WHO GUIDE TO IDENTIFYING THE ECONOMIC CONSEQUENCES OF DISEASE AND INJURY

Department of Health Systems Financing
Health Systems and Services

WHO GUIDE TO IDENTIFYING THE ECONOMIC CONSEQUENCES OF DISEASE AND INJURY

Department of Health Systems Financing
Health Systems and Services
World Health Organization
Geneva, Switzerland

WHO Library Cataloguing-in-Publication Data

WHO guide to identifying the economic consequences of disease and injury.

1.Public health - economics. 2.Cost of illness. 3.Health expenditures. 4.Financing, Health. 5.Wounds and injuries - economics. I.World Health Organization. Dept. of Health Systems.

ISBN 978 92 4 159829 3 (NLM classification: WA 30)

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The views expressed are those of the responsible officers and do not necessarily reflect those of the World Health Organization.

Responsibility for the content in this document lies with staff members of the Department of Health System Financing in the Health Systems and Services Cluster at the headquarters office of WHO in Geneva, Switzerland (Dr David Evans, Dr Tessa Tan Torres Edejer, Dr Dan Chisholm and Dr Anderson Stanciole).

The Department wishes to acknowledge and thank the following for their critical inputs:

1) for the conceptual foundations section, we build on an earlier draft WHO discussion paper by Dr Kenji Shibuya, Dr Ajay Tandon, Dr Arnab Acharya, Dr Chris Murray, Dr Tessa Tan Torres Edejer and Dr David Evans; 2) for the macroeconomic impact section, we build on a draft literature review conducted by Ms Silja Baller and Dr Taghreed Adam on behalf of the Department; 3) for the microeconomic impact on firms, we acknowledge the input of Professor Glen Pederson (University of Minnesota, USA), who provided the first draft of this section.

We also wish to thank all participants of the *WHO consultation on methods for identifying the economic consequences of disease and injury*, held in Venice, Italy on 12-13 June 2008:
Dr Arnab Acharya (London School of Hygiene & Tropical Medicine);
Professor Clive Bell (University of Heidelberg)
Professor David Bishai (John Hopkins University)
Dr Jane Chuma (Kenya Medical Research Institute)
Professor Indrani Gupta (University of Delhi)
Professor Chuc Nguyen Thi Kim (Hanoi Medical University)
Dr Mead Over (Centre for Global Development)
Professor Glenn Pederson (University of Minnesota)
Dr Lorenzo Rocco (University of Padua)
Professor Richard Smith (London School of Hygiene & Tropical Medicine)
Dr Rosalia Vazquez-Alvarez (University of St Gallen)
This meeting could not have proceeded without the outstanding technical and administrative support of staff in the WHO European Office for Investment for Health and Development in Venice: Dr Erio Ziglio, Dr Marc Suhrcke and Ms Bianca Bortot.

Finally, we wish to acknowledge the WHO regional focal points for health financing (a number of whom were able to attend and contribute to the above consultation): Mr D Bayarsaikhan, (WPRO); Dr Alaka Singh (SEARO); Dr Joses Kirigia (AFRO); Dr Eduardo Levcovitz (AMRO); Mr Joseph Kutzin (EURO); Dr Hossein Salehi (EMRO).

EXECUTIVE SUMMARY

Introduction

Distinct from but complementary to clinical or epidemiological approaches to disease burden assessment, analysis of the economic impact of ill-health can address a number of policy questions concerning the consequences of disease or injury. Some of these questions relate to the microeconomic level of households, firms or government – such as the impact of ill-health on a household's income or a firm's profits – while others relate to the macroeconomic level, including the aggregate impact of a disease on a country's current gross domestic product or its future growth prospects. Resulting estimates - for a particular disease, injury type or for diminished health status in the population generally - can usefully inform decision makers about the overall magnitude of economic losses and their distribution across a number of key drivers or categories of cost. Although insufficient as a basis for setting priorities and allocating resources in health – for which data on effectiveness are also needed – economic burden studies may help to identify possible strategies for reducing the cost of disease or injury via appropriate preventive action or treatment strategies.

The number of economic impact studies in health has grown exponentially since the codification of a 'cost-of-illness' framework in the mid-1960s. Although most studies continue to use some variant of this methodology (which combines the 'direct' costs of medical care, travel costs etc. with the 'indirect' cost of lost production because of reduced working time), macroeconomic growth models have increasingly been used to better understand the dynamic and multifaceted nature of losses at the societal level. There has also been increasing policy and research interest in better understanding the microeconomic consequences of ill-health, particularly at the household level in lower-income countries.

Looking across the large body of existing literature, it is apparent that there is a considerable degree of methodological heterogeneity, and also that many studies suffer from a range of conceptual deficiencies. In light of these methodological shortcomings, as well as the strong continuing demand for economic impact studies in health, WHO is proposing a defined

conceptual framework within which the economic impact of disease or injury can be considered and appropriately estimated, with a view to enhancing the consistency and coherence of economic impact studies in health.

Conceptual foundations

Ill-health can contribute to losses in individual utility or social welfare in a number of defined ways, both directly (because people prefer to be more healthy than less healthy) and indirectly by reducing the enjoyment or utility associated with the consumption of goods and services unrelated to health, or by compromising other economic objectives such as producing income that allows people to consume market goods. Since the consumption of health goods and services in general does not yield utility or welfare directly - people would prefer not to incur these expenses in terms of money and time - the key direct determinants of economic welfare can be summarized as the consumption of *non-health* goods and services, leisure, and health itself. It is the impact of disease or injury on these domains of economic welfare that should form the basis of estimation.

In order to have a clear economic meaning, it is vitally important that economic impact studies be explicit and consistent about which of these domains of welfare are to be captured. Is it the impact on economic welfare in its entirety that is being assessed, or just some specified component(s) of it? At the macroeconomic level, for example, the impact of illness on gross domestic product (GDP) - both now and in the future - is something that is measurable and has a clear economic meaning. On the other hand, combining health expenditures with (market and some non-market) production losses and expressing this in relation to GDP – as cost-of-illness studies tend to do – does not have a clear economic meaning.

Economic impact studies also need to be clear as to the appropriate counterfactual being used (the comparator situation against which economic losses can be assessed). Is it being assumed that the disease or risk factor never existed, or just that no new cases are assumed to occur in the current period and/or in the future? Again, the decision to adopt a particular approach will be determined by the underlying question; for example, a prevalence-based approach (in which new

as well as pre-existing illness in a given year is assessed) is more suitable for ascertaining the total current economic burden of a disease, whereas an incidence-based approach (in which only new cases are included) is more useful for ascertaining the expected impact of a disease in the future (and its potential prevention). Traditional cost-of-illness studies to date have broadly employed a variant of the prevalence-based approach, in that they estimate disease-related intervention costs for a given year (but *not* future years), plus the present value of lost production in future years associated with deaths in the current period. This seems to be a peculiar and also inconsistent approach to cost estimation and counterfactual analysis.

Macroeconomic analysis of the consequences of disease and injury

A macroeconomic approach to assessing the impact of ill-health should be concerned with establishing the aggregate impact of disease and injury across different economic agents on three areas related to economic welfare (both now and in the future): non-health consumption possibilities, leisure time and health status.

Most economic impact studies at the societal level have focused on gross domestic product (GDP), which represents market consumption opportunities. While this has a clear meaning, it is important to note that GDP includes expenditure on health goods and services, so this component should be omitted and the focus of analysis be redirected towards establishing the present value of discounted aggregate flows of current and future consumption of non-health related goods and services linked to disease. Key channels through which disease or injury can impact on macroeconomic performance or output include increased health expenditures, labour and productivity losses, and reduced investment in human and physical capital formation.

Although the cost-of-illness approach concerns itself with the societal impact of disease or injury, it would appear to fall some way short of providing an adequate model at the macroeconomic level. By focusing on health sector spending and lost labour productivity only, CoI studies provide only a very partial picture of the true macroeconomic impact of disease, and fail to consider the contribution of depleted capital accumulation, investment in human capital and

demographic change to diminished economic growth. Instead, we recommend the application of a more general and dynamic assessment of forgone consumption opportunities.

The choice of modeling framework should be guided by the question that is being asked. If the focus lies on the overall (non-health) GDP impact only, an economic growth model based on either calibration or estimation might be used. The calibration approach can be particularly useful if detailed data is lacking. The evaluation of more disaggregated impacts, such as on specific sectors or categories of macroeconomic agents might require the specification of a computable general equilibrium (CGE) model. This type of model can also fully incorporate issues of dynamic adjustments and the inter-linkages across settings. However, the data requirements and computational costs tend to be high. If the interest lies in the determination of ill-health on overall economic welfare, models based on willingness-to-pay (WTP) valuation techniques - which enable non-marketed goods and services as well as health itself to be valued in economic terms - would need to be adopted. Conceptually, such an approach provides a more appropriate measure of total economic welfare losses resulting from disease or injury, but there are a number of empirical concerns around its practical implementation which mean that considerable uncertainty surrounds derived estimates. It is therefore recommended that (empirically-based) estimates of market losses be separately identified and reported from (hypothetically-based) estimates of foregone welfare.

Microeconomic analysis of the consequences of disease and injury

Distinct from concerns over the societal or population-level impact of ill-health, there is increasing policy and research interest in the microeconomic impact of disease or injury, focusing in particular on the impoverishing and other effects that ill-health or injury can have on the consumption possibilities of households. Illness typically leads to increased household expenditures on health services and goods, and may also reduce time spent producing income that allows them to consume market goods. In response to this change in income and/or expenditure, households may reduce their consumption of non-health goods and/or liquidate household savings or assets (and by so doing diminish their opportunities to generate the stock of financial and physical capital that will enable it to maintain or increase its consumption possibilities in the future). Furthermore, ill-health can interfere with the consumption of non-market activities (e.g.

giving up unpaid housework or leisure time to look after a sick household member) and of course reduces the stock of health itself. Accordingly, analysis of the trade-offs that households make between consumption in current and future time periods, and in their time allocations to market production, non-market production, health improvement and leisure need to reflect *time preferences* (consumption today versus consumption next period) and also *risk preferences* (certain consumption versus uncertain consumption).

An important economic consequence of disease or injury at the microeconomic level of households, firms and government is that, through its impact on functioning, individuals are unable to perform their usual day-to-day activities. Economic impact studies have largely relied on an input-based approach to the measurement and valuation of these production losses (which assumes that the duration of an individual's absence from work fully corresponds to the market value of those lost days). Use of such an approach can be expected to overestimate these economic losses, because it overlooks the 'coping strategies' or compensating mechanisms used by households or firms to mitigate the adverse circumstances imposed by sickness. An output-based approach that measures actual (rather than potential) net losses in income or market / non-market production provides a more robust and accurate basis for estimation.

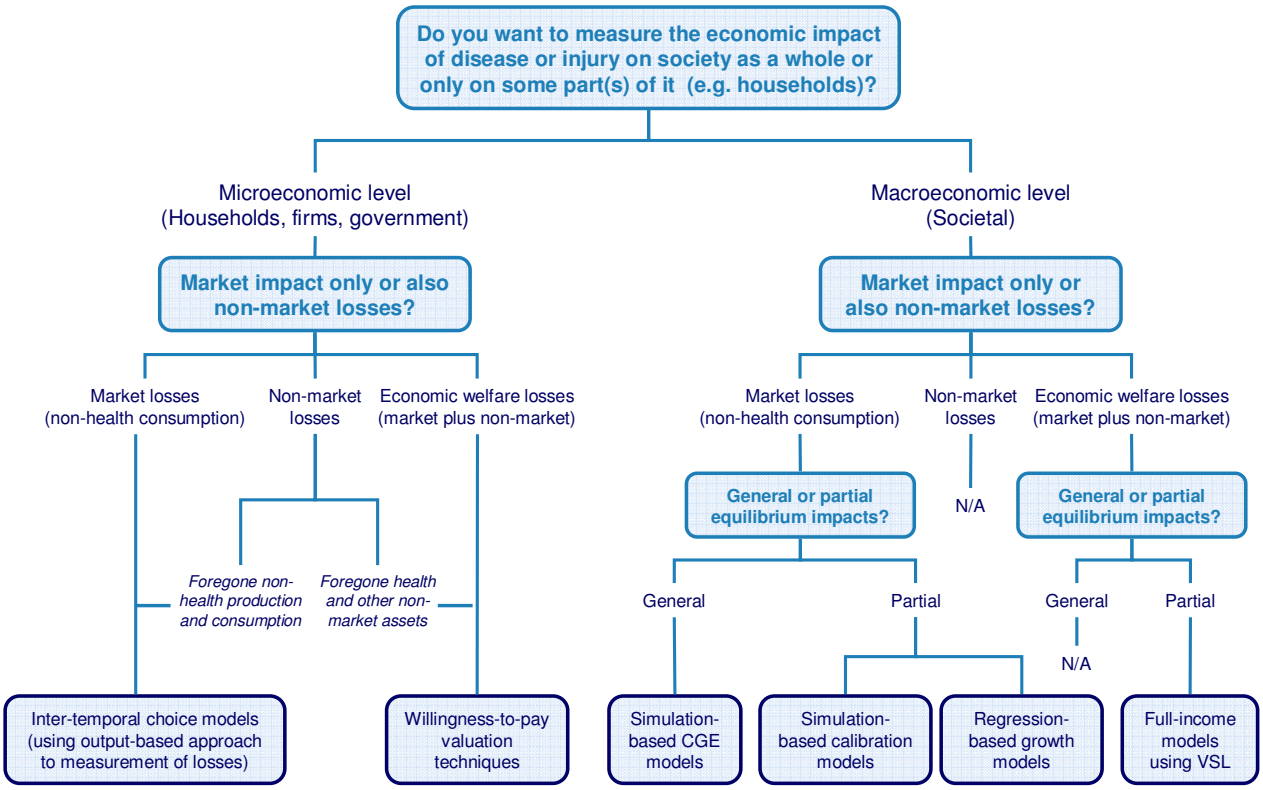
Valuation of the full economic welfare impact of illness on households - including the value of reduced health itself - requires the application of willingness-to-pay measures. Since there remains considerable uncertainty around the appropriate value to apply to these health-related welfare losses, we recommend separate reporting of market and non-market losses in economic impact studies and advise great caution in the use and interpretation of WTP-based measures.

