidiaries of multinationals are mainly engaged in local manufacturing of drugs across a broad range, but not necessarily those that are technologically demanding due to the difficulties of organizing labour, laboratory facilities and other infrastructure requirements (among other firm level strategic reasons). It is rather the local firms that have acquired the capacity to conduct API work and aspire to branch out into more demanding areas of health innovation that drive the process ahead. The factors that impede its growth are the unavailability of human capital, regulatory barriers to conduct biotechnology based research and lack of entrepreneurial support.

5.4.4 Mismatch of skills and resources in new product and process innovation

Table 5.5 presents descriptive statistics of full time employment in firms, PRIs and training and teaching hospitals, education level of staff and the proportion of staff engaged in R&D. It shows that the mean employment over the period 2001–2005 for firms is not statistically and significantly different (on average) from that observed in the PRIs and hospitals. Second, firms, hospitals and PRIs are statistically and significantly equally old. Furthermore, the staff working in the firms and hospitals is on an average less educated than those employed by PRIs. The percentage of staff working in private firms with a PhD and MSc degree is statistically and significantly smaller than that amongst PRIs with similar degrees, and the percentage of staff in firms that have only a BSc degree is statistically and significantly larger than that in PRIs. More specifically, the mean share of staff with a PhD and MSc degree is statistically and significantly larger in PRIs than in hospitals and firms. Furthermore, the mean share of staff with a BSc degree is statistically and significantly smaller in PRIs than in firms but larger than in



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

hospitals. Finally, mean R&D staff and the percentage of R&D performers are not statistically and significantly different across firms and PRIs. These figures are equal to zero in hospitals.

Table 5.5 Descriptive statistics: key actors of innovation

Variable	Mean	(Std dev.)	Mean	(Std dev.)	Mean	(Std dev.)	Mean	(Std dev.)
	Firms		Hospitals		PRIs		All	
Employment 2001–2005 (FTEs)	41.814	(66.858)	89.812	(258.01)	34.600	(44.164)	55.647	(155.502)
Age (in years)	12.561	(9.770)	14.813	(12.699)	8.760	(19.023)	12.327	(13.630)
% of staff with PhD	0.000	(0.000)	0.002	(0.009)	0.082	(0.139)	0.021	(0.078)
% of staff with MSc	0.015	(0.038)	0.007	(0.017)	0.103	(0.121)	0.035	(0.077)
% of staff with BSc	0.276	(0.288)	0.020	(0.052)	0.103	(0.159)	0.148	(0.275)
R&D performers 2001–2005	0.098	–	0.000	–	0.160	–	0.082	–
R&D personnel in 2001–2005	0.015	(0.087)	0.000	(0.000)	0.056	(0.161)	0.021	(0.100)
Number of observations	41		32		25		98	

Source: author's field survey, 2007.

Table 5.6 presents estimation results of a bivariate probit model that explains jointly the probability of being involved in new product and new process development in Kenya's health innovation sector. The results suggest that employment has a positive and significant, albeit small, effect both on new product and new process development. Firms, hospitals or PRIs with no R&D staff over the period 2001–2005 are less likely to be involved in new product and new process development than those that have some R&D staff in any year of the same period. *Ceteris paribus*, PRIs are more often involved in new product development than firms and hospitals, which confirm the descriptive statistics of Tables 5.4 and 5.5, and firms are more often involved in new process development than pharmaceutical hospitals and PRIs. Finally, the share of staff with a PhD, MSc or BSc degree has a negative and significant effect on both new product and new process development. This is a surprising result, as one would expect the opposite, but it explains the mismatch of skills and activities in the local system. This points to the nature of innovation itself – the firms which have staff mainly engaged in quality assurance and quality control for manufacturing various pharmaceutical products are more engaged in innovation, whereas the R&D going on at the PRIs and the universities (which have the more qualified staff on an average) does not result in much local innovation. The two equations of new product and new process development are shown to be positively and significantly related with a cross equation correlation of 0.879. This confirms the descriptive result of Table 5.6.



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

Table 5.6 Bivariate probit ML estimation results: main actors of innovation

<i>Variable</i>	<i>Coefficient</i>	<i>(Std error)</i>
<i>New product development</i>		
Mean employment (FTEs) 2001–2005	0.150 [†]	(0.089)
Staff having university degree	−0.987*	(0.411)
Non-R&D performers 2001–2005	−1.417**	(0.513)
Pharmaceutical PRIs	1.017*	(0.433)
Pharmaceutical firms	0.561	(0.455)
Intercept	0.063	(0.623)
<i>New process development</i>		
Mean employment (FTEs) 2001–2005	0.167 [†]	(0.095)
Staff having university degree	−0.993*	(0.437)
Non-R&D performers 2001–2005	−1.760**	(0.614)
Pharmaceutical PRIs	0.343	(0.497)
Pharmaceutical firms	1.226*	(0.490)
Intercept	0.143	(0.720)
<i>Extra parameter</i>		
ρ	0.879**	(0.089)
No. of observations	98	
Log-likelihood	−68.469	

Source: author's survey, 2007.

Notes

Significance levels: [†] 10%; * 5%; ** 1%.

5.4.5 Policy vision and guiding framework

The legal framework for scientific and technological research and development are guided by the Science and Technology Act (cap 250) laws of Kenya. The Act establishes the machinery through which the government can avail advice on all matters relating to science and technology activities in the country in order to be able to coordinate research and experimental development. The Act also creates a National Council for Science and Technology under the present Ministry of Science and Technology, comprising all the Permanent Secretaries of the relevant ministries and 12 other members representing eminent scientists derived from various disciplines. The National Council on Science and Technology in Kenya has a broad mandate focusing on innovation and new technologies of importance to the country, such as biotechnology and ICTs. The Kenyan government has also been discussing a new Science, Technology and Innovation bill since 2006. Although the



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

bill has not yet been decided upon, in 2008, a Science, Technology and Innovation Strategy was approved. However, Kenya presently spends less than 0.5 per cent of its GDP on science and technology and several sectors of the economy that have demonstrated significant capacity, such as ready-made garments and the cut-flower sector, are all being held back from progressing to more knowledge intensive domains due to the country's weak knowledge capacity.

Since the Science and Technology Act has no specific provisions on biotechnology, a specific framework to address issues of biotechnology development has been formulated. A National Biotechnology Policy was enacted in 2006 that sets out the safety procedures for biotechnology in the context of research and development, technology transfer and commercialization of products that could result from research undertaken in Kenya. The policy also identifies infrastructure development and the allocation of financial and human resources as key priorities for the growth of local health and agricultural biotechnology industry and research. As part of these developments, agricultural biotechnology and Kenya's biosafety framework have received much more attention than health biotechnology per se.

The Kenyan Health Policy Framework of 1994 sets out the health targets to be achieved by the country by the year 2010, which includes adopting explicit strategies to reduce the burden of disease within the country. Over the past few years, the government has also sought to encourage entrepreneurship in the field of drug development using natural resources, further support research through national institutions like KEMRI and also forge alliances between biomedical researchers and traditional healers. The Pharmacy and Poisons Board and the Kenya Bureau of Standards, discussed earlier in Section 5.2, are expected to play an important role in achieving this. The government has also sought to encourage entrepreneurship in the field of drug development using natural resources in recent years. This is being done by supporting more research in the public sector institutions like KEMRI and by enabling local alliances between biomedical researchers and traditional healers.

However, on a sectoral scale, the policy framework on health biotechnology is fragmented and there is no strategic policy vision in place to promote health biotechnology-led development, especially one that takes into account the technological needs (in terms of both physical and knowledge infrastructure) of the process. While Kenya has a national biotechnology policy, intellectual property



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

law and a framework for establishing knowledge infrastructure of relevance to biotechnology research, there is no broader vision that links these to science, technology and innovation policy for the sector (or national science, technology and innovation policy for that matter), local health needs and competitiveness. The amount of financial resources that are allocated to public sector organizations for research is negligible. The framework is reminiscent of a rather familiar state of innovation in latecomer countries (see Oyeyinka and Gehl Sampath, 2009b), wherein the emphasis has been mainly on replicating a set of agencies rather than to enable the knowledge basis required for the sector to flourish. Even within the agencies that are in place, organizational competence is lacking due to the absence of sustained funding and skilled manpower to steer the organizations in productive directions.

The need for agencies that perform umbrella functions of coordinating innovation policy for the country as well as for the sector cannot be stressed enough. Even if one were to disregard this at the present, the lack of organizational competence as a result of sustained funding and skilled manpower to steer the existing organizations in productive directions is a troublesome reality. A striking difference when compared to the Indian and Bangladesh cases is the lack of a sector specific strategic vision (even partial) to promote local production of medicines. The fact that there is a pharmaceutical sector in the country within which several local firms like Cosmos Pharmaceuticals have excellent API manufacturing skills is a clear success story. This issue of resilience, at least in pockets, within local innovation systems that leads to sporadic success stories is a very interesting phenomenon, which is followed up in the Nigerian example (Chapter 6) and Part III of this book. On the part of the government, there even seem to have been several incentives, which for whatever reasons, have run against the development of pharmaceutical capacity in the country. For example, the survey found that local manufacturers who chose to import ingredients and manufacture drugs locally paid taxes that were much higher than those who chose to simply import the finished forms for distribution and retail. Several smaller firms (and even medium and large firms that were small at their inception) complained that these rules on taxes and tax rebates on importing finished forms were a contributing factor to their firm level strategies to branch out into distributing and retailing and not manufacture and product development. Such rules, over time, tend to get institutionalized in pervasive ways and help explain sectoral growth as we observe it in Kenya today.



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

5.4.6 Local health system specifics and impetus to innovation

As mentioned earlier in this chapter, Kenya has one of the lowest scores on the human development index and life expectancy has remained the same over the

Box 5.2 Cosmos Pharmaceuticals

Cosmos Pharmaceuticals manufactures health products in various dosage forms like tablets, capsules, liquid orals, externals and powders both for human and veterinary use. Having been in existence for 32 years now, its main focal therapeutic categories are anti-malarial, anti-retroviral, anti-tuberculosis, anti-diabetic, cardiovascular and gastrointestinal. Cosmos's manufacturing facilities are approved by the Pharmacy and Poisons Board, the National Drug Authority of Kenya and are subjected to regular quality audits by the National Regulatory Agency, Ministry of Health Uganda, Tanzania, Yemen and NGOs like MSF. The drug formulations are manufactured in conformity with the current GMP as laid down by the WHO as well as the drug manufacturing and quality regulations of the Ministry of Health, Republic of Kenya.

Cosmos is the recipient from GlaxoSmithKline and Boeringer Ingelheim for the manufacture of several ARVs and also is amongst the top ten suppliers of Artemesinin based therapy for the cure of Malaria to the Global Fund. For the ACT, it imports the Artemesinin from Tanzanian suppliers and makes the APIs in-house. At the time of the survey, Cosmos had just begun to operate a new facility that conforms to higher drug manufacturing standards of the FDA and some other regulatory agencies.

Source: author's survey, 2007

past 15 years. General performance of the local health sector can be broken up into two main phases. A first phase ending in the early 1990s has achieved some positive trends in health indicators; for example, life expectancy improved from 44 years in 1962–1960 to 60 years in 1993, infant mortality rate reduced from 120 per 1000 in 1963 to 62.5 per 1000 in 1993. However, a second phase from 1993 up until now has a different story to tell. Life expectancy decreased from 60 years in 1993 to 47 years (largely due to HIV/AIDS), while infant mortality has risen from 64 per 1000 in 1993 to 72 per 1000.

There have been several important policy initiatives in the health sector such as the launching of a Kenyan Health Policy Framework in 1994, which articulates the role of the government in improving the health of the population. This was followed by a first National Health Strategic Plan, which was implemented



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

between 1999 and 2004 and aimed to streamline the provision of health services including redistribution of services to the rural areas, with more emphasis on preventive health services and health awareness.

In the present situation, the Ministry of Health provides public health care services and controls 52 per cent of all medical facilities available in the country (Export Processing Zones Authority, 2005). However, there is immense disparity in the provision of health care services between the urban and rural areas. There is an acute shortage as well as unfavourable distribution of medical personnel, i.e. 1:33,000 in rural and 1:1700 in urban in total as of 2005. The survey also shows that the percentage of hospital staff with Masters or PhD degrees is no more than 2 per cent, which is a really worrying result and calls for health systems strengthening in the most fundamental way (see Table 5.6). A health manpower policy was formulated in 2004 with the aim of addressing the development and retention of human resources in the health sector.

Private health spending is close to 78 per cent (WHO, 2007) and health expenditure is 8 per cent of the country's GDP. Modern medicine still caters to a small percentage of the population, when compared to traditional medicine, which is not well integrated into the modern health care system. To increase access to health care, a national health insurance policy has been proposed but consensus is yet to be reached. Some other policy initiatives seem to be working quite well, which have been initiated locally in collaboration with international funders and donors (see Box 5.3). Although the Ministry of Health is one of the largest purchasers of drugs sold locally,³ the survey did not find much evidence of a direct linkage between local innovation capacity and better access to drugs and services in the country both in the public and private sectors. The rural areas, especially, are highly neglected and under-privileged in this regard and private sector based distribution systems still have a long way to go in this regard. The Pharmaceutical Society of Kenya is the agency in-charge of issuing licences to pharmacists and ensures that drug dispensaries and stores perform according to pharmacy practitioner standards. Additionally, it is also responsible for ensuring the distribution of pharmaceutical and non-pharmaceutical products amongst local dispensaries and drug stores, monitoring and advising members on new disease control programmes and promoting high quality training of pharmacists. However, the survey found low evidence of well-qualified personnel in pharmacies and drug stores across the country.



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

5.4.7 The failure to establish local collaborations and linkages

Table 5.7 shows the distribution of collaboration intensity between firms, hospitals and public research institutes in the sector as a whole. As in all the other surveys, all respondents were asked to rate their collaboration with other actors on a scale of 1 to 5, where 1 denoted very weak and 5 denoted very strong.