Conclusion

If undertaken in a defined manner with reference to a coherent set of conceptual foundations, economic impact studies in health can usefully contribute to health policy dialogue. Too often in the past, however, studies have not been founded within a clear, logical framework, meaning that they produce results that can be misleading or spurious. Much greater attention is therefore called for when considering or planning an analysis of the economic impact of disease or injury.

Is an economic impact study needed in the first place (what will it bring in addition to clinical or epidemiological indicators of disease burden)? What is the policy decision that it addresses or bears upon? What is the explicit purpose, scope and perspective of the study? What are the key channels through which economic impacts are expected to be felt, and what are the data (or other) constraints to their appropriate measurement and valuation? Dealing with these questions at the outset will encourage a more rigorous approach and a more meaningful assessment to economic impact studies in the future. It will also help to identify which of a number of specific measurement approaches or models might be appropriate to use in a given context (see for example the simple algorithm below; detailed explanation of identified models can be found in the main body of the guide).

Algorithm for determining what methodological approach to use for economic impact studies in health



Abbreviations used: N/A: Not available; CGE: Computable general equilibrium; VSL: Value of statistical life

1. INTRODUCTION

1.1. Policy context: why measure the economic consequences of disease?

Measuring morbidity and mortality are key considerations for estimating the burden of disease in populations. However, only focusing on morbidity and mortality effects provides an incomplete picture of the adverse impact of ill health on human welfare. In particular, the economic consequences of poor health can be substantial. Health 'shocks' – such as unexpected increases in health expenditure, reduced functional capacity and lost income or productivity – are often a primary risk factor for impoverishment (WHO, 1999; Xu et al., 2003). Poor levels of health may also adversely impact educational attainment and consequent levels of future income. At a societal level, poor population health is associated with lower savings rates, lower rates of return on capital, and lower levels of domestic and foreign investment; all of these factors can and do contribute to reductions in economic growth (Ruger et al., 2006). Measurement of these various adverse impacts provides decision-makers with an indication of the extent to which a specific disease or, more generally, depleted health status disrupts or reduces economic production or consumption opportunities at the household or societal level.

Set against these negative impacts is the contribution that health can make to economic growth (WHO, 2001; Bloom, Canning and Sevilla, 2004). The positive association between health and wealth constitutes a vital argument in the justification for greater investment in health systems and services. Therefore, appropriate measurement and valuation of the economic benefits that accrue from the reduction or elimination of disease represents a further important, and more positive, reason for undertaking economic impact studies in health.

There are in fact a number of health policy questions that are or can be addressed by economic impact studies, not only at the microeconomic level of households, firms and government agencies, but also at the aggregate macroeconomic level in terms of the impact of ill-health on gross national income or product (see **Box 1** below for some examples of adverse impacts).

Interest in measuring the economic consequences of illness has a long history and shows no sign of abating. The first mention of this type of work in official documents of the World Health Organization (WHO), for example, is found in 1951, only three years after the Organization was established. Winslow (1951) advocated the collection and dissemination of evidence about the potential economic benefits associated with public health interventions as a method to persuade governments to allocate more resources to public health. Since then, the literature has expanded rapidly, particularly in the last two decades. By 1980, Hu and Sandifer (1981) could identify just over 200 studies. Based on a search of the Library of Congress Web of Science database, using the subject heading 'cost-of-illness', we have found 191 new cost-of-illness papers published during the 1990s, and another 450 between 2000-2005 (mostly in medical or public health, rather than in economics or health economics journals).

Even though there is general agreement about the channels via which poor health can have economic effects (see, for example, Ruger et al., 2006), there have been many problems with the methods used to obtain empirical estimates. Most studies undertaken to date have set out to derive a national-level cost estimate of the impact of a specific disease, based on some version of the cost of illness (CoI) approach, which was formalized by Dorothy Rice and colleagues in the late 1960s and subsequently revised on several occasions (Rice 1966, 1967; Cooper and Rice, 1976, Rice, Hodgson and Kopstein, 1985). Using this approach, the possible economic consequences of specific illnesses are divided into 'direct costs' - the expenses incurred because of the illness (including medical care, travel costs, etc.) and 'indirect costs', the value of lost production because of reduced working time. Rice and colleagues did not try to measure the costs of pain and suffering, describing them as 'intangible costs'. Direct and indirect costs are then summed to provide the overall cost the illness imposed on society, often expressed as a percentage of current period gross domestic product (GDP).

The variety of different approaches taken to estimating direct and indirect costs has limited the comparability of results across studies (Ettaro et al., 2004; Hu, 2006; Segel, 2006), while the use of a 'human capital' approach to the measurement of lost production - which essentially multiplies the total period of absence and by the wage rate of the absent worker - lacks a robust theoretical foundation and might well overestimate actual economic losses (Hodgson and Meiners, 1982;

Koopmanschap et al., 1995). These criticism have not prevented the very rapid growth of the literature and the belief among many people in the public health community, echoing Winslow, that information on the economic consequences of illness will help to attract greater funding for the control of particular diseases. We agree that sound economic analysis can provide important arguments for greater investments in health, but argue that such analysis needs to be undertaken in a credible and scientific way.

One reason for this methodological disorientation relates directly to a lack of clarity in what such studies think they are measuring. Put another way, studies are not clear about the precise nature of the question that they are trying to address or answer.

Box 1 Illustrative health policy questions addressed by economic impact studies

Level	*Question / topic*	*Section*
Macro: Society	1. What impact does ill-health have on gross domestic product or its rate of growth?	**3.1 - 3.2**
	2. How much does society pay for medical and other expenses because of illness?	**3.1.1**
	3. What is the impact on social product (i.e., both market *and* non-market consumption lost opportunities), or on social welfare more generally?	**3.3**
Micro: Households	1. What impact does ill-health have on a household's income or consumption patterns (over a single year, or for a longer period of time)?	**4.1**
	2. How much do households pay for medical or other expenses because of illness (for an episode, over a year, or over a lifetime)?	**4.1.2**
Firms	1. What impact does ill-health have on a firm's operating costs, output or profit?	**4.2**
	2. What is the impact of ill-health on productivity in the work place (including impaired performance while still at work, as well as absenteeism) ?	**4.2.3**
Government	1. What proportion of government expenditure could have been saved and directed to an alternative use in the absence of illness? *(e.g. what social security payments could be avoided by the prevention of or cure for disease?)*	**4.3**
	2. What impact does ill-health have on the government workforce and on the government's ability to provide services?	**4.3.2**

In relation to **Box 1**, for example, the first household-level question - the effect of disease or injury on household income - could be interpreted to include only market production, in which case the underlying quantity of interest is the value of monetary income lost because of illness. If, on the other hand, there is interest in ascertaining the economic losses associated with both market and non-market production (e.g. by including informal care-giving for a sick household member), the quantity of interest would extend beyond purely financial losses to a broader concept of foregone economic welfare. Similarly from the macro perspective, where policy makers are interested in the effect of ill-health on GDP, the underlying quantity of interest is market production foregone because of illness. For the final macro question, however, the quantity of interest is clearly a different and broader notion of economic loss (social product or social welfare).

1.2. Motivation, scope and objectives

In order to redress some of the issues raised above, WHO is proposing a defined conceptual framework within which the economic impact of diseases and injuries can be considered and appropriately estimated, with a view to enhancing the consistency, comparability and coherence of economic impact studies in health. Specifically, we focus in Section 2 on the conceptual foundations of economic impact studies, which covers a number of core issues such as study perspective, timescale and scope, and then turn in Sections 3 and 4 to the implications that choices around these issues have for the measurement and valuation of costs at the macroeconomic and microeconomic level. We summarize above in **Box 1** which sections of this guide contain particularly relevant discussion relating to different health policy questions.

The scope of these guidelines should be noted. The aim is to provide an overview of the different approaches that have been used to measure the economic impact of health, consider their merits and shortcomings with respect to a specified conceptual framework, and offer recommendations for improved practice in the future. In this sense, it is closer to a methodological review and proposal than a step-by-step user guide to the implementation of an economic impact study. The production of such practical guides or manuals will follow from the application of these general

guidelines to specific disease areas, which are likely to differ with respect to the key questions to be addressed and the consequent measurement approaches to be employed.

The target audience for this guide comprises two groups. First, it responds to a demand from WHO departments, country offices and disease programmes interested in or requested to carry out studies on the economic consequences of specific diseases. This guide will assist economic impact studies in this area by providing guidance on choices about the available methods, objectives, limitations, data availability and the level of technical expertise that is required. The second group includes decision makers and technical specialists from international organizations, national governments and academic institutions, which will also hopefully benefit from the presentation of an uniform framework for assessing the economic consequences of disease and injury. Overall it is hoped that this guide will contribute to improve standards of practice and, by increasing the degree of comparability and uniformity between studies, promote more informed discussion in this field.

It is very important to stress at the outset of this guide that the purpose of studying the economic impact and burden of disease is distinct from analysis linked to the allocation of scarce resources to a range of possible health interventions. The resource allocation question requires information on both costs and benefits of the interventions and is aimed at setting priorities among different interventions by employing cost-effectiveness analysis or cost-benefit analysis (see, for example, Drummond et al., 2005 or the *WHO guide to cost-effectiveness analysis*; Tan Torres et al., 2003). On the other hand, the study of the economic impact of disease estimates costs by comparing the current status with a disease with a hypothetical counterfactual (such as the situation without disease). While these two analyses are inter-related, the purposes of these two exercises should not be confused. Priority cannot be given to diseases or health problems simply because they have the largest economic impact. Additional information is required on the costs required to reduce the burden and the marginal effect of each additional unit of expenditure (Shiell et al., 1987; Drummond, 1992; Wiseman and Mooney, 1998, 2000). Economic impact assessment is a way of highlighting the importance of a particular disease to balance the traditional epidemiological assessment of mortality or morbidity (Drummond, 1992). Its primary purpose should be to inform decision makers by providing descriptive indicators of the magnitude of a

disease or a health problem as a complement to methods of deciding how scarce resources should be used to improve health. A secondary aim of a small sub-set of economic impact studies has then been to isolate the fraction of economic costs attributable to a certain disease or risk factor that is *avoidable* via a range of effective policy measures or interventions (see, for example, Rehm et al., 2006, 2008). From a health policy perspective, estimation of these avoidable costs is a highly relevant undertaking that should be strongly encouraged.

2. CONCEPTUAL FOUNDATIONS

In this section, we discuss a number of key conceptual issues or questions that need to be addressed at the outset of any planned economic impact study in health, since they have a direct influence on what a study can expect to achieve and how it will be carried out. Some of the main questions are: What are the main channels through which ill-health impacts on households, firms or on society as a whole? What is the analytical perspective of the study, and the corresponding quantity of interest that underlies this perspective? What is the scope of the study, in terms of the consequences that are included or excluded from analysis? What is the appropriate comparator situation or counterfactual scenario to be used, against which measured economic consequences can be assessed? And what is the timeframe for the study, in terms of capturing future as well as current economic impacts of disease or injury?

2.1. Ill-health and its economic consequences

In order to articulate the possible ways in which disease or injury may lead to economics losses, it is necessary to start by considering what it is that people or societies value. According to welfare economic theory, and subject to various constraints including income and time, individuals or populations seek to maximize utility (the term economists use to describe economic welfare). They do this by combining to best effect their consumption of a range of goods and services - some of which can be bought and sold (including health care), and some of which cannot but nevertheless have discernible value (e.g. home-grown produce that is directly consumed rather than sold). In addition to the consumption of goods and services, individuals or populations also generate utility via other means, such as taking care of others (without financial compensation), and spending time with family and friends or in other forms of leisure.

Health contributes to individual utility or social welfare in three ways. First, people prefer to be more healthy than less healthy (i.e. health directly affects utility). In economic terms, it is an argument in the utility or social welfare function. Second, the enjoyment of consumption of other goods and services is partly influenced by the level of health (i.e. marginal utility derived from

consumption is partly a function of health status). Third, without good health other economic objectives, such as producing income that allows people to consume market goods, stand to be compromised; in other words, it is instrumental to an individual's or community's capability to undertake desired activities or functions (Sen, 1985). One qualification that needs to be made is that while the consumption of most types of goods and services yields welfare directly, the consumption of health goods and services does not. People would prefer not to incur these expenses in terms of money and time, but do so because they believe it will protect or promote their health. Accordingly, the key direct determinants of economic welfare are the consumption of 'non-health' goods and services, leisure, and health status.[1]

On the basis of this standard framework - for a formal exposition, see Appendix A - it is possible to identify a number of ways in which ill-health could negatively influence the choices and preferences upon which different economic agents seek to maximize utility. The specific channels through which these consequences of disease or injury impact households, firms and governments are expected to differ somewhat, and are discussed at length in Section 4 of this guide. Here, we provide a brief summary of the main channels and impacts:

Households: The impact of ill-health on a household can be measured in terms of its impact on the consumption of non-health goods and services (market and non market), leisure, health status - which represent the essential components of welfare, as explained above - or in terms of the overall change in welfare. The mechanisms through which it influences current and future consumption are manifold. For example, and particularly in lower-income countries with a high proportion of direct out-of-pocket health spending, ill-health will drive up household consumption of health-related services and goods at the expense of non-health goods and services. By increasing the time spent seeking care or in states of health that prevent work, it can also reduce production of both market and non-market goods, and through this, consumption. The impact is not just limited to the current time period; health services and goods may be paid for out of current income, but could also be financed from cash savings if available, or if not, via

[1] More complex formulations which involve altruism and bequests to future generations, for example, could be formulated, but they do not change the essential features of the conceptual framework. 'Non-health' consumption is defined as all consumption unrelated to the disease or injury in question; in the case of certain conditions like substance abuse which impose wider social costs, this would also exclude, for example, expenditure on criminal justice and security services.

a loan or the sale of household assets (e.g. dis-savings). Reduced household income, savings and assets resulting from the consumption of health services and goods may in turn lead to depleted investment in (physical, financial and human) capital. These factors influence consumption possibilities in the future.