Box 5.3 Health voucher scheme for rural and urban health in Kenya

In July 2006, the Kenyan government launched "Vouchers for Health" in a three-year pilot targeting poor women in three rural districts and two urban slums. The programme provides access to safe deliveries, family planning and gender violence recovery services. The initial budget was about €6.5 million.

The project is based on the concept of reimbursing providers for services rendered rather than paying for service inputs beforehand. In this project, providers are accredited and contracted to deliver high-impact services to voucher-bearing patients. Vouchers are sold at highly subsidized prices to poor women, allowing them to choose where to access safe deliveries, family planning services and gender violence recovery services. The subsidized vouchers broaden access for the poor to public, private, NGO and faith based hospitals. Detailed planning was done before the onset of the project through participant observation, community interviews and provider assessments. It provided data for the initial design and benchmarks against which results are measured.

An impact evaluation tool has been developed with the main focus to establish how well the poor have been targeted and to measure impact on maternal and infant mortality and morbidity.

Only 18 months into the project 97,500 vouchers (72,800 for safe motherhood and 24,700 for family planning) were sold to poor women for equivalent of US\$2 and 70,600 vouchers were reimbursed. Reimbursed delivery services are worth US\$60–200. Due to the project's initial success, the extension to new districts and additional services is planned with additional €10 million. As a national project, it has the potential to inform the creation of a national social health insurance programme. One can view the voucher as an insurance card (with clearly outlined benefits) giving access to services built similarly to insurance. Common elements include accreditation, quality assurance, reimbursement system, claim processing, costing and pricing, integrating private sector, client choice, provider competition, access and equity. The project "Vouchers for Health" and the poverty-grading tool can serve as an international model.

Source: author



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

Table 5.7 Intensity of collaboration in Kenya's health innovation system

<i>Collaboration intensities</i>	<i>Firms</i>	<i>Hospitals</i>	<i>PRIs</i>
With PRIs	1.43	1.5	3.24
With industrial associations	1.87	1.43	–
With universities	1.41	1.68	3.08
With private laboratories	1.60	2.25	2.64
With hospitals	2.19	2.62	2.60
With traditional medical practitioners	1.41	1.40	1.52
With firms	1.56	–	1.45
With governmental agencies	1.75	–	2.80
With international organizations	–	–	3.32
Total number of observations	41	32	25

Source: author's survey, 2007.

More specifically, Table 5.7 reveals several important collaboration trends in the country's health innovation system that shed more light on product and process innovation patterns. First, the patterns of collaboration intensity of firms are similar to those of hospitals while both these patterns are different from those of PRIs. More specifically, collaboration intensity is on average rated weak by firms and hospitals (less than 2), but fairly strong, strong and very strong by PRIs (around or above 2.5). Second, universities tend to collaborate and have much less interaction with other actors in the innovation system than PRIs, which once again helps to highlight the funding and research difficulties peculiar to university centres of excellence within the Kenyan public sector. Third, firms collaborate mostly with hospitals for the sale of their products. All three key actors – firms, hospitals and PRIs – demonstrate very weak collaboration with traditional medicinal practitioners who treat a large majority of the local population. This is a very important result that shows that traditional medicine, despite its importance in the local health system, is not well integrated into the ongoing innovation activities in any significant way. It also helps to point to the lack of trust between traditional medicinal practitioners and other actors in the local health innovation system. Firms collaborate equally weakly both with universities (1.41) and PRIs (1.43), which points to the lack of sufficient linkages between the public sector and private enterprise for health innovation; usually assumed to be the biggest precondition for successful health innovation. Firms also have extremely low collaboration with other firms (1.56) and no collaboration with international organizations (both in the public and private sectors), which also helps to underscore the low level of product development activities happening within the



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

country. This is a very different result from what was observed, for example, in the collaboration trends in the Indian sector (Chapter 3). Finally, a very important result is that only PRIs have collaboration with international organizations. This issue is discussed further in the next section.

More rigorous econometric analysis to explain new product and new process development by collaboration intensity variables (a bivariate probit model not presented here) confirms these results further. The analysis reveals that collaboration intensity with PRIs and private laboratories affects positively and significantly the probability of being involved in new product development. At the same time, the probability of being involved in new process development is positively and significantly influenced by the intensity of collaboration with hospitals. Finally, *ceteris paribus*, pharmaceutical firms are more often involved in new process development, which once again shows the kind of innovation (when present) in the local pharmaceutical sector.

5.5 Global pull factors and their impact on innovation patterns

Two global pull factors are significant in the context of Kenya's health innovation, one negative and the other positive. Kenya has been considerably successful in attracting several international partners for collaborative research, especially in the area of biotechnology (although agricultural biotechnology is a greater focus of such collaborations than health, see Oyeyinka and Gehl Sampath, 2009b). At the same time, Kenya's intellectual property regime, which is TRIPS compliant, is not necessarily a positive impetus to local innovation given its present status.

5.5.1 International collaborations

Kenya's experience with international collaborations is interesting and calls for a re-look at several levels. At a first glance, it is striking to note, as Table 5.7 shows, that the only set of actors that rate international collaborations as an important factor for new product and process innovations are PRIs (a rating of 3.32). However, there is limited mobility of labour between the public and private sector enterprises (or even within different public sector enterprises) and a large amount of tacit knowledge acquired in the PRIs remains within the organizations. Predominantly, the survey finds that this is because of the lack of linkages between the various actors on the one hand within the local innovation system, and second because of the low product development emphasis on the other.



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

Kenya has been a partner in several very acclaimed international collaborations in health, especially the Kenyan AIDS Vaccine Initiative. IAVI focused on localizing its R&D efforts in various latecomer countries in order to promote ownership of the initiative (amongst the southern partners) and communication (Chataway *et al.*, 2007). However, most of these projects have not contributed to building local research capacity in ways that spill over substantially to the other actors in the innovation system because of the observed tendency of international private sector companies to bring into the country finished (already modified) products for trials and limited production, thereby circumscribing the active participation of local public sector institutions and their researchers in the research and product development process. This points attention to a criticism that Kenyan international collaborations have met previously (especially in biotechnology), deeming them to be exogenous, driven largely by international private sector interests and supported by the donor community or international private foundations.

The lack of absorption of capacity through international collaborations into the local innovation system can be traced back to the lack of knowledge infrastructure in addition to collaboration incentives, as the survey data reveal. The survey finds a strong relationship between national strategy for biotechnology development, availability of relevant human skills in the local research institutes and international collaborative efforts. In other words, if there were more relevant human skills that could be deployed and if national and organizational strategies for biotechnology were more clearly set out and implemented to make capacity building a priority in international research collaborations (both of which rely on policy capacity, see next section), strategic involvement of local researchers in international collaborations could have been effected, which is now not the case. The survey also found that most researchers who take part in such capacity building and training are disgruntled by the low state of innovation capacity in the local system which constantly acts as a hindrance to applying their acquired skills to research and innovation activities to the local context.