Firms: Ill-health can reduce the productivity and efficiency of a firm, which may negatively affect its earnings and profits, its ability to invest profits into new capital accumulation and thereby reduce the wealth or consumption possibilities of its owners. Productive capacity is partly determined by factors other than human capital (such as technology), and potential economic losses from work absence due to ill-health may be partially or fully compensated for by other workers. However, this implies higher costs for the firm through the need to keep some 'excess capacity' in the event of illness to some staff members. Firms may also devote a proportion of their operating activities to health-related expenditures and benefits for their employees.

Government: Governments essentially produce public goods and redistribute income, although the definition of what constitutes a public good is often very broad. Illness in its employees can reduce the output of public goods or increase the cost of producing them in the same way as with firms. However, governments are often more concerned with the impact of ill-health in the population on its financial expenditures and receipts. These relate to the increased costs of providing or financing health services, increased social security payments including disability or unemployment benefits, and reduced tax receipts. From an economic perspective, it is important to note that such redirection of financial flows represent transfer payments and therefore do not appear in national income accounts (because no goods or services were required in return).

As argued earlier, the key direct determinants of economic welfare at the *societal* level are the consumption of non-health goods and services, leisure and health status; the pathways through which these can be affected by good or bad health mirror the impacts on households. The economic impact of disease on society, however, is often expressed in reference to gross domestic product (GDP), which is simply the aggregation of household consumption, government expenditure and investment by firms (in non-financial products). However, GDP is a partial measure of overall economic welfare, reflecting the command over the consumption of

market goods and services. Even then, it includes expenditure on health which would not add directly to welfare except to the extent that they lead to increases in health status (discussed below in Section 3.3). GDP does not include non-market consumption, the value of leisure or the value of health itself, which are all components of the broader concept of economic welfare.

One important distinction between the concepts of household utility versus social welfare is that societies also care about distributional issues, and quite complex functional forms have been developed to express social welfare under different assumptions of social preferences for equality (Atkinson, 1975; Sen, 1973). This implies that measuring the full effect of ill health on economic welfare would require monitoring who is affected by illness and the subsequent changes in non-health consumption flows and leisure. To our knowledge this has not been attempted to date, and we do not elaborate further on this point ourselves in this guide.

2.2. The quantity of interest for economic impact studies in health

One of the first principles of any costing study is to define the viewpoint of the analysis, because the way costs should be measured depends crucially on the reason for measuring them, i.e., on the question being asked. This principle can also be applied when developing methods for measuring the economic consequences or costs of disease where the underlying quantity of interest can differ according to the question being asked. This is clear from **Box 1** where the economic consequences from the perspective of a household may imply a different quantity of interest to a study on the consequences to the overall economy.

For example, the effect of disease or injury on a household's income or consumption opportunities (household question 1 in **Box 1**) could be interpreted to include only market production and cash purchases, in which case the quantity of interest is the value of market consumption opportunities foregone because of illness. On the other hand, the quantity of interest could be redefined to include the economic losses associated with both market *and* non-market production or consumption opportunities. Finally, it could be defined even more broadly to include the value of the lost welfare associated with illness, which would include not just the lost

production and consumption, but the value of the changes in leisure and the decrement in health status itself. Similarly from the macroeconomic perspective, policy makers might be interested in lost market consumption opportunities, often proxied by GDP. They might also be interested in some measure of 'social product' (incorporating market and non-market production and consumption) or overall welfare. On the other hand, they might also be interested in quite narrow questions, such as the extent to which paying for health diverts consumption from other possible uses. Each of these questions implies a different quantity of interest. As long as they are measured correctly, each of them has a clear economic meaning. In general, however, the estimates derived from answering different questions should not be combined because the quantities of interest are different.

One of the most common problems emerges from application of the traditional cost-of-illness (CoI) methodology and the combination of direct and indirect costs. The implicit quantity of interest for direct costs is the 'non-health market consumption' possibilities foregone (because expenditure on health services and goods reduces opportunities to consume other services and goods, such as food or clothing). Indirect costs reflect some measure of lost 'social product' in that they typically add some estimate of lost non-market production/consumption to lost market production/consumption - but since this notion of 'social product' includes lost consumption of health goods and services, a preferable concept would be that of 'non-health' social product. In short, the underlying quantities of interest for direct and indirect costs are different, so the resulting estimate is difficult to interpret. It certainly cannot be interpreted as a loss of GDP because health expenditure forms part of GDP.

It would be possible, at least in theory, to define a quantity of interest that allows the addition of direct and indirect costs if that is desired. As implied above, it is non-health market consumption opportunities or non-health social product. Illness requires households (and society) to pay for health services, preventing them enjoying the consumption of other types of goods and services. Disease reduces household earnings (or economic production at the aggregate level), which in turn reduces present and future consumption of health and non-health goods and services. To add both together, the health component of consumption would need to be removed from the indirect

cost calculations. These calculations, although feasible, are somewhat complex - for example, the impact of direct costs on savings (or borrowings) and investment would have to be included.

The alternative would be to directly measure the welfare loss associated with the illness. This can be done for either the household or the society as a whole using a variety of techniques. The validity of these techniques are addressed in subsequent sections dealing with the measurement and valuation of the economic impact of disease or injury.

2.3. The choice of counterfactual for economic impact studies in health

All economic impact studies, regardless of whether they focus on households, firms or the population as a whole, require that the economic burden or costs associated with a particular health condition be measured against an alternative scenario or 'counterfactual'. The chosen counterfactual in an economic impact study provides the comparator situation against which the current or future burden of a disease can be calculated.

For many disease entities, the choice of counterfactual is simply stated as the situation that would prevail in the absence of the disease. For example, what would be the economic situation in a country if no cases of HIV had occurred in a given year (or ever)? Concerning risk factors for disease such as tobacco use or unsafe sex, the choice of counterfactual is not so obvious, since the total absence of the risk may not be possible, feasible or desirable. To illustrate, smoking or unsafe sex have harmful health consequences, but can also create some sort of immediate utility or satisfaction for the individual. Consequently, individual rational decision-making can result in a non-zero welfare-maximizing level of engagement in these risky behaviors. Similar arguments can also be made from the strict medical sense. For instance, excess consumption of salt represents a risk factor for cardiovascular disease, but reducing levels of daily salt intake to zero is not desirable because sodium makes up one of the essential minerals required for human survival. Therefore, the theoretical minimum risk of exposure to a disease or risk factor may be quite different from the plausible or feasible minimum level of risk, which implies that a range of

counterfactual situations can be considered and that the choice may be determined by the desired policy relevance of the study (Murray et al., 2003; Greenland, 2005).

Identification and specification of an appropriate counterfactual requires careful consideration of temporal issues, such as whether to exclude cases of disease that occurred in the past (that may or may not be amenable to intervention), to include all prevalent cases or only new (incident) cases of disease in a given year, or to include downstream future costs of disease. Again, the decision to adopt a particular approach will be determined by the underlying question; for example, a prevalence-based approach is more suitable for ascertaining the total current economic burden of a disease, whereas an incidence-based approach is more useful for ascertaining the expected impact of a disease in the future (and its potential prevention). The tradition that has emerged from the cost-of-illness literature has been to distinguish two approaches:

- *Prevalence-based studies*: The economic impact of all disease cases (new as well as pre-existing) in a given year is included, together with any deaths attributable to the cause in that year discounted back to present values. The implied counterfactual is that no (new or pre-existing) illness or death occurred in this time period. A variant of this would be to assume that no new or pre-existing illness / deaths were to exist from this time onwards.

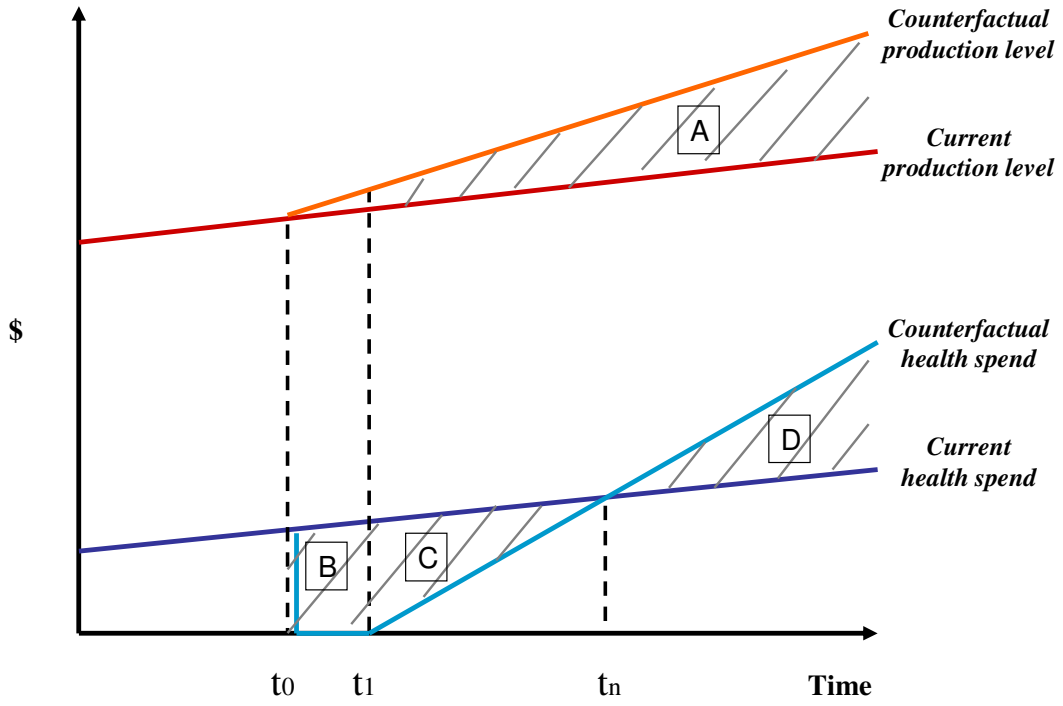
- *Incidence-based studies*: Economic impact is measured in terms of the present value of lifetime (or other specified time period) costs of new illnesses or deaths in the current/index year, so the counterfactual is that there is no new morbidity or mortality in the specified time period (but pre-existing morbidity would remain). A variant on this would be to assume that no new illness or deaths were to exist from this time onwards.

A third option, found for example in studies of the economic consequence of smoking (Barendregt et al., 1997), is to assume that the risk factor or disease had never existed. Cost-of-illness studies to date have broadly employed a prevalence-based approach to costing, i.e. they estimate disease-related intervention costs for a given year (but not future years), and the present value of lost production in future years associated with deaths in this year. This seems to be a peculiar and also inconsistent approach to cost estimation and counterfactual analysis.

To illustrate this see **Figure 1**, which shows the impact of eliminating all new (and, in the prevalence approach, pre-existing) illness, injuries or deaths in the index year. The upper panel shows production over time - the source of consumption opportunities, both market and non-market. Without the disease, the labour force would have been larger and probably more productive, so the counterfactual production path is higher than the current path. Area A represents the production gains. The lower panel represents the opportunity costs of the health expenditures linked to this disease - the diversion from non-health consumption. Without any new or pre-existing cases this year, expenditures fall immediately to zero. However, in subsequent years health costs related to this disease or condition would slowly rise back to exceed previous levels because the people who would have died due to the disease in the index year now survive and increase the size of the household or total population. These people still face the same age- and sex-specific risks of disease because of the assumption that cases are eliminated for one year only. The initial gains from eliminating the disease this year (areas B + C) might well be offset by the eventual increased expenditure (area D). For example, Barendregt et al. (1997) used life table analysis to show that if all smokers were to quit, population-level health care costs would actually be greater in the long term, due to their increased survival and exposure to other risks to health. Traditional cost-of-illness studies do not include this consideration and therefore overestimate the present value of health resources that would be saved by the elimination of the disease for one year.

To rectify this, using the traditional prevalence approach, the present value of resources required to treat additional morbidity and mortality related to the disease in question in the future needs to be deducted from the savings in health intervention costs in the index year. In addition, some of the health-related resources (direct costs) freed for other uses because of the absence of one disease would have leaked back into health-related goods and services unrelated to that disease. This leakage would also need to be deducted from direct cost estimates for a specific disease if they are to be expressed in terms of non-health-related national product. Similarly, some of the increase in production (indirect costs) associated with reduced disease would also leak back into health-related goods and services. Again, from a consistent conceptual viewpoint these should be deducted to allow the appropriate estimation of costs for a specific disease.

Figure 1 **Economic gains and losses in a prevalence-based cost of illness study**
(counterfactual: no new or pre-existing cases of disease occur in index year)



Notes:

A	Increased production resulting from absence of disease X
B	Decreased health-related expenditure on disease X in time t_0 - t_1
C	Decreased health-related expenditure on disease X in time t_1 - t_n
D	Increased health-related expenditure on disease X after time t_n (survivors from year t_0 - t_1 increase population at risk in future years)

Figure 1 can be extended to each of the other possible counterfactuals. If we use, for example, the version of the incidence counterfactual in which all new cases of illness and death are prevented from this time forwards, the production gains would be as shown in the upper part of the figure, though eventually larger. The counterfactual health spending would fall more slowly to zero because pre-existing cases remain, but then remain at zero. From the viewpoint of public health planning and prevention of a disease or health problem, the incidence-based counterfactuals (i.e., no new cases this year or this year forward) are most appropriate, since they are more sensitive to current epidemiological trends and are better equipped to assess the

potential benefits of health interventions plus the expected long-term impact of current ill-health in the future (Hartunian et al., 1981; Oster et al., 1984). The most conceptually attractive of the two incidence-based scenarios would be to calculate the present value of the expected long-term impact of current ill-health in the future, especially for diseases with a possibility of elimination. This would capture the static and dynamic economic impact and burden of a disease.

2.4. Conceptual foundations: key points

Ill-health and its economic consequences

- Ill-health can contribute to losses in individual utility or social welfare in a number of defined ways, both directly (because people prefer to be more healthy than less healthy) and indirectly by reducing the enjoyment of consumption of goods and services unrelated to health or compromising other economic objectives, such as producing income that allows people to consume market goods.

- While the consumption of most types of goods and services yields utility or welfare directly, the consumption of health goods and services does not. People generally would prefer not to incur these expenses in terms of money and time, but do so because they believe it will protect or promote their health, i.e. they can increase utility indirectly through their impact on health, but face an opportunity cost in the sense of reducing potential non-health consumption. Accordingly, the key direct determinants of economic welfare are the consumption of non-health goods and services, leisure, and health

- The economic impact of disease and injury is not just limited to the current time period; health services and goods may be paid for from savings or additional borrowing or the sale of household assets, and thereby lead to depleted investment in (physical, financial and human) capital. These factors influence consumption possibilities in the future.

Quantity of interest in studies of the economic impact of disease

- The question being addressed by the study, and the associated quantity of interest, need to be stated explicitly. It is legitimate to ask a variety of questions (e.g. all of those in **Box 1**), and quantities of interest that have a clear economic meaning can be defined for each.

- The microeconomic and macroeconomic impact of disease or injury can be assessed in terms of lost non-health consumption, lost non-health social product, or lost economic welfare. These are all different, but each has a clear economic meaning (whether it is feasible to measure the impact of illness on them is addressed in Sections 3 and 4).

- Where interest lies with understanding the economic impact on a household, taking into account the 'direct' costs and the impact on lost production, non-health consumption opportunities represent an appropriate quantity of interest (or in a multiple period framework, the discounted present value of non-health consumption opportunities). Alternatively, an indicator of the overall welfare loss could be used, a broader concept.

- At the macro level, the impact of illness on GDP both now and in the future is something that is measurable, at least in theory, and has a clear economic meaning. However, it can be somewhat confusing conceptually, because increasing investments in health goods and services increases GDP, yet do not increase welfare directly.

Choice of counterfactual in economic impact studies in health

- As with the choice of quantity of interest, a number of counterfactual scenarios can be identified, so again economic impact studies need to be clear as to the question that they seek to address and the appropriate counterfactual to use.

- A prevalence-based approach is more suitable for ascertaining the total current economic burden of a disease, whereas an incidence-based approach is more useful for ascertaining the expected impact of a disease in the future (and its potential prevention or elimination).

3. MEASUREMENT AND VALUATION: MACROECONOMIC LEVEL

This section focuses on the impact of illness at the national level and evaluates the channels through which disease and injury impact overall economic welfare. As discussed above and in Section 4, efforts to identify, measure and value the economic impact of disease and injury from the *micro*economic perspective focus on economic agents - households, firms or government. By contrast, a *macro*economic approach is concerned with establishing the *aggregate* impact across these economic agents and by so doing derive a societal or economy-wide assessment.

As we argued previously, disease can reduce non-health consumption opportunities (market and non-market), both now and in the future. In addition, it can reduce leisure time as well as health status. All influence economic welfare. A key question is therefore to ascertain the extent to which different analytical approaches adequately capture any or all of these possible impacts. For example, many economic impact studies at the societal level have focused on gross domestic product (GDP), which represents market consumption opportunities. While this has a clear meaning, it is not totally consistent with the welfare theory described above because GDP includes expenditure on health goods and services. Ideally, we would want to omit this component from the analysis and identify the present value of the discounted aggregate flows of current and future consumption of non-health related goods and services linked to disease.

The most common methodology that has been used in order to derive a societal estimate of the impact of disease or injury is the cost of illness (CoI) approach, which combines 'direct costs' (medical care, travel costs, etc.) and 'indirect costs' (the value of lost production because of reduced working time) into an overall estimate of economic impact on society, often expressed as a percentage of current GDP. Although the CoI approach concerns itself with the societal impact of disease or injury, it would appear to fall some way short of providing an adequate model at the macroeconomic level. A more detailed critique of the CoI approach is provided in Appendix B, but three of the key limitations can be summarized as follows:

- It is not clear what overall economic impact CoI studies are trying to measure. Indirect costs involve lost opportunities for non-health market consumption, though not

necessarily in the current period alone. Indirect costs of formal sector workers represent lost health and non-health market consumption opportunities, again not necessarily in the current period alone. By including non-market production losses (such as the lost productive time of homemakers / informal caregivers), but valued at some sort of average shadow price to avoid discrimination, the attempt is to move closer to a form of 'social product', but the number emerging from this would only be a true assessment of non-market opportunity costs if shadow prices were normally distributed. The end result is that it is impossible to define an implied quantity of reference for CoI studies.

- By focusing on health sector spending and lost labour productivity only, CoI studies provide only a very partial picture of the true macroeconomic impact of disease, and fail to consider the contribution of depleted capital accumulation, investment in human capital and demographic change to diminished economic growth.

- Use of a 'human capital' approach to valuing production losses due to illness, disability or premature death (which involves multiplication of the total period of absence by the wage rate of the absent worker) is not realistic in settings where a pool of underemployed or unemployed labour exists, perhaps seasonally. Restricting the estimation of these potential production losses to the 'friction period' that it take to find a replacement worker provides a level of correction in the short term (Koopmanschap et al., 1995), but this still omits from consideration longer-term effects such as changes in the supply of labour.

In short, traditional cost-of-illness studies employ a static, partial and inconsistent approach to estimating the macroeconomic impact of disease and injury at the societal level. A more general and dynamic assessment of the present value of forgone consumption opportunities is required. Accordingly, the next section focuses on the mechanisms through which health shocks might affect GDP, and in Section 3.2 we turn to the two analytical approaches that have been used to identify these impacts on GDP - cross country regressions and simulation-based approaches, including calibration and computable general equilibrium models. Section 3.3 moves on to consider possible approaches to measure the overall economic welfare losses associated with

health shocks. Section 3.4 presents the key points related to measurement and valuation at the of the impact of disease and injury at the macroeconomic level.

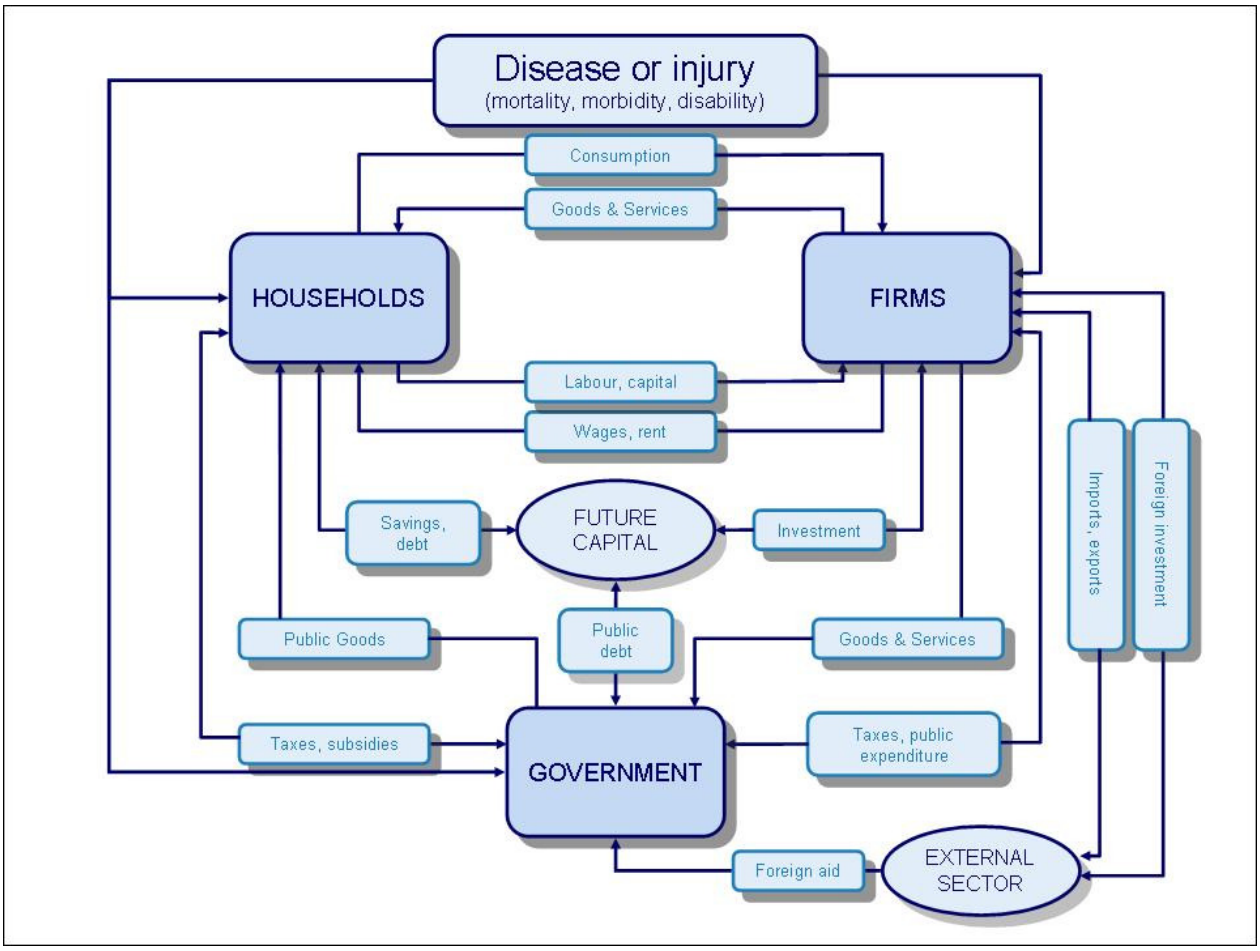
3.1. The impact of health shocks on national economies

We illustrate how the microeconomic agents fit together in **Figure 2**, using a 'flow of income' framework to depict the main channels of transmission of the economic impact of disease on households, firms and the government. Households provide labour and capital for firms, which employ these inputs to produce goods and services. In return, firms pay wages and rents to households for their productive inputs, and households use this income to consume the outputs of production and to invest to finance future consumption. We have simplified the system somewhat because households also provide labour to the government for the production of public goods, but this figure illustrates the basic framework of the circular flow of income, showing how aggregate income depends on the interaction between different agents.

In addition to providing inputs to production, and consuming goods and services with the income earned from these inputs, the framework depicted in **Figure 2** illustrates other economic functions or activities performed by households. One of these functions is as a financing agent for government activities via the payment of direct and indirect taxes; a further function relates to savings and investment, namely the transfer of disposable income (i.e. net of consumption and taxation), which is invested in the firms' productive activities and the formation (or diminution) of future capital stock. Disease and injury may challenge the household economic capacity to such an extent that it is forced to resort to loans and debt. Governments might also run surpluses and deficits; firms might be net borrowers or net lenders. These effects are shown in the diagram, forming the "dynamic" part of the framework - capital stock at the end of the period is an important determinant of output (and consumption) in future periods. For its part, the government generates revenue through the collection of direct and indirect taxes, loans and any foreign aid made available by the external sector. The government uses this income to purchase goods and services from the firms for the production of public goods. These goods are either provided or sold to households and firms. Closing the circular flow of income is the external

sector, which in addition to foreign aid also makes direct transactions with the private sector (exports, imports) and makes investments in production (foreign direct investment).

Figure 2 Conceptual framework for identifying the macroeconomic impact of disease



Key drivers of the macroeconomic impact of disease and injury can be understood in terms of injections into and leakages out of the economic system. In equilibrium, the aggregate income derived by all economic agents should be equal to aggregate expenditure; if total output is to remain constant, all leakages and injections in the system should even out. So if injections are in fact greater (or conversely, smaller) than leakages, total output increases (or decreases). For example, one implication of reduced labour productivity due to disease and injury is that the household will reduce the effective labour supply and consequently will earn less income. If there is no counterbalancing injection in the system (such as increased private debt, the liquidation of

private assets or an increased supply of public goods), this initial leakage may well have a multiplicative or knock-on effect and potentially lead to further reductions in GDP. The reduction in labour supply also exerts some influence on the operating activities of the firm, even if the firm is initially able to maintain its current levels of production by hiring additional workers to replace those that got sick. This is because over time the firm can be expected to experience a reduction in aggregate demand or market turnover due to lower disposable incomes. Moreover, feedback effects can lead to ever declining levels of economic growth, unless counterbalancing injections, such as an increase in foreign direct investment, can contribute to break the cycle.