5.5.2 IPRs

The country's intellectual property regime is compliant with standards set out by the Agreement on Trade Related Aspects of Intellectual Property Rights of the WTO. The IPRs in Kenya are covered under four Acts of Parliament, namely: the Intellectual Property Act (cap 509), the Trademarks Act (Act 506), the Seeds and Plant Varieties Act (cap 326) and the Copyrights Act (cap 150). Kenya has a very low



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

score of local patent applications when compared to foreign ones, which once again acts as a measure of the low level of local entrepreneurial activity. Patent registrations amounted to 61 in the year 2001, with two of these being registered by residents and the rest by non-residents. The number of international patents operational in Kenya as of 2002 amounted to a total of 89,180 according to 2002 data available with the WIPO (WIPO Patent Information 2007). The extremely stringent intellectual property regime certainly poses limitations on reverse engineering activities in health innovation and Kenya is yet to enact several flexibilities that are allowed and important under the TRIPS Agreement,⁴ such as a research exemption to universities and public research institutions. There is also very little awareness of patenting possibilities in the public sector. The survey also reveals that all actors advocating and negotiating for royalty-free access to biotechnologies in Kenya (such as the International Service for the Acquisition of Agri-Biotech Applications and the African Agricultural Technology Foundation) are non-governmental in nature, underscoring the lack of awareness amongst policy makers and KIPi on the impact of the present intellectual property regime on local entrepreneurial ventures.

Although Cosmos Pharmaceuticals has acquired voluntary licences for the production of some ARVs, the Kenyan government has not compulsorily licensed the production of any drugs in the interest of public health.

5.6 Conclusion

The key actors and their constraints identified in this chapter point towards a very low capacity to carry out health biotechnology innovation in Kenya. The policy framework is highly fragmented and unable to coordinate the technological requirements of health biotechnology with the local health needs of the population. The universities, local research institutes and governmental agencies suffer similarly from lack of funding, organizational competence and knowledge infrastructure of the kind required to embark on health biotechnology. The private sector remains stunted due to difficulties of accessing relevant technologies, venture capital and risk-sharing mechanisms. These limitations in the innovation environment also limit knowledge spillovers from international collaborations into Kenya's health biotechnology innovation system. In particular, the following challenges will need to be overcome in order to promote the growth of innovation capacity in the sector.



KENYA'S HEALTH INNOVATION CAPACITY

Excerpted from *Reconfiguring Global Health Innovation*

5.6.1 Challenge 1: promoting a vibrant private sector for innovation

Kenya's local private sector enterprise in health and biotechnology is a critical asset in promoting innovation capacity. It is presently highly challenged by foreign competition and lack of technological support from the public sector and the government. Despite the intent of the National Biotechnology Policy of 2006, little has been done to promote the emergence of a vibrant private sector for health biotechnology. Emergence of a vibrant private sector is a reflex to the presence of risk attenuating mechanisms that offer business support and promote enterprise. Innovation experiences in health biotechnology both in the frontier countries and relatively successful latecomers find evidence of several innovative financing activities for both technological and health care services. There is little governmental infrastructure for technological incubation, acquisition of technology required to pursue innovation and funding, especially small-firm risk sharing and financing initiatives of the kind a country like Kenya would need.

5.6.2 Challenge 2: a highly fragmented policy framework for health and biotechnology

The policy framework is highly fragmented and unable to coordinate the technological requirements of health biotechnology with the local health needs of the population. The universities, local research institutes and governmental agencies suffer similarly from lack of funding, organizational competence and knowledge infrastructure of the kind required to embark on health biotechnology. The private sector remains stunted due to difficulties of accessing relevant technologies, venture capital and risk-sharing mechanisms. These limitations in the innovation environment also limit knowledge spillovers from international collaborations into Kenya's health biotechnology innovation system.

5.6.3 Challenge 3: lack of skilled human capacity

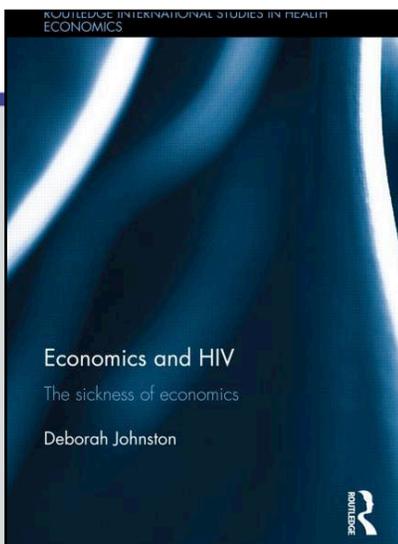
A review of previous studies all agree on this point: Kenya still has a shortfall of adequate laboratory capacity and facilities needed to effectively exploit biotechnology (Wafula and Falconi, 1998, and Odame *et al.*, 2002 to mention a few). Human skill, the other component of knowledge infrastructure, is highly limited, a point that has been discussed at length in the previous section.



CHAPTER

6

A POLITICAL ECONOMY PERSPECTIVE ON THE ECONOMIC IMPACT OF AIDS



This chapter is excerpted from
Economics and HIV: The sickness of economics
by Deborah Johnston

© 2013 Taylor & Francis Group. All rights reserved.



[Learn more](#)



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

If mainstream approaches to the macroeconomic or sectoral costs of AIDS do not do a good job of illuminating the impacts of AIDS, what is the alternative? This chapter sets out a political economy approach that helps us analyse the effect of AIDS on the economies of the most highly affected countries. Drawing on a structural approach to growth supplemented by both political economy and feminist economics, this chapter will suggest the key issues for an understanding of AIDS impact. These issues could provide important information for both national policy-makers and international organizations.

To understand how AIDS might impact on overall economic growth and development, we first need to think about economic growth. We have already seen in Chapter 6 that the single-sector approach of the neoclassical growth model has been widely criticised. This is because it fails to explain empirical patterns (Kenny and Williams 2001) and because of its questionable theoretical assumptions (Fine 2000b). These latter concerns about the unrealistic theoretical assumptions are many and varied. At their heart is a concern that standard theory uses a single-sector model and is inherently unconcerned with a central feature of growth and development – rapid economic and social structural change. In the following section, we will look at an alternative approach to economic growth that is fundamentally concerned with changing economic structure. Once this is set out, it is the work of a later section to rethink the relationship between AIDS, growth and economic development.

A structural model of growth

Alternatives to mainstream growth theory have been widely suggested. As Fine (2000b) has shown though, the main contender, endogenous growth theory, shares many of the same underlying assumptions as standard growth theory. A more promising alternative draws on the work of both Keynes and Kaldor, and is often labelled in current writing as a structuralist approach. One of the best examples of this approach is the work by Ocampo *et al.* (2009), which sets out several key characteristics of the approach, starting out by recognizing that convergence in income levels among countries is the exception rather than the rule. There has been a clustering in time of successes and failures, and it is argued that this is because international factors play a crucial role in the overall growth dynamics of the developing world. Here is a key contrast to the mainstream emphasis on domestic policies and institutions. These international factors play out through



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

terms of trade and interest rate shocks, debt relief and aid, as well as trade flows.

A second contrast with the mainstream is, as was suggested above, a focus on economic structure, specifically with high productivity, high value-added manufacturing, suggesting that 'intelligent sector-level policies can facilitate the development process' (ibid. 144). As Tregenna (2008) argues, in its simplistic form, the intersectoral reallocation of labour from low- to high-productivity activities is central to increases in overall productivity. However, this reallocation of labour happens through changing patterns of demand, rather than through supply-side changes. Growth is seen as being demand-led, and this focus on demand again contrasts with the mainstream approach, which is supply-focused. Instead Ocampo *et al.* (2010) say productivity growth is as much a result as a cause of economic growth, as increasing domestic or international demand for high-value added activities allows the fuller utilization of resources and also promotes dynamic productivity increases.¹ Here the standard supply-side stories about productivity growth arising from improvements in the factors of production are thought of as, at best, weak. Thus, nutrition, education and health are a necessary 'framework condition' for growth, but not sufficient to ensure that it occurs (Ocampo *et al.* 2009: 146). Economic success in this approach is seen as being as much about States as about markets, as government is seen as being responsible for a favourable structural transformation of the economy and for managing external shocks.

Thus, the manufacturing sector is accorded a special place in the growth process, but the rate of growth is also affected both by international factors and demand trends. This approach has been increasingly concerned, then, about the evidence of deindustrialization in many African economies following the adoption of neo-liberal policies (Noorbakhsh and Paloni, 2000; Rodrik, 2006; Shafaeddin, 2005; Tregenna, 2008). The liberalization policies referred to here were the macroeconomic stabilization and structural adjustment policies that were the conditionality for receiving IMF and World Bank funding, in the wake of oil price crises and global recession of late 1970s. These were intended to improve static and dynamic efficiency. While there are methodological problems in unpicking their effects (see Noorbakhsh and Paloni 2000), most writers agree that there has been a clear lag in sub-Saharan African performance in terms of growth in manufacturing value added. Shafaeddin (2005) argues that the impact of trade liberalization was negative for African countries, because the competition from imports and removal of state support led to deindustrialization and increased



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

reliance on imports in manufacturing. Structural adjustment and stabilization programmes failed to encourage private investment, particularly in the manufacturing sector. Structural adjustment policies failed to take the sub-Saharan Africa specificity into account (Noorbakhsh and Paloni, 2000; Bigsten *et al.* 1998; Rodrik 2006).

Sub-Saharan African countries are now seen as being severely constrained in terms of manufacturing development, due to increasing competition from other developing countries, such as textile competition from Vietnam and China (Kaplinsky, 2006). Tregenna (2008, 2012) has argued that this pattern of deindustrialization has characterized South Africa, and that the economy is predominantly dependent on domestic demand. This dependence on domestic demand expansion as a source of growth has emerged since 2000, especially for manufacturing, which has generally performed relatively poorly. This kind of decomposition is useful in illuminating the key issues for policy but has rarely been carried out for African countries, not least because of the unreliability of African growth data (Jerven 2010).

While the structuralist approach is invaluable in highlighting the importance of structural, international and demand-side factors, it does have some lacunae. One is the failure to fully consider where domestic demand comes from and here the work of political economy is useful in setting out how political and social factors affect the pattern of incomes. Bujra (2004: 632) uses a Marxist focus on the reproduction of capital to remind us that this reproduction depends on both the economic sphere (in which capital exploits labour) and the social sphere (in which labour is reproduced). Marxists have always been concerned about the tension in these processes, as it is in capital's interest to exploit labour power for profit, but it needs labour to reproduce itself and to consume its products. Importantly, Bujra (2004: 632) argues that capitalist production in Africa remains dominated by both 'the looting of resources and the exploitation of labour still anchored in rural semi-subsistence economies'. This would paint a picture of a capitalist development barely changed from the colonial past. However, Marais (2005) reminds us that the present character of capitalism in Africa is fundamentally affected by the character of neo-liberal policies that swept away the old state interventions we discussed in the previous few chapters and instead installed a new set of interventions designed to increase profits. What both authors would agree on is that, although inherently diverse, the system of production in sub-Saharan African countries has been based on casual labour, low wages and



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

limited benefits. Where growth has occurred, as a result of metals and oil booms, it is not clear that this has led to widespread employment creation.

The discussion in this section has presented an alternative view of the growth process, very different from that discussed in Chapter 6. Rather than the ahistorical, one-sector supply-side model of neoclassical economies, we have painted a picture of the need for demand-generated structural change. Indeed, we have explicitly generated a picture in which the health of labour does not itself lead the growth process. Structuralist approaches see it as a facilitating condition, while political economists emphasize the way that capitalist development has been underpinned by low wage, casual labour. This paints a very different picture to the Sachs' approach, also described in Chapter 6, where the supply-side impacts of good (or bad) health generate growth through increases (or decreases) in labour productivity. How can the impact of AIDS be conceptualized? To understand this, we might first want to look again at the relationship between AIDS and social reproduction, where the gendered characteristics of the epidemics begin to take on new meaning.

AIDS, social reproduction and gender

In Chapters 1 and 3, we saw that in Africa there was a predominance of women living with HIV. This appears to reflect a combination of biological, economic and social vulnerability has led to a greater susceptibility of women to HIV/AIDS (UNAIDS/UNFPA/UNIFEM 2004; Urdang 2006). This has a clear effect on the modelled demographic impact. For example, in Namibia, South Africa and Zimbabwe women will have lost at least six more years from their life expectancies than their male counterparts by 2025 (United Nations 2003).

This differential impact of HIV/AIDS on women makes it especially important for macroeconomic studies to consider the effect of higher female illness and death. Macroeconomic approaches would have to take into account the impact of HIV/AIDS on women's activities and the implications of this impact for the economy as a whole. However, the vast majority of studies in Table 6.2 (in Chapter 6) do not separately model female and male prevalence rates or mortality, though there are two exceptions.

The Bell *et al.* (2003) model of the relationship between education and the macroeconomy predicts higher mortality rates for mothers within households



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

resulting from higher female HIV/AIDS prevalence. However, they assume that mothers and fathers have identical preferences and activities related to child rearing, so these differential prevalence rates have no influence on their results. In contrast, Young's (2005) study separately models the impact of HIV/AIDS on men and women, with the difference being that women bear the costs, in terms of time, of child rearing; however, Young assumes equal HIV/AIDS prevalence rates for men and women. HIV/AIDS reduces fertility rates because women have a greater disinclination for unprotected sex and an increase in the market value of a woman's time reduces her willingness to spend time raising children. Overall, Young finds that the impact of HIV/AIDS is to increase income per capita due to a rise in the capital-labour ratio and an increase in the education level of children (with lower fertility leading to better-educated children).

Young's assumptions seems to bear little resemblance to the stories told in micro-level or qualitative work, about declining education preferences of households and about the limitations to women's decision-making (see Chapters 3 and 8). Moreover, both studies contain a very limited treatment of differences in the activities of men and women. Drawing from a review of available time-use and mortality studies regarding women's reproductive activities, I will show that the micro evidence from sub-Saharan Africa suggests that the burden of reproductive activities is not reduced by HIV/AIDS in the way Young proposes and that the higher mortality of mothers may be more noteworthy than Bell *et al.* assume.