In addition to the several channels discussed below, there are other ways through which disease and injury can impact the economy. This includes, for instance, the wider economic impact of disease and injury resulting from 'feedback loops' between epidemiological and economic factors. For instance, the outbreak of rapidly spreading infectious diseases and the public perception of the risk of becoming infected can lead to dramatic changes in patterns of social and economic interaction. During the outbreak of severe acute respiratory syndrome (SARS) in China in 2003, this has been shown to reduce the level of social contact, which in turn reduced the volume of economic transactions and by extension affected the economy (Smith, 2006). On the other hand, less social interaction also contributed to slow down the disease dispersion and therefore to control the negative economic impacts of the outbreak. Where relevant, these feedback loops have to be taken into account by models of the economic impact of disease, possibly via a general equilibrium framework which is able to control for their simultaneous determination.

3.1.1. The impact of health expenditures

Disease and injury impose economic consequences in the form of increased health expenditures for almost all groups of economic agents. As discussed in Section 4, this is the case, for instance, when there are increased health care expenditures resulting from disease and injury at the household level, when firms make investment in health and safety to protect the wellbeing of their workers, or when governments are forced to provide care to prevent or treat disease. When aggregated at the macroeconomic level, these different categories of health expenditures define the overall amount that the society pays for medical and other expenses as result of illness. This

amount represent an opportunity cost in terms of reduced availability of resources for consumption of non-health related goods and services at the current period. In addition, increased health expenditures as result of disease and injury also imply an opportunity cost potentially associated with lower economic growth, which can be seen in a number of ways.

In most countries private consumption represents by far the largest component of aggregate demand. Moreover, the macroeconomic effect of fiscal policy on economic growth depends critically on the properties of the consumption function (Agénor and Montiel, 1996). Therefore any factor that affects aggregate consumption can also be expected to have major implications for the overall economic performance. The onset of disease, especially in developing countries with limited coverage of social protection systems, often forces households to adapt patterns of consumption in order to cope with the additional burden of health expenditures. The increase in the consumption of health-related services and goods frequently leads to a corresponding reduction in the resources available for savings and investment.

In many developing countries the average consumption of non-health related goods and services is close to the subsistence level, so households are often forced to decrease domestic savings, to sell assets or to increase commercial or informal borrowings. The aggregation of these separate transfers frequently implies that increased health expenditures exert an overall negative impact on savings at the societal level. Several studies have assessed this negative impact on savings and have shown that it is likely to be considerable, especially in the case of epidemics such as HIV/AIDS (Cuddington, 1993; Cuddington and Hancock, 1994; Arndt and Lewis, 2000; Haacker, 2002). On the other hand, lower aggregate savings are likely to lead to increases in interest rates and the opportunity cost of investment, which can then impact negatively on the formation of capital and ultimately on economic growth.

A number of studies that have looked in particular at the demand-side of the economy have suggested that, by stimulating aggregate demand, increased health expenditures can have a positive influence on economic growth (Quattek and Fourie, 2000). In line with our perspective of considering the impact on potential flows of non-health related consumption, however, any increase in health-related GDP should generally not be considered as an increase in economic

welfare (see also Greener, 2002). Although an increase in health-related GDP represents an increase in economic activity with potential further multiplicative effects, the immediate effect is to reduce the amount of resources that can be directed into production and consumption of utility-generating non-health GDP.

Increased health expenditures are also foreseen to impact on government's resources, by reducing the amount of household taxable income and increasing the amount of expenditures that have to be covered out of public funds. The resulting impact on public budgets limits the government ability to spend resources in other areas, particularly in social expenditures such as education and in public investments in infrastructure projects, sanitation, science and technology, and other strategic areas. Public investment is a major source of economic dynamism and economic growth in many developing nations (Agénor and Moreno-Dodson, 2006). Consequently, reductions in public investment resulting from increased health expenditures can seriously compromise long-run economic growth potential. On the other hand, tighter overall budgets can compel governments to increase the level of taxation in order to meet increased additional health expenditures. This in turn can have macroeconomic repercussions by depressing aggregate demand and therefore limiting the growth potential of the economy.

3.1.2. Labour and productivity losses

Health status can impact labour productivity and economic growth via a number of different channels and, as suggested by Weil (2007), it is useful to distinguish between direct, indirect and demographic effects.

Direct impacts

Several studies suggest that improved health conditions exert a direct impact on economic growth by increasing worker productivity (Bloom et al., 2004; Alsan et al., 2006; Weil, 2007; Bloom and Canning, 2008). A healthier workforce enjoys greater physical and mental ability and therefore is able to produce more with the same amount of inputs. The aggregate impact on economic growth of increased labour productivity can be quite substantial. For instance, in an analysis of the

economic impact of historical improvements in nutrition in England and Wales, Fogel (1994) shows that around one third of the growth of the British economy over the period 1780-1980 can be attributed to better nutrition and greater efficiency in transforming nutritional input into economic value. Some authors have argued that this effect is especially important in the context of developing countries, since a high proportion of labour force in these countries is engaged in manual work and productivity tends to be more reliant on the sheer physical capacity to work (Strauss and Thomas, 1998; Bhargava et al., 2001; Bloom et al., 2003). In these cases, health improvements that increase the number of healthy workers for a fixed stock of land and capital in the economy might lead to higher unemployment or lower wages.

Indirect impacts

Weil (2007) identifies a number of indirect channels through which health can affect economic development and increase labour productivity. First, improvements in overall health conditions can contribute to lower absenteeism and worker turnover, and to increase cognitive functioning. In the same way, these factors might have important implications for production costs. Absenteeism and increased worker turnover often force firms to hire additional workers, who then need to receive additional training and require time to adapt to their new functions. On the other hand, improvements in cognitive abilities can contribute to improve the quality of education for a given level of schooling, which in turn can be expected to improve labour productivity.

In addition, productivity losses and lower working capacity resulting from disease and injury often lead to reduced household earnings. This can have important aggregate implications, even though some countries have counterbalancing mechanisms in place, such as systems of disability insurance, that can help to mitigate the effects for the household. From a macroeconomic perspective, however, the resulting lower economic efficiency implies that the overall level of production for a given amount of inputs is reduced. Lower productivity and income generation can exert knock-on effects on aggregate demand and supply, which can hinder the pace of economic growth. For example, reduced spending power at the household level might discourage firms from investing in productive capacity and contracting workers, further intensifying the reduction in aggregate demand.

The second indirect channel of impact of better health on economic growth identified by Weil (2007) is via increased incentives to delay retirement and to save for retirement. Some diseases such as HIV/AIDS affect mostly working age population groups and can contribute to reduce labour market participation and force people into early retirement, reducing the average retirement age. This in turn can have important impacts on the dependency ratio in the economy and, by extension, on potential economic growth. Moreover, the additional numbers of individuals that are forced into early retirement often require special medical care and specialized services. This can contribute to reduce the amount of resources that can be directed into productive activities and can also exert cost pressures on businesses, via increased health care costs and insurance premiums. On the other hand, the expectation of longer retirement periods resulting from mortality improvements often leads to higher savings during active life, which in turn can contribute to increase the stock of physical capital per worker. As discussed below, the third channel through which health improvements affect economic growth is via improved incentives to investing in education due to longer payoff periods in the labour market.

Demographic impacts

Changes in the demographic composition of a nation can have significant impacts on economic performance. For instance, changes resulting from mortality variations that increase the dependency ratio can impact on economic growth. Reductions in the size of the working age population due to sexually transmitted pandemic diseases such as HIV/AIDS are a case in point. Since the effects of changes in the demographic composition are intrinsically dynamic, however, it is possible that negative impacts in the present can be transformed into positive shocks in the future and vice-versa.

Bloom and Canning (2000, 2008) describe the concept of "demographic dividend" as countries undergo a process of demographic transition from high to low rates of mortality and fertility (Lee 2003). Improvements in health and sanitation conditions cause significant declines in rates of child mortality and consequent demographic explosion with exponential rates of population growth. Over time, as parents adapt to the new environment with low child mortality, the fertility

rates also tend to decline, with the effect that we observe a baby boom cohort that is much larger than both the preceding and the following birth cohorts. Bloom and Williamson (1998) examine the significance of this effect in explaining the rates of economic growth in East Asia during the period 1965-1990. The results suggest that population dynamics account for 1.4%-1.9% of annual GDP per capita growth in the region during this period (approximately half of the overall growth observed). Conversely, Bloom and Malaney (1998) estimated that reduced population size caused by the mortality crisis seen in Russia during the 1990s contributed significantly to reduced economic growth over this period.

Some diseases such as HIV/AIDS affect mainly the working age population and can exert significant pressure on the demographic structure. High mortality rates among working age populations can lead to increasing dependency ratios, and lower population and labour force growth, with quite substantial economic impact in terms of loss of market output. There are also possible implications for reduced investments in human capital formation, since the young might be forced to engage earlier in market productive activities and also in non-market production, for instance acting as informal carers (see Greener 2002).

Impact on GDP per capita and overall GDP

Recent studies have argued that faster population growth resulting from improvements in health and life expectancy can also have significant implications for the impact of health on income per capita as opposed to the impact on total or overall GDP. Acemoglu and Johnson (2006), focusing on exogenous health improvements during the 20th century, argue that extended life expectancy has increased population growth, which in turn has contributed to partly offset income growth. Therefore, the impact of health improvements on economic growth might turn out to be lower than has been suggested by other studies. Similarly, the report by the United Nations Department of Economics and Social Affairs / Population Division (2004), focusing on the impact of HIV / AIDS, suggests that "in a typical neoclassical growth model, AIDS affects total output directly, by decreasing the number and efficiency of workers, and also indirectly, by decreasing savings and investment. Since HIV/AIDS also results in a lower population than would otherwise have existed, the effect on GDP per capita is smaller than the effect on total output; at least in

principle, there could be situations in which the net effect on GDP per capita would be nil or even positive". Over (1992) also addresses this question in relation to the impact of HIV/AIDS in sub-Saharan Africa by noting that "if the only effect of the AIDS epidemic were to reduce the population growth rate, it would increase the growth rate of per capita income in any plausible economic model". However, the results of his analysis suggest that specific characteristics of the AIDS epidemic (namely, that it reduces savings and affects both skilled and non-skilled labour) are sufficient to reverse this trend and actually reduce per capita income in a 35-year projection of 30 countries.

The evidence provided by this literature strand has challenged the dominant view that health improvements are conducive to economic growth, both in absolute and per capita terms. This emphasizes the need to take into account the implications of a faster growing population for the process of economic development and points out the complexities involved in establishing causal links between improvements in health and economic outcomes. Overall, no consensus has been established as to whether health has either a positive or negative impact on incomes. In particular, it is very difficult to determine that sizable income changes can be attributed to relatively small changes in health in a country. More research is needed and the studies on this area should make sure that the conclusions are robust to the hypothesis that accelerated population growth might counterbalance the positive impact of health improvements on income growth.

3.1.3. Effects on human, physical and financial capital formation

In addition to the direct effects through increased productivity described above, there are also a number of other mechanisms by which health impacts on economic growth. Some of these are to a great extent related to changes in the incentives to human capital formation, savings and returns to investments.

Human capital formation

Bloom and Canning (2008) suggest that there are two mechanisms through which health affects the formation of human capital and education. First, better health can increase school attendance

among children and improve cognitive ability and learning capacity. For instance, Miguel and Kremer (2004) show that the treatment of intestinal worms in children increases schooling attendance and educational achievement. The second effect is that longer life spans and reduced mortality increase the incentives to acquiring valuable skills due to longer payoff periods. In other words, better health can increase the returns to investments in education and worker experience, because longer life spans increase the payoff period for investment in human capital productivity.

Recent studies have considered the impact of disease and illness on health capital investments, focusing specifically on the endogenous determination of mortality and life expectancy (Bloom, Canning and Graham, 2003; Hazan and Zoabi, 2006). Environments with low life expectancy discourage individuals from investing in health and education due to relatively low payoff periods and high uncertainty (Kalemli-Ozcan et al., 2000; Chakraborty, 2004; Finlay, 2006). This can lead to feedback effects that reinforce poverty-mortality traps whereby high mortality discourages investments in human capital, slowing down economic development and further intensifying the processes that determined high mortality in the first place. A theoretical model developed by Castelló-Climent and Doménech (2008) examines the implications of this mechanism, showing that under certain circumstances a low steady state poverty trap emerges where children from poor households receive low levels of human capital investment and consequently end up with low life expectancy and working in non-skilled jobs.

A related effect is the impact of child health on actual investments in human capital and the trade-off between quality and quantity of offspring. Lower child mortality frequently leads to reductions in fertility because it reduces the need to have more children in order to guarantee economic capacity in the future. In addition, having fewer children encourages parents to increase the investment in human capital formation and the transmission of skills and knowledge for each child. In an analysis of 44 African countries in the period 1985-2000, Kalemli-Ozcan (2006) evaluates the impact of HIV/AIDS aggregate prevalence rates on fertility and school attendance. The results suggest that the HIV/AIDS epidemics has had very significant effects on these decisions, encouraging fertility but reducing investment in human capital accumulation. For

instance, compared to a country with low prevalence of HIV/AIDS but otherwise similar, Congo has 2 more children per woman but 38 percentage points lower rates of school attendance.

Another factor contributing to higher return to human capital investment is that low absenteeism and longer life expectancy increase job experience and can contribute to delayed retirement. Greater returns to education and longer pay-off periods make people more willing to invest in education and the development of on-the-job skills, potentially conducing to greater economic growth. Similarly, longer life expectancy increases prospects of retirement and can induce individuals to save more in order to guarantee retirement income. From the macroeconomic point-of-view, higher volumes of aggregate savings can contribute to reduced interest rates in the economy. This can be a powerful instrument to increase aggregate investment and long run economic growth.