Reproductive activities are not included in this measure of economic activity.² These activities are excluded because the unpaid provision of domestic and personal services – such as doing the laundry, preparing meals and caring for adults and children – falls within the United Nations System of National Accounts (SNA) general production boundary, but is outside the SNA's economic production boundary. Thus, it is termed non-SNA production (Young 2000). Data on non-SNA production activities are not collected by the statistical instruments used to collect information on GDP. However, some countries have collected this data separately and have used it to construct satellite accounts, which suggest that the value of non-SNA production is quite large.³ Consequently, the macro models reviewed in Chapter 6 focused on SNA economic production (with the exception of the paper by Young 2005).

However, Chapter 8 showed that AIDS has an important impact on reproductive activities, with the demand for reproductive activities rising particularly sharply in



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

the later stages of AIDS-related illness. Although men and children are also engaged in the reproductive economy, it is clear that women are the primary providers of such activities. Akintola (2004), United Nations Development Fund for Women (UNIFEM 2005), and Urdang (2006) all discuss the fact that caring for those with AIDS remains predominantly a female task in Africa. Similarly, Steinberg *et al.*'s (2002) study of South African households affected by AIDS found that in two-thirds of households surveyed, the primary caregiver was female.

More generally, household time-use studies from sub-Saharan African countries provide general evidence of the burden of reproductive activities borne by women, with available studies suggesting that female respondents spent between 2.5 and 3 times as many hours on reproductive activities as male respondents (Budlender and Brathaug 2002; Latigo and Mohammed 2005). However, although these time-use studies provide useful general information, the results give only average indications of the burden of reproductive activities, despite evidence that women's reproductive work is not distributed evenly across class and social groups. Fontana's (2002) gendered CGE model of Zambia attempts to disaggregate women's burden of reproductive activity by both location and educational status. Some studies give limited glimpses of the use of hired caregivers by some households, which illustrates the importance of knowing how different households will meet HIV/AIDS-related care needs.⁴ For example, Steinberg *et al.* (2002), in their South African study, provide the qualitative example of a South African woman who hired a carer for her ill adult son. However, no micro studies offer overall data on the percentage or characteristics of households who engage hired carers.⁵ Thus, in considering the increasing demand for reproductive activities arising from HIV/AIDS, it should be remembered that some better-off households may be able to meet these demands by hiring paid carers.

For households that are unable to hire paid carers, Chapter 8 showed the increased demand for reproductive activities will lead to a change in the pattern of all other reproductive and productive activities in the household. The findings from micro studies suggest that HIV/AIDS will increase the burden of women's reproductive activities and that this will often affect the productive activities. How should we think of this affecting economic growth?

One way is to incorporate the effects on the supply side, through changes in the labour supply. However, the model by Young was the only model reviewed in Chapter 6 that explicitly includes reproductive activities, and even this model



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

assumes that women's reproductive activities would fall due to higher market wages and lower fertility (2005). While other models focus on the productive economy, few make any explicit allowance for changes in the burden of reproductive activities.⁶ Only three could be found in this review: In their models using evidence from Tanzania and Malawi, Cuddington (1993b) and Cuddington and Hancock (1994) note that some individuals may withdraw from the labour force in order to care for the sick. Both models make a range of assumptions, the extreme one being Cuddington's assumption that for every individual who withdraws from the labour force due to illness in the last stages of AIDS, another individual withdraws to provide care. In a study of South Africa, Quatteck revises Cuddington's number to suggest that a lower value of 1.33 should be adopted for South Africa, as the country's medical care system is relatively sophisticated, assuming: '[F]or every person with full-blown AIDS, four months of person-year equivalent labour supply will be lost' due to increased burden of reproductive activities (1998: 37). This, however, looks overly conservative given the, albeit limited, results of Booyesen and Bachmann's (2002) study.

While these three models do at least consider the impact of increasing reproductive activities on the labour market, none of the models for the productive economy look explicitly at other impacts of increased care, such as the withdrawal of children from school. This point becomes clearer if the consequences of illness or mortality of adult female caregivers are considered. The evidence in Chapter 8 suggests that the death of a mother often has differential impacts to the death of a father. However, this picture has not been captured even by those macro-models that do include assumptions about reproductive activities. For example, Bell *et al.* (2003) assume that there would be no differential impact on education arising from a mother's as opposed to a father's death. Cuddington (1993b), Cuddington and Hancock (1994), and Quatteck (1998) focus only on the impact of higher demands for personal care during the final stages of AIDS illness and do not look at the impacts that occur after mortality. In order for these models to accurately assess the impact of AIDS on a larger scale, it must explore the ways in which the epidemic has ushered in changes to time devoted to reproductive activities and the subsequent impact on household members.

This section has suggested that AIDS has a particular impact on household reproductive activities, which will tend to draw women from the labour market and children from school. These effects will be strongest for households that cannot afford hired carers and do not have unemployed members who can undertake such



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

activity. These effects were not fully incorporated in the macroeconomic models reviewed in Chapter 6. The implication is that the current size and nature of the labour force will be affected, as will future skill levels. These effects occur in addition to those of mortality itself. We saw in Chapter 8 that, under certain conditions (e.g. especially with maternal mortality), there are long-run impacts on educational access and nutritional status for affected children. We saw in Chapter 8 that poor households struggle to ring-fence food expenditure and are sometimes unable to maintain adequate food consumption.

Overall then the disruption to reproductive activities caused by AIDS is likely to have both short- and long-term impacts on the quantity and quality of labour supply. We will discuss this further in the next section. Before moving on, it is important to consider the implications for social reproduction from this discussion. As Bujra (2004: 633) concludes, AIDS exposes the way that African economies are gendered. As we saw in Chapter 8, it also exposes the class inequalities between households. The shifting of the burden of social reproduction from the State and from large companies onto households, as employment benefits and state intervention to ensure welfare have been cut back, has particularly hit poor women. AIDS is not a leveller but an amplifier of inequality.

How sustainable are these pressures on social reproduction? Poor households are initially hard hit, and may have some kind of rebound in consumption to levels immediately prior to AIDS mortality, but this is hardly a laudable outcome. The period during which AIDS leads to appreciable illness has already had significant impacts on an individual's ability to work. Illness reduces household income and eventually raises demand on the other (female) household members. This shift in the level and demands of household reproductive activities has important social and economic implications, especially for women in poor households.⁷ The startling impacts of ARVs on an individual's productivity and employment – reported in Chapters 7 and 8 – have significant implications for policy-makers. The provision of ARVs not only fulfils people's right to health, but also has clear impacts on household income and clear impacts on the gender burden of household reproductive activities.

Labour, demand and long-run growth

There are several interrelated issues that we will discuss in this section: that



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

labour markets exhibit complex segmentation; that issues of skill and productivity are poorly understood; and responses to AIDS impacts by firms will be crucial for the long-run welfare of households. However, at the end of this section, we will turn our attention to issues of aggregate demand, and the affect of AIDS.

The structuralist approach to growth would ask us to consider if sectors with the highest value added, especially manufacturing, have high prevalence rates. However, when thinking about the labour market effects of AIDS, in terms of worker morbidity and mortality, it is difficult to make generalizations. Chapter 4 suggested that although limited in the quantity of studies, it is apparent that there are complex and distinct sectoral patterns of HIV prevalence among African countries. We can see this in a simple way – by looking at the relationship between wealth and HIV prevalence. Patterns in East Africa (with higher rates for higher wealth groups) would imply that it is skilled labour that has the highest prevalence. In Southern Africa, this pattern is quite different and it might be that HIV prevalence is highest for unskilled or semi-skilled labour.

At the same time, it is clear that with higher prevalence rates for women, sectors dependent on female workers might be especially vulnerable to labour force disruption from AIDS. Combined with our conclusions on reproductive activity in the last subsection, we might conclude that female labour is particularly affected both by the direct and indirect impact of AIDS. However, it is even more difficult to unpick the sectors most dependent on female labour. Female labour force data is poor in many countries. Budlender (2004: 19–20) reminds us that while there have been specific attempts to improve the data on women's productive activities in countries such as South Africa and India, in many others, labour force data continues to undercount women's work in the areas of informal wage work, subsistence production and self-employment. Labour force data often poorly represent the employment experiences of the poorest women (see Sender and Johnston 1996 on South Africa).

The garment industry and other light industry often relies on female unskilled or semi-skilled labour. One illustration of the higher rates of HIV prevalence among female workers comes from a joint ILO-UNIFEM report on the garment industry in Lesotho (reported in UNIFEM 2005). In the context of very high prevalence of HIV/AIDS, absenteeism among the mainly female workers had risen dramatically, with one factory manager reporting that each month almost half of the 6,500 workers visit the factory health clinic. This impact is likely to be particularly severe



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

or long-lasting where gender segmentation of the labour force is slow to change.⁸

What will the overall economic costs of these patterns of prevalence be? As we saw in Chapter 7, the direct costs to employers are often small and can be offloaded over time, as firms reduce worker benefits (such as sick leave, health care and life assurance). The greatest supply-side costs are likely to come from the indirect costs of lower productivity, absenteeism and high worker turnover. This is an area in which it is clear that very simplistic assumptions are made in the usual macroeconomic models. The standard neoclassical model has a single sector and was unable to incorporate stories of overall productivity impact within the model—productivity impacts occurred exogenously when some modellers adapted estimates of labour productivity to account for worker absenteeism or illness or when it was assumed that capital was substituted for labour. The dual sector models discussed in Chapter 6 were an advance, as they allowed a more nuanced discussion of the impact of AIDS on labour productivity. For example, Cuddington (1993a) assumes that the informal sector is generally more labour-intensive and its workers have lower labour productivity than the formal sector. He also assumes that workers in the informal sector are paid their average product for social reasons, with enterprises being family run (1993a). Markus Haacker (2002) also assumes that the informal sector is more labour-intensive and that workers have lower labour productivity. Rather than formal and informal, Cornia and Zagonari (2002) model ‘agriculture’ and ‘all other’ sectors. However, they again see ‘agriculture’ as using unskilled labour with low productivity.

This assumption about low-productivity labour drives the results of all the dual-sector models discussed here, which find that the informal or agricultural sector will experience less of the impact of HIV/AIDS than the formal sector. This occurs because all studies assume that the productivity of all workers falls by the same percentage (due to effects like absenteeism and the loss of more experienced workers). However, because these models assume that workers in the informal sector are less productive than other workers, the absolute impact felt by this sector is smaller than for other sectors. Thus in Cuddington’s approach, the impact is greatest for the formal sector because it has higher absolute productivity loss (1993a). In the Cornia and Zagonari model (2002), the non-agricultural sector shows the greatest impact because the model assumes that productivity is highest among (non-agricultural) skilled workers. Haacker (2002) argues that the informal sector is more insulated to the impact of HIV/AIDS because it is less affected by productivity shocks to labour.



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

Clearly, this assumption of a low-productivity, undifferentiated sector is open to two criticisms: first, it is not sensitive to differences within the informal sector, which encompasses a wide range of rural and urban activity. Second, it is not sensitive to differences within the informal sector workforces and underestimates the role of skills and experience in some roles. The UN (2003) says that the impact of HIV/AIDS may be underestimated because knowledge levels may be high within the informal sector, and workers may be difficult to replace.

There is a wider problem here. These approaches say very little about the way that employers will react to these direct and indirect AIDS costs. In the models above, it often assumed that as labour becomes less efficient, there will be a shift towards more capital-intensive technologies. However, as we saw in Chapter 7, this was not the only response that employers can make: they can also choose to offer employees greater benefits (to maintain their health for longer); or they can choose a casualization policy along with pre-employment checks or other measures to reduce their exposure to AIDS-related costs. There was abundant empirical evidence that the vast majority of firms in high-prevalence countries were providing little for workers, with very few providing even basic information or advice. Surveys suggested that it was only some high-profile, international firms who opted to provide a wide range of health benefits. State regulation also has a role in this, and the work of Marais (2005) suggested that AIDS may have intensified a trend to reduced employee benefits and labour casualization that began when African countries adopted neo-liberal policies.

Of course, casualization has its limits. There are recruitment and retraining costs for firms, and its success is dependent on there being a widely available supply of labour at the right prices. For skilled and semi-skilled workers, there is a limit to this assumption. However, the pressures of AIDS are likely to have mixed impacts on unskilled workers. As the poorest households struggle to reproduce themselves in the face of reduced income and higher reproductive activity, there are two countervailing factors. On the one hand, labour is needed to help care for ill individuals, while on the other hand, the desperate need for more income to cover increased spending demands is likely to propel more people into the labour force. It is hard to see what the outcome will be in terms of the overall supply of unskilled workers, although it is clear that the pressure on household incomes means that implications for long-run education and nutrition are poor.

This discussion has so far focused on the supply-side impacts of AIDS. However, it



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

is the demand-side factors that might be most important in terms of inducing longer-run productivity changes. Our structuralist approach would suggest that trends in demand, both domestic and international, for manufacturing production is crucial in explaining if dynamic economies of scale can be secured. The earlier discussion also suggested that the manufacturing sectors of African economies were decimated by liberalization and are now constrained in their ability to compete with exports from Asian countries.

What impact will AIDS have? We have seen in Chapter 7 that household disposable income is likely to be severely squeezed as income falls and health-related expenditure rises. Non-essential expenditure will be reduced, and for the poorest households, even food expenditures may be reduced. Low-income consumer goods, such as cigarettes, personal items and low-end consumer durables will be most affected. For the wealthiest households, employment benefits or medical insurance may cushion some of these changes – and importantly access to ARVs will mean that afflicted individuals can maintain their income-earning activities. The spending patterns of the wealthiest households might see little change as a result. This divergent pattern of impacts is likely to have important implications for the domestic manufacturing sector. Assuming that African manufacturing activities usually concentrate on low-value, unbranded consumer goods, these stories about domestic demand are bad news, and suggest that manufacturers will experience a downturn in demand. This will have long-run impacts on productivity growth, unless enterprises can find new more buoyant markets internationally. To do this, they will have to contend with competition from Asian countries and so it may not be a feasible option.