Physical and financial capital formation

Moving to physical and financial capital, the amount of resources invested in an economy over time is one of the main determinants of long term economic growth. Increasing the stock of capital can contribute to raise efficiency in production and has also been shown to produce externalities effects that are conducive to growth. Disease and injury change the economic incentives conditioning the choices of microeconomic agents that can affect aggregate levels of savings and investments in physical capital in a number of different ways.

First, as discussed above, by extending the relative length of retirement and the planning horizon, higher longevity and life expectancy exert lifecycle impacts on individual behavior and encourage people to increase savings for retirement. Increasing aggregate levels of domestic savings can in turn lead to higher investment in capital formation and faster economic growth. Improvements in health conditions have also been suggested to affect the inter-temporal discount rate of individuals and households, which can have an effect on saving and investment decisions. Reinhart (1999) and Sanso and Aísa (2006) argue that longer life expectancy affects the household's degree of patience and willingness to smooth consumption over time, increasing savings and therefore exerting a positive impact on economic growth.

On the other hand, as suggested by Bloom et al (2003), higher savings during young ages can be offset by increased payouts during retirement, thereby rendering the net effect indeterminate. However, during the adjustment phase the gains in aggregate savings can be quite substantial. In other words, the boost to savings rates can be temporary and counter-balanced in equilibrium by a higher old-age dependency rate.

Zhang and Zhang (2005) estimate the effect of initial life expectancy on the ratio of investment to GDP for 76 countries in the period 1960-1989, showing that a 10% increase in initial life expectancy can raise the investment ratio by 2.4% and lead to nearly 1% increase in the growth rate (in a sensitivity analysis, however, the effect of life expectancy on the savings rate is inconclusive). This supports the suggestion made by Bloom et al (2003) about an off-setting effect of higher retirement expenditures. In addition, the effects of longevity on savings estimated by Zhang and Zhang (2005) are decreasing on life expectancy - that is, they become weaker as life expectancy increases. This is in line with the theoretical results presented by Aísa and Pueyo (2004), which suggest that health and growth compete for resources as life expectancy increases.

Another way that health could explain increasing investment is by attracting additional foreign direct investments to a country. Studies have suggested that the level of human capital in the host country can affect the distribution of foreign direct investment flows. In many cases the success of investments in physical capital depends on the existence of a complementary skilled labour force that is able to efficiently employ the equipment and can deal with the complexity associated with modern economic transactions. In this sense, a healthy labour force is an important factor to guarantee that capital investments can be recovered. Human capital in the form of workers' health contributes to increase worker productivity, for instance by lowering absenteeism and raising on-the-job experience; this can help to attract foreign direct investment flows. One implication suggested by Noorbakhsh and Paloni (2001) is that countries relying on low-skilled labour and natural resources to attract foreign investment may find it difficult to attract investments into high added-value sectors, which could boost economic growth.

Alsan et al. (2006) evaluate the impact of health on foreign direct investment inflows to low- and middle-income countries, highlighting that some of the same factors through which health affects the operations of firms in general also influence decisions to invest in countries with high burden of disease morbidity and mortality. Alsan et al. (2006) also argue that the endemic prevalence of communicable diseases may deter foreign investors from investing for fear of their own health and of expatriate staff. The main result of their analysis is that the inflow of foreign direct investment is strongly and positively related to measures of population health, and that each additional year of life expectancy yields 9% increase in foreign direct investment.

In summary, current evidence about the effect of health on savings and investments is still ambiguous. Although there are some indications that higher levels of health may contribute to attract inflows of foreign direct investment, the incentives effect on higher domestic savings are ambiguous and at best temporary. Nevertheless, as discussed above, a direct effect of health on savings exists if disease and injury force households to draw on existing and future savings in order to finance increased health expenditures.

3.2. Measuring the impact of disease on national income or product

To accurately assess the macroeconomic impact of disease or injury on national income, a dynamic multi-period model which can take into account the inter-temporal linkages of all of the mechanisms described above is required. Our review of recent studies assessing the macroeconomic impact of disease has identified that models can be generally classified into two categories:

- Regression-based econometric growth models using either single-country or cross-country panel data;

- Macroeconomic calibration and computable general equilibrium simulation models;

Below we provide an overview of the main analytical approaches that are used to assess the macroeconomic impact of disease and injury, focusing on the main advantages and drawbacks of each approach.

3.2.1. Regression-based estimation models

Most economic growth studies to date have used regression analysis in the style of Barro (1991), who empirically tested various contemporary growth models and by so doing established a standard set of variables that should be included in growth regressions. Special emphasis is given to the role of human capital in the growth process; for 98 countries between 1960-1985, Barro finds that the growth rate of real GDP per capita is positively related to initial human capital, and negatively related to the initial level of real per capita GDP. Human capital is in turn negatively correlated with fertility and positively with the ratio of physical investment to GDP. Other variables relevant for explaining a country's growth performance are the share of government consumption in GDP and market distortions (inverse relationships) plus measures of political stability (positive relationship). Some studies have employed more or less developed versions of a production function (Bloom, Canning and Sevilla, 2004), with the more sophisticated studies defining a system of equations to take into account the endogeneity of health, i.e. the feedback mechanism between health and income (Bonnel, 2000). Much of the current literature has focused on the impact of overall health, as measured by life expectancy or adult survival rates. Disease-specific analyses have focused mainly on the negative impact of malaria on economic growth in sub-Saharan Africa (Gallup and Sachs, 2001; McCarthy, Holger and Wu, 2000).

This approach is largely based on the Mankiw-Romer-Weil empirical growth model (Mankiw et al., 1992), which extends the Solow model to include measures of human capital in addition to physical capital and labour as determinants of economic growth (Knowles and Owen, 1995; McDonald and Roberts, 2002). This framework assumes the existence of a production function that uses physical capital, human capital and labour as inputs in the production of aggregate output. The parameters of the production function are estimated using a cross-sectional panel dataset of countries with the aim of evaluating whether differences in health indicators have an impact in subsequent income growth.

Considering first studies that focus on the impact of changes in general mortality and life expectancy, Bloom et al (2004) find the coefficient on life-expectancy to be 0.01, implying that increasing life expectancy by one year raises output by about 1%. However, the effect is not well determined and the coefficient is not statistically significant. Once country-specific total factor productivity (TFP) is included (via fixed effects), they find that each extra year of life expectancy adds 4% in output. Lorentzen et al. (2005) also show that adult mortality is a robust and statistically significant predictor of economic growth. They carry out two types of estimation - instrumental variable (IV) as well as three-stage least squares (3SLS), whereby the latter is meant to separate out and capture the effects of three different hypothesized transmission mechanisms (physical capital investment, human capital investment and fertility rates). From the IV estimation they find that an increase in adult mortality of one standard deviation is associated with a 1.2% reduction in growth. The results from 3SLS estimation, however, indicate that a one standard-deviation increase in adult mortality leads to a 0.85% decrease in growth through the three pre-defined channels. This suggests that there are other channels through which mortality affects growth. It is worth noting that adding an adult-mortality variable to a standard growth regression makes the sub-Saharan Africa-dummy insignificant (which otherwise accounts for a 1% growth difference), suggesting that high mortality can explain a large part of Africa's low growth performance.

Chakraborty (2004) also finds a very strong positive relationship between health and income growth. When regressing only life expectancy on the 1990 GDP level per worker, his results suggest that longevity explains about 81% of the cross-country variation in income levels. In his growth regression, the coefficient on life expectancy in 1970 is around 6 times higher than that on school enrolment rates, pointing to a much bigger role for health than education. Sala-I-Martin et al. (2004) employ a Bayesian framework to assess the relative contribution of several variables in explaining economic growth. The authors find 11 variables that are robustly correlated with growth and health (as proxied by life expectancy in 1960) is found to be the 5th most significant.

Another way to gain insight into the economic impact of disease and injury is to apply a cross-country regression framework to specific diseases (as opposed to ill-health generally). The

evidence about the impact of specific diseases on economic growth is very mixed. As described below, while there is a clear evidence base to suggest that malaria can have an impact on economic growth, the same is not always the case with respect to other diseases.

Turning to studies estimating the effect on income of malaria, Gallup and Sachs' (2001) find that countries with intensive malaria grew 1.3% less per person per year. They calculate that a 10% reduction in malaria was associated with 0.3% higher growth. Growth between 1965-1990 for countries with malaria has been 0.4% per annum compared to average growth for other countries of 2.3%. Further, more than one third of countries with severe malaria (11/29) had negative growth between 1965 and 1990.

McCarthy, Wolf and Yu's (2000) growth regressions suggest a significant negative impact of malaria morbidity on GDP per capita growth. This result is robust to extensive sensitivity analysis, which includes the estimation of both a pooled specification and seemingly unrelated regressions and the inclusion of generic dummies for Africa and for sub-Saharan Africa to pick up country-idiosyncrasies. They further estimate the hypothetical growth effect of eliminating malaria morbidity by country and year, which exceeds 0.25% for about 25% of the sample and peaks at 3.22% for Malawi. Artadi and Sala-I-Martin (2003) predict that had Africa had no Malaria over the last four decades, its annual growth rate would have been 1.25% higher than it actually was. Their estimate of foregone growth from lower life expectancy is 2.07% per year. Gallup et al. (1998) put the reduction in growth rates due to malaria at 1.3%, while Shepard et al (1991) find a growth reduction of 0.6%.

The evidence from cross-country regressions about the impact of other health conditions is more limited. For instance, the study by Gyimah-Brempong and Wilson (2004) reviewed below reports significant effects of child mortality rates on economic growth. However, the literature review by Wilhelmson and Gerdtham (2006) find no convincing evidence to suggest that maternal and newborn health affect economic growth. Very few studies have been carried out in this area. Wilhelmson and Gerdtham's (2006) review did not find any study that directly reflects the impact on economic growth of the health status of children and young mothers. The proxies that have been used for maternal and neo-born are very crude and likely to correlated with other health

indicators, including for example infant mortality (which reflects the survival of children but not of young mothers) and life expectancy (which covers both child and mother health, but also includes other groups such as adult males). The wide variation with respect to the methods and variables that are used imply a large degree of uncertainty around the estimates.

Econometric and statistical issues

A number of issues limit the applicability and policy relevance of cross-country growth regression models. One of the main limitations is that most studies focus only on the impact of changes in mortality to estimate the impact of health on economic growth. Life expectancy is only a partial measure of the underlying population health status. In particular, morbidity indicators related to the prevalence and incidence of several diseases are associated with health status and can impact economic performance.

Another major limitation of cross-country growth regressions is the heavy reliance on a 'black-box' model, largely ignoring some of the complex relationships that exist between disease and economic development. In particular, it is very important that economic growth regressions correctly address the issue of endogeneity, which as discussed in **Box 2**, is very likely to be a concern in the relationship between health and income. The most common strategy to deal with endogeneity is the use of instrumental variables and/or the estimation of systems of equations that allows for the simultaneous determination of endogenous variables, including health, education and economic growth. Bloom et al. (2004) use lagged values of input and output as instruments to identify the effect of life expectancy on current income growth. Bonnel (2000) control for endogeneity issues by estimating a system of three equations for economic growth, institutional variables and the determinants of the HIV epidemic. In this model HIV/AIDS is assumed to reduce economic growth. In its turn, economic growth can either slow down the epidemic spread (by providing employment for women, increasing education and improving physical infrastructure) or facilitate it (in the sense that large investment projects frequently rely on labour migration, which can contribute to the spread of the disease). The net effect on GDP per capita is indeterminate ex ante because lower labour force growth would increase income per capita

(assuming constant GDP), while lower investment and savings would reduce stock capital and income per capita.

Box 2 Endogeneity in the relationship between health and economic growth

Endogeneity or reverse causality is defined as the simultaneous determination of variables within an economic system. Endogeneity is likely to be a very important issue in cross-countries growth models that aim to estimate the economic impact of disease and injury in economic growth. While better health will lead to higher income through various specified channels, it is also generally the case that higher income will lead to better health, as more money is available to prevent and cure diseases and people have access to better quality health goods and services. Moreover, it is also possible that unobservable factors affect both health and income simultaneously, causing further difficulties for the identification of the effect of health on economic welfare.

Econometrically, ignoring endogeneity is problematic as it can lead to biased coefficient estimates. In the macroeconomic literature several approaches can be used to properly identify the causal effect of health on economic growth and to control for the endogeneity between health and income (Jack and Lewis, 2008), including the use of instrumental variables, structural estimation and exploring the impact of exogenous health shocks.

Other studies rely on exogenous variations in health indicators in order to identify the impact of health on economic growth. For instance, Acemoglu and Johnson (2006) exploit the exogenous improvements in population health resulting from the development of vaccines and new drugs during the 20th century international epidemiological transition. Exogenous health improvements have increased rates of economic growth. On the other hand, fertility rates have not decreased as much (although still contributing to extend life expectancy). The authors estimate that 1% increase in life expectancy leads to 1.5-2% increase in population, but a much more modest impact on total GDP. Consequently, the increased pace of population growth has contributed to offset the faster growth in income. There is no evidence that the considerable life expectancy increases have led to gains in GDP per capita. Acemoglu and Johnson's (2006) results suggest that the impact of life expectancy on economic growth might be lower compared to other studies. However, this study covers a very specific context and consequently the results cannot be used to draw general conclusions about the impact of health on economic growth. First, there is the criticism of lag times (Bloom and Canning, 2008), meaning that it takes time until child health

improvements are translated into increased adult worker productivity. This is reinforced by the fact that most of the technological improvements taking place over the period of analysis affect mainly children. Moreover, the analysis does not cover the rise of the HIV/AIDS epidemics, which has heavily reduced life expectancy and the working-age population in several countries. Another limitation relates to the assumption of a fixed capital stock over the long run. Although this is reasonable for the amount of land, it is certainly not the case for other types of capital. Finally, it is important to recognize that even if it is the case that better health does not increase income, there are still significant impacts in terms of economic welfare improvement, as discussed below.

In addition, some studies have suggested the existence of diminishing returns to health improvements, in the sense that there is a threshold beyond which health improvements cease to have positive effects on economic growth. Barghava et al (2001) estimate the impact of adult survival rates (ASR) in GDP growth in a sample of both developed and developing countries. The analysis estimates a sizable and positive impact of ASR increases on the percentage growth rate in low income countries, but a negligible or even negative impact in middle and high income countries. In addition to the interpretation of diminishing returns to health improvements mentioned above, Barghava et al (2001) suggest that in developing countries ASR are determined by levels of nutrition, smoking prevalence rates and infectious diseases. In high income countries, by contrast, other aspects such as genetic factors, access to care and the cost of services have greater importance.

In a comparison between Sub-Saharan Africa and OECD countries, Gyimah-Brempong and Wilson (2004) report that the stock of health human capital (as measured by the inverse of the child mortality rate) explains between 22%-30% of the per capita GDP growth rate. Although this impact is positive and significant in both regions, the results suggest a non-linear relationship between health and economic growth whereby increases in health increases the growth of per capita income but the marginal effect eventually diminishes. These results highlight the importance of choosing appropriate definitions of health indicators and taking into account the uncertainty surrounding theory, measurement and specification in cross-country growth regressions (Aghion and Durlauf, 2008).

3.2.2. Simulation approaches: Calibration and CGE models

Economic growth regressions are an important method to estimate the overall macroeconomic impact of diseases and injury. However, in some situations it is also important to consider other possible approaches. Below we consider simulation models based on calibration and computable general equilibrium (CGE) models as additional methods that can be used to evaluate the macroeconomic impact of disease.

Calibration models

The calibration or growth accounting approach starts off from microeconomic estimates of the exogenous effect of health on income at the individual level (Bloom and Canning, 2008). This stable relationship is then combined within a macroeconomic model with country-level health indicators to obtain an estimate of the direct impact of health on income differences across countries. The results can then be further used to decompose the differences in GDP across countries into input factors as implied by an aggregate production function. Using this approach, Weil (2007) estimates that male ASR (adult survival rate) is an important determinant of GDP gaps between countries. Eliminating ASR differences across countries would reduce the variance of GDP per worker and significantly reduce the level of inequality in GDP per worker between the richest and poorest countries. However, the author notes that these effects are substantially smaller compared to estimates derived from cross-country regressions. Shastry and Weil (2003) carry out a similar analysis to estimate separately the effect of the female prevalence of anemia and ASR on GDP variance. Decomposing GDP variance into its contributory factors shows that differences in anemia prevalence account for 1.3% of the log GDP variance, while ASR account for 19%. The latter result is very important, considering that physical capital contributes to 20% of log GDP variance, education human capital contributes to 21% and differences in productivity account for 40%.

Others studies use a similar calibration approach to estimate the macroeconomic impact by simulating the effect on GDP growth resulting from projected changes in economic inputs due to disease and injury. This approach is based on the specification of a structural growth equation

with a Solow-style production function that depends on the capital stock, labour force size and aggregate measures of human capital, such as education and health indicators. The production function uses parameters estimates from the literature. The use of different measures of population health, such as life expectancy, ASR or standardized mortality rates due to specific diseases allows the estimation of the effects of changes in these variables on the amount of effective labour in the economy and, by extension, on economic growth. This approach is very flexible and suitable for applications with limited data availability. Moreover, since it is based on structural production functions it avoids many of the issues that affect other empirical models of economic growth. The main limitation, however, is that this approach more readily applies to changes in the effective labour force resulting from mortality, but is less sensitive to capturing morbidity effects.

Using this approach, Abegunde et al (2007) estimate the impact of chronic diseases - cardiovascular diseases, cancer, chronic respiratory diseases and diabetes - on economic growth in 23 countries that account for 80% of the total burden of chronic disease mortality in developing countries. This allows the estimation of the impact in the period 2006-2015 given current and projected rates of mortality in these countries and also to evaluate the impact of reducing chronic disease annual death rates by 2% per year in the period considered. Also using this approach, Cuddington (1993) estimates that HIV/AIDS reduces 2010 GDP in the United Republic of Tanzania by 15%-25% compared to a counterfactual no-AIDS scenario. Similarly, Cuddington and Hancock (1994) compare the impact of HIV/AIDS in Malawi in three scenarios: a no-AIDS counterfactual versus medium and extreme AIDS projections. The results of the analysis for the period 1985-2010 suggest that annual GDP growth would be 0.2%-0.3% lower in the medium scenario and 1.2%-1.5% lower in the extreme scenario. This implies that real GDP in 2010 was projected around 4%-5% lower compared to the no-AIDS scenario. However, a more recent analysis taking into account the full impact of HIV/AIDS on mortality would perhaps reveal a more sizeable impact.

The analysis of the economic impact of HIV/AIDS in Botswana by Jefferis et al. (2008) uses simulations to estimate net and dynamic impacts on different economic outcomes. Although the impact on GDP growth is likely to be negative, the impact *a priori* on other variables such as

GDP per capita, unemployment and public finances is indeterminate. The impact on GDP per capita will depend on whether the reduction in GDP outweighs the decline in population. Likewise, the overall impact on unemployment results of the interaction between lower growth of labour supply, which tends to increase wages and decrease unemployment, and lower labour demand resulting from lower investment and economic growth, which tends to have the opposite effect. The authors capture these dynamic impacts in a growth model that combines demographic and macroeconomic projections to explain GDP growth, average income, investment, employment, wages and government health expenditures. The simulations suggest that the epidemics reduces the yearly rate of real GDP growth between 1.5%-2% and GDP per capita growth between 0.5%-1%, the difference due to lower population growth. The employment situation also deteriorates, with higher unemployment and lower wage growth. The model also compares the impact of public provision of antiretroviral treatment (ART) on government budget and economic performance. In spite of the high cost of ART provision, over time the net effect on fiscal budget is relatively small because ART increases government revenue from economic growth and reduces the cost of other services, such as orphan care and inpatient costs.

Computable general equilibrium (CGE) models

In some other cases a more disaggregated approach might be needed. For instance, in addition to the overall macroeconomic impact, some studies might also aim to assess the economic impact across markets and sectors of the economy. Alternatively, it might be important to consider the distributional effects of policy interventions across different agents. In recent years, the use of CGE models has become increasingly popular to answer this type of questions.

Recent advancements in theory and analytical techniques and improvements in computational and data processing capability have made it possible the estimation of complex CGE models, which provide very complete and accurate characterizations of the economy, starting from disaggregated individual categories of agents and building up to the macroeconomic equilibrium. This kind of modeling approach provides a conceptually appropriate framework that can be used to estimate the current impact of specific diseases on the economy, to compare alternative case-scenarios, and to forecast the likely order of magnitude of the effect of policy interventions.

Working under a general equilibrium framework, CGE models are used to find simultaneous price-guided clearing of all sectors of the economy (Smith et al., 2005). The economy consists of representative agents, a sufficient set being consumers and producers. The more elaborate studies also take into account the heterogeneity within the sets of representative agents, for instance, by considering different types of consumer, such as young- and old-age households, and different sectors of the economy, such as health and non-health. The implementation of CGE models is based on social accounting matrices (or input-output tables), which map out the flows of resources across sectors and agents in the economy. The information required to construct such matrices can be very hard to collate, especially in the context of developing countries, and might require substantial expertise to adapt the data into the necessary format for analysis.

Consumers maximize their utility by trading-off time spent working against leisure as well as consumption against saving. Producers choose that combination of inputs that meets their target output and maximizes profits. Equilibrium is then characterized by those prices at which the level of production and consumption within each sector ensures that demand equals supply across all sectors and market. This framework explicitly accounts for all inter-linkages across different sets of sectors of the economy - households, firms, government and the external sector, where applicable. Equilibrium requires that an additional set of constraints is imposed on these sectors so that "expenditures do not exceed income and income, in turn, is determined by what factors of production earn"; Piermartini and Teh, 2005: p. 4). The parameters of the model are then chosen using the method of *calibration* which yields a set of values that support the benchmark equilibrium. Calibration is further supplemented by specifying values for exogenous elasticity parameters, which are typically obtained from other studies as they cannot be calibrated using values in the benchmark dataset.

Up to this point, the application of CGE models to evaluate the macroeconomic impact of disease are few in number. While a handful of CGE studies were carried out in the early 1990s (e.g. Cuddington, 1993; Kambou, Devarajan and Over, 1992), they mostly found insignificant effects of a health shock on the economy. More recent general equilibrium studies, however, predict dramatic economic effects, especially for HIV/AIDS (Arndt and Lewis 2001; Bell, Devarajan and

Gersbach, 2003). For example, Arndt and Lewis (2001) project GDP growth rates and levels for South Africa for 2010 and find a growth rate of 3.75% in a 'no AIDS' scenario versus a rate of 1.25% in the AIDS scenario. In terms of levels, they predict that GDP will be 17% below the 'no AIDS' level in 2010. Bell, Gersbach and Devarajan (2003) predict that if nothing is done to combat the epidemic in South Africa,, a complete economic collapse will occur within 3 generations. The simulations suggest that if nothing is done GDP per capita in 2080 will be only 14% of what could be expected in the absence of HIV/AIDS.

Smith et al (2005) use a CGE model to assess the macroeconomic impact of antimicrobial resistance (AMR) in the United Kingdom of Great Britain and Northern Ireland. AMR is introduced as a shock that impacts on labour supply, productivity and the cost of providing health care. However, since estimates of the partial impact of AMR on these factors were not available, the authors assumed a range of values for the parameters based on estimates from other countries: AMR is assumed to reduce labour supply between 0.1%-0.8%, to impose productivity losses in the range 0.5%-10% and to increase health care delivery costs by between 0.5%-10%. Working through the inter-linkages between the different sectoral impacts it becomes clear that AMR implies a reduction in overall societal welfare amounting to a real GDP fall between 0.4%-1.6%, which corresponding to around 6%-20% of the budget of the National Health Service in 1995.

In assessing the health impact on economic growth, some recent simulation studies have also evaluated the general equilibrium and demographic impacts on population growth. The simulation model proposed by Ashraf et al (2008) uses similar assumptions as Acemoglu and Johnson (2006) to conclude that the impact of health improvements on economic welfare are lower than suggested by other studies, because the impact on worker productivity is offset by an increase in population growth. Moreover, the productivity benefits of health improvements can require quite long periods before being felt in terms of GDP per capita. The analysis by Young (2005) on the impact of HIV/AIDS in South Africa assumes that the epidemic will slow down population growth by increasing adult and child mortality and reducing fertility rates. The results suggest that the reduced population growth will be large enough to offset the impact on reduced human capital formation and in the long run will contribute to a net increase in income per capita.

As the name implies, a main advantage of CGE models is that they focus on *general equilibrium* effects, as opposed to other models that are based on *partial equilibrium* analysis. The main difference is that partial equilibrium models focus on only one sector of the economy, with the assumption that *ceteris paribus* any effects or interactions between the sector under consideration and the rest of the economy are either very small or inexistent. For instance, reducing the price of a given good or service increases the available effective income and may lead to increased demand for other goods and services; a partial equilibrium analysis ignores such effect and assumes that the initial price change has no effect on the demand for other goods service, which is assumed to remain constant. Similarly, in general, increasing production in one sector can be expected to divert resources away from other sectors; however, a partial equilibrium analysis assumes that such transmission effects are negligible. In addition, public health emergencies of international concern are a case in point (see Beutels et al 2008). The immense scale and scope of such emergencies can lead to behavioral changes and capacity problems which are likely to violate the implicit assumptions behind partial equilibrium analyses.

In many cases, partial equilibrium can be considered a reasonable working assumption, for instance, if the focus of interest lies on the determination of the short run impact of disease on a specific sector. In other cases, however, it is important to consider general equilibrium impacts such as in analyses involving more than one sector, assessing the relevance of income effects and to gauge the impact on substitutability and complementarity of products and factors of production across sectors (Piermartini and Teh, 2005). By properly identifying the linkages between markets, the general equilibrium framework might be used to model ripple effects observed when changes in one market affect the equilibrium in other markets which, in turn, might have implications also for the original market. Contrary to partial equilibrium models, which normally concentrate on the economic impact only to the health care sector, CGE models have a broader focus on the social costs of disease and injury that includes the impact on other sectors of the economy as well (Smith et al., 2005).

A crucial step in the application of the CGE framework is the calibration of the main parameters and elasticities of the model. This often relies on estimates provided by microeconomic studies, which specify, for instance, what is the immediate impact of disease and injury on the labour

supply provided by the households. The next sections of these guidelines discuss in depth the issues related to the microeconomic estimation of the economic impact of disease and injury. For the moment, it suffices to say that CGE models are very dependent on the choice of parameters. Therefore, it is very important that empirical applications check the robustness of such assumptions by means of extensive sensitivity analysis of the results. This might include, for instance, running replication checks and re-estimating the results using different sets of parameters and with different assumptions about the feedback effects between sectors. Moreover, CGE models are frequently reported to suffer from major computational difficulties that might require considerable effort in order to obtain stable results.

3.3. Measuring the impact of disease on economic welfare

The analytical approaches considered in the previous section focus on market-based components of the macroeconomic impact of disease and injury, and therefore provide only a partial estimate of the full impact of disease or injury on economic welfare. To illustrate, Nordhaus (2005, p. 374) argues that, according to conventional measures of national income, "an economy in which people have a per capita income of $20,000 with lives that are nasty, brutish, and short would be ranked as equivalent to one with the same per capita income and lives that are healthy, civilized, and long (…) the key point is that the same *annual* income with a long and healthy life has a higher living standard than that income with a short and diseased life". Health-related improvements are one of the main activity areas neglected if the consequences of disease are measured solely in terms of GDP.

In order to place an economic value on health or indeed any other desired outcome, welfare economists posit that a change in welfare can be measured by observing how much a person is prepared to give up other consumption opportunities in order to get it, which amounts to the amount they are willing to pay. The use of willingness to pay (WTP) as an approach to the economic valuation of goods and services is attractive to the extent that it provides an appropriate underlying quantity of interest by which all possible impacts of disease can be added together. While the human capital approach focuses on impacts on market consumption and labour

productivity, the WTP estimates a more comprehensive measure of the benefits of living which includes in addition the value of non-market production and consumption, non-labour income, leisure time and any premiums attached to the avoidance of pain and suffering (see Landefeld and Seskin, 1982). As a valuation technique, WTP has been applied widely in the economic evaluation of transport, agriculture and environment, and has also been increasingly used in health (Olsen and Smith, 2001).

Willingness to pay for market transactions is straightforward enough since the prices agreed upon for these transactions are observable, whereas for non-market goods and services more indirect measures are needed (either via revealed choices in the market place or via stated preference). Stated preference techniques such as contingent valuation surveys ask people the maximum amount they would hypothetically be willing to pay for a particular outcome or commodity (or willing to accept in compensation for not having it), such as the amount they would be prepared to spend in order to avoid a life-threatening disease. By contrast, the revealed preference approach uses market observations to value the trade-offs that individuals are prepared to make in order to satisfy their preferences. This technique has been specifically used to place a value on human life by observing the trade-offs that individuals make between income and the risk of death (for example, an oil rig diver will accept an increased probability of death if he is sufficiently compensated with 'danger money'). The revealed amount of money needed to accept a specified increase in the chance of death is used to estimate a (market-based) estimate of the value of a statistical life. In a review and meta-analysis of the literature, Viscusi and Aldy (2003) derived a predicted value of statistical life of US$ 5-6 million in the US, equivalent to at least 100 times GDP per capita.

Such estimates have been used to convert lives that have been lost to disease or saved by an intervention into a monetary value (Nordhaus, 2002; Bloom, Canning and Jamison, 2004). Dating back to an initial application some 35 years ago (Usher, 1973), studies using a 'full-income' approach attach monetized values to the mortality or morbidity prevented by health programmes, and by so doing extend the analytical scope beyond market-based impacts alone towards the fuller notion of economic welfare. For example, Nordhaus (2002) placed economic values on changes in the probability of death in order to generate an expanded notion of national income

that explicitly incorporates changes in longevity into measures of overall economic welfare[2]. The results suggest that the health improvements during the 20th century may be worth as much as the totality of the measured increase in non-health goods and services. Murphy and Topel (2006) assess the economic value of mortality reductions in different periods in the United States. In line with Nordhaus (2005) the results reported by Murphy and Topel (2006) suggest that cumulative longevity gains during the 20th century are valued in 2000 at around US$1.3 million per person. The estimated value of projected future improvements in health, such as those resulting from lower cancer prevalence, are also very large. Laxminarayan et al. (2007) apply the full-income approach to model the expansion of DOTS coverage for TB treatment in Africa in the period 2006-2015, showing roughly a 10:1 ratio of benefit to costs under two alternative scenarios. Becker et al (2005) estimate the full income gains in a panel of 49 countries over a 30-year period in order to assess the degree of convergence between them, taking into account both income gains and improvements in life expectancy. The results suggest that full income has been converging across countries during this period, with an estimated increase of 140% for developed countries compared to 192% for developing countries.

Although full-income models offer the prospect of deriving estimates that more closely resemble the economic welfare losses of disease and injury, they currently face a number of challenges to their validity and reliability. Concerning revealed preference techniques used to derive the value of statistical life, a key measurement challenge is that wages are determined by many other factors besides the (perceived or actual) risk of death, which poses statistical problems for accurately determining the specific income-risk trade-off. A further problem in the context of economic burden studies - where the implied counterfactual is typically the elimination of disease - is that valuations of human life have generally been done by eliciting WTP responses for small reductions in the risks of death, which are translated into the value of a statistical life by assuming a linear relationship between WTP and risk. WTP is likely to be a non-linear function of risk – and as the risk of death approaches 1, WTP is likely to be bounded only by an individual's call on resources. Identifying the area under a non-linear WTP curve, which would be required for using the technique in economic impact studies, poses serious practical problems.

[2] In order to avoid double-counting, medical care is treated as an instrumental input into the health sector and subtracted from aggregate measures of consumption expenditures.

In addition, efforts to value human life typically sum individual valuations, thereby assuming that the individual would benefit from the reduction in risk or the health improvement. This method incorporates individual risk preferences into the total measure of benefit. Although the benefits of an intervention can be risky for an individual, such benefits at the population level are much more predictable, and a measure of social benefit from elimination should take this information into account as far as possible. Moreover, asking only the individuals who would benefit from an intervention ignores the fact that people are willing to pay to improve the health of others (Olsen & Donaldson 1998; Smith and Richardson, 2005).[3] A related issue, which also applies to the measurement of market-based consequences of disease and injury, refers to the assessment of external costs or benefits associated with ill-health. In cases where individuals do not bear the full costs or benefits associated with their actions, the resulting externalities can make it harder to assess health valuations based on WTP (Suhrcke et al 2006).

A further measurement and valuation challenge concerns the valuation of non-fatal health outcomes. Although estimates of the value of non-fatal outcomes are beginning to emerge, for example with respect to injury (Kuhn and Ruf, 2008), full-income studies to date have only incorporated changes in mortality. For many leading contributors to the global burden of disease and injury - including sensory and mental disorders - years lived with disability rather than years of life lost represent the key measure of diminished health. An apparent solution to this current shortcoming would be to derive an economic value for a composite measure of health loss, such as the disability-adjusted life year (DALY), but it would not appear to be possible to derive this using revealed preference techniques because they rely on observations made with respect to the trade-off between income and death (rather than loss of healthy life).

An alternative approach to the valuation of a DALY would be to employ stated preference techniques, but one of the fundamental challenges facing the implementation and validity of contingent valuation surveys relates to the hypothetical nature of the valuation, because although

[3] WTP techniques also assume that people have the information required to value benefits appropriately. Yet information failures are accepted to be one of the major imperfections in the health care market and one of the major justifications for government intervention. If we accept that governments must intervene because people are unable to make informed decisions about their own health, we suggest there is a logical inconsistency in using uninformed preferences to value the aggregate welfare loss associated with ill health.

WTP should be constrained by the ability to pay, this is easily overlooked if the individual will not actually have to make the payment out of his or her own pocket. There are a number of other, well-documented practical measurement challenges with contingent valuation (which we do not repeat here), ranging from the description of the scenario(s) to be valued and choice of respondent, through to framing effects and biases that may be inadvertently introduced into the valuation exercise (Drummond et al., 2005; Olsen and Smith, 2001).

Conceptually, therefore, a willingness to pay approach provides an appropriate measure of total economic welfare losses resulting from disease or injury, but the empirical concerns around its practical implementation means that there is considerable uncertainty around the appropriate social value to be used for assessing changes in overall economic welfare. Accordingly, we propose that empirically-based estimates of market losses be identified, measured and reported separately from hypothetically-based estimates of full economic welfare loss. Because WTP-based measures are supposed to take into account total economic welfare losses arising from disease or injury, one way of reducing the dependence on hypothetically-based measures of economic loss would be to restrict its use to domains of social welfare loss that cannot be adequately captured by other, empirically-based approaches. This could be achieved by subtracting from WTP-based estimates all of the market-based losses that have been estimated via an economic growth model. Such a hybrid approach would retain the objective of capturing full economic welfare losses, but would avoid double-counting problems and minimize the reliance on hypothetically-based estimation.

3.4. Macroeconomic impact: key points

Based on the above discussion of main analytical approaches used to assess the macroeconomic impact of disease and injury, we provide in Appendix C a summary overview that describes the main assumptions, underlying quantity of interest, data requirements, main outputs of analysis, advantages and limitations of each method, as well as a non-exhaustive compendium of possible data sources that can be used. In addition to models with a specific macroeconomic scope, the CoI approach is also included in this inventory because it is often used to come up with estimates

that supposedly reflect the economic impact of disease and injury in terms of GDP losses (see also Appendix B for further discussion on the limitations of the CoI approach).

- Macroeconomic studies assessing the impact of disease and injury should focus on three areas related to economic welfare: non-health consumption possibilities (both now and in the future), leisure time and health status. Studies in this area should clearly recognize how the chosen framework of analysis identifies each (or all) of these components.

- The use of the cost-of-illness approach is not recommended. Traditional cost-of-illness studies employ a static, partial and inconsistent approach to estimating the macroeconomic impact of disease and injury at the societal level. Instead, we recommend the use of more general and dynamic assessment of forgone consumption opportunities.

- Disease and injury impact the macroeconomic performance via a number of different channels, which include increased health expenditures, labour and productivity losses, and reduced investment in human and physical capital formation. Macroeconomic studies should take into account how these channels affect different macroeconomic agents.

- Studies that aim to estimate the macroeconomic impact of disease should carefully consider the issue of endogeneity in the determination of health and income. Several empirical approaches are available that can be used to mitigate or reduce this problem.

- Likewise, studies should take into consideration the several aspects of uncertainty surrounding current models, which relate to theory, measurement and specification. Any assumptions about current levels of prevalence/incidence, the impact on particular variables, economic elasticities and other economic relationships should be tested and validated through extensive sensitivity analysis. The results of sensitivity analysis should be documented and reported alongside the main results.

- The choice of modeling framework should be guided by the question that is being asked. If the focus lies on the overall GDP impact only, an economic growth model based on either calibration or estimation might be used. The calibration approach can be particularly useful if detailed data information is lacking.

- The evaluation of more disaggregated impacts, such as on specific sectors or categories of macroeconomic agents might require the specification of a computable general equilibrium (CGE) model. This type of model can also fully incorporate issues of dynamic adjustments and the inter-linkages across settings. However, the data requirements and computational costs tend to be high.

- Finally, if the interest lies in the determination of overall welfare, an approach based on willingness-to-pay (WTP) methodology is indicated. Currently, the most common application of this technique is via the specification of a 'full-income' model, which attaches monetary values to life years lost as a result of disease or injury. Conceptually, such an approach provides a more proximal measure of total economic welfare losses resulting from ill-health, but there are a number of empirical concerns around its practical implementation which mean that considerable uncertainty surrounds derived estimates. It is therefore recommended that empirically-based estimates of market losses be separately identified and reported from estimates of full economic welfare loss.

4. MEASUREMENT AND VALUATION: MICROECONOMIC LEVEL

In this Section, we discuss the implications of the conceptual framework outlined in Section 2 for the appropriate measurement and valuation of the economic burden of disease or injury at the *micro*economic level of households, firms and the government. As emphasized there, the costs that *should* be included in an economic impact study vary according to the question being asked, as well as the scope or perspective of the analysis. For example, the chosen scope or perspective of the analysis will determine whether only financially quantifiable costs (market losses) are included or whether non-market losses are also considered.

4.1. Microeconomic impact of disease and injury at the household level

Distinct from concerns over the societal or population-level impact of ill-health, there has been a steadily increasing level of policy and research interest over the last two decades in the microeconomic impact of disease or injury, focusing in particular on the impoverishing and other effects that ill-health or injury can have on households. Much of the policy concern and research literature relates to low- and middle-income countries, where the share of total health expenditure paid for out of pre-payment mechanisms such as social or universal insurance schemes is typically less than in high-income countries. Instead, households are required or forced to 'pay as they go' when illness or injury occurs, and for poorer households this can quickly become unsustainable or a major drain on household income. In fact, it is estimated that each year 44 million households worldwide face catastrophic health expenditures - defined as spending more than 40% of their non-subsistence income on health care payments - and about 25 million are pushed into poverty (Xu et al., 2003, 2007).

In addition to the 'direct' financial expenditures that households incur in obtaining the health services and goods that they need, economic impact studies have also studied a more 'indirect' set of consequences that may potentially befall households. Most notable are the losses in production or income which then translate into lost current consumption, and possibly lost future

consumption because of the impact on savings (or debt). A recent literature review in the context of low- and middle-income countries identified 62 research studies that assessed these direct and indirect costs to households. The authors found there to be considerable heterogeneity, not only with respect to the methods used in different studies but also across different disease entities (McIntyre et al., 2006). The most commonly studied diseases have been malaria and HIV/AIDS (Chima et al., 2003; Russell, 2004). Where indirect costs were included, they tended to exceed direct costs; in the case of malaria, for example, direct costs of treatment and prevention have been found to be in the range of 2%-3% of household income, compared to 2-6% for indirect costs (Russell, 2004) - but differences in methodological quality and approach mean that such findings need to be interpreted with caution.

An important line of enquiry underlying both the direct and indirect costs of disease at the microeconomic level has been articulation of the 'coping strategies' that households utilize in order to mitigate the unwanted consequences of illness, including substitution of labour within the household (to preserve production and income flows) and the disposal of assets (to pay for health care) (Sauerborn et al., 1996). As pointed out by Ahlburg (2000) and Russell (2004), however, there remains a limited understanding about how effective these coping strategies are in terms of longer-term asset and/or livelihood preservation; coping mechanisms are not costless and may entail costly consequences for the household in the future (i.e. negative impacts are deferred). For instance, the education of children can be seriously compromised if they are taken out of school and required to start working in order to compensate for the death or serious sickness of a household member. This can have obvious implications for employment and income generation opportunities that in turn might affect the household's consumption choices in the future.

Another important aspect refers to the impoverishing and consumption impact of disease and injury. In the absence of social security and other forms of formal / informal compensating