Conclusions on macro-modeling approaches

Part II of this book began by showing the weaknesses of macroeconomic models of AIDS impact. These weaknesses ranged from general problems with mainstream growth modelling to specific problems with the data on both HIV prevalence and the impact of AIDS. While macro studies cannot reduce the experience of those affected by HIV/AIDS to statements of mathematical formula, to modify Waring's caution regarding macroeconomic enquiry (1989), it is apparent that such models have policy importance, despite their focus on GDP per capita, which this article has identified as a very imperfect measure of welfare. I have argued that structuralist approaches would improve the realism of macroeconomic



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

assessments of AIDS. When combined with a feminist economics approach that allows us to picture the impacts of higher reproductive activities, we can illustrate more clearly the aggregate impact of the HIV epidemics.

This chapter offers an explanation for the small impact that HIV/AIDS has on productive activities found by most macro studies. Mainstream models have only had limited ways to factor in the impact of AIDS and to understand the operation of labour markets. It is clear that mainstream approaches give cursory attention to the nature of social reproduction, include limited changes for productivity impacts and ignore the implications for consumer demand. The combination of these three factors suggest that the long-run impact of AIDS will be greater than mainstream macroeconomic predictions often suggest. Long-run productivity growth is likely to be slower due to both supply- and demand-side factors. The strains of household reproduction for poor households will be one factor in this, as they reduce their expenditure on non-food items, such as consumer durables and even education. The empirical evidence suggests that in some cases, even food consumption cannot be maintained and that AIDS has long-run impacts on child nutrition – another pathway by which long-run productivity will be affected. Without legislation to protect worker rights, most firms (and farms) are likely to deal with the potential supply-side costs of AIDS by intensifying the casual nature of employment, at least for unskilled workers. This is likely to further affect household reproduction, by reducing pay, job security and worker benefits. A vicious cycle could be created, where a low-wage, low-rights model only intensifies the deflationary affects of AIDS.

The widespread provision of ARVs promises to bring a host of benefits – around worker productivity, reduced reproductive burdens and also crucially (and remembering our discussion in Chapter 5) in reduced HIV transmission. Local production of ARVs and condoms would also lead to greater demand-side benefits, as the provision of imported ARVs (and condoms) means that public funding serves to increase the demand for imports. Local production of generic ARVs in South Africa has however been problematic due the privatized model adopted. If local production could be restructured under a more transparent model then it might yet serve to inject demand into moribund manufacturing sectors.

Part II of this book has suggested an extensive research agenda. The limited nature of the empirical evidence has been shocking. We know very little about the impact of AIDS and do not have some of the data building blocks (for example around



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

sectoral prevalence levels or AIDS morbidity) that we would need to tell a more accurate macroeconomic story. There has been huge unevenness by country, with the best data being available for South Africa, Kenya, Uganda and Tanzania. The same is true for sectors, and we know very little about the impacts, for example, on large-scale farming or manufacturing. Furthermore, we need to think carefully about the methodology used to assess impact. A political economy approach would suggest that calculation of average impacts across either a variety of households or sectors tells us little. If we are concerned about the impacts on growth, we need to have a clearer focus on sectors of higher value added. If we are concerned about household reproduction, we should focus on the impact on different classes and genders.

Notes

1. Dynamic economies of scale are thought to come about through Kaldor–Verdoorn effects, whereby production itself leads to productivity improvements. The rationale is that more rapid output growth leads to the introduction of more productive technologies and the realization of economies of scale. These economies of scale are not only statics in nature (i.e. one-off benefits as scale is increased) but also dynamic (i.e. there are continuing increases in productivity over time) (Ocampo *et al.* 2009: 122). This relationship has been tested empirically but found its theoretical roots in the work of Kaldor (Tregenna 2008), who argued that manufacturing is imbued with special characteristics not shared by the other sectors. Manufacturing growth ‘pulls along’ economic growth in a unique way (Tregenna 2008: S176) because of its dynamic economies of scale, such that the growth of productivity in manufacturing is higher the higher the rate of growth of output. ‘Learning by doing’ is seen as more important in industry than in agriculture or services, as are innovation and intersectoral linkages. Technological change in the rest of the economy is seen as generated by the manufacturing sector, as other sectors put into production higher productivity manufacturing inputs (Tregenna 2008: S176).
2. Reproductive activities can be thought of as the household level components of social reproduction. Jefferson and King (2001) note that the terminology and definitions surrounding ‘reproductive activity’ are



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

problematic; however, I eschew the alternative term 'caring activities' here for two reasons. First, as Urdang (2006: 166) notes, the term 'caring activities' carries a certain mystique that obfuscates the debate. Second, caring activities can be wrongly envisaged as being related only to child or adult personal care, particularly in a discussion of the impact of HIV/AIDS, when in fact reproductive activities include a wide range of care and household maintenance tasks.

3. The method and rationale of construction of satellite gross household production accounts is discussed in Young (2000) and Budlender (2004). Budlender and Brathaug (2002) estimate that for South Africa, non-SNA production activities would be equal in value to between 11 and 55 per cent of GDP. Using a more limited set of time-use information, Fontana (2002) estimates that non-SNA production might be valued at 21 per cent of GDP in Zambia. See Young (2000) for a review of the various methodologies used to measure the size of the reproductive economy. Different methodologies provide different estimates of the size of the reproductive economy. A common approach, which usually yields the highest estimates, is to apply a specialist wage rate to input activities. However, Marilyn Waring reminds us that we need to consider carefully the measures of value obtained in this way, as the market-equivalent functions of reproductive work are 'sex-segregated or sex-stereotyped jobs,' so the wages paid to such workers will reflect gendered disadvantage (1989: 280).
4. As the SNA economic production boundary includes paid care work carried out by employed carers, it should therefore be included in GDP measures, provided data is collected appropriately.
5. Indeed, some studies of households affected by HIV/AIDS, such as the study by Booyesen and Bachmann (2002), select respondents through NGO or government-care organizations, which may lead to a bias against the identification of those relying on private carers.
6. The models by Arndt (2003) and MacFarlan and Sgherri (2001) make passing reference to the productivity lost by workers due to time taken from work to care for sick relatives. However, this is not discussed in any detail, and no information is given about how this relates to other factors leading to a fall in productivity.
7. As Çağatay *et al.* (1995) have shown, a shift between the productive and



A POLITICAL ECONOMY PERSPECTIVE ON THE IMPACT OF AIDS

Excerpted from *Economics and HIV*

reproductive activities has important implications due to the different social and economic implications in terms of who carries out activities and the income that they bring. Poor women have borne the brunt of increased reproductive burdens, trying to replace lost public provision of services.

8. Micro examples of the impact of AIDS in light of gendered labour market segmentation can be found in studies of smallholder agriculture in areas of high HIV/AIDS prevalence. The impact of HIV/AIDS is likely to be determined by the gender-segmented nature of agricultural tasks. Baier's (1997) study placed particular emphasis on the impact that higher female prevalence rates have on food production, given that women dominate food production in the areas of East Africa studied. Conversely, Yamano and Jayne (2002) found that household agricultural production suffered the greatest impact when the male head of household died but was able to recover from the effects of another adult death. One reason may be the important role of male household heads in cash-crop production in the regions they studied. Haslwimmer's (1994) study of agriculture in Rakai, Uganda, suggested that with the death of a spouse, a husband or wife does not automatically assume the dead spouse's role, because of gender norms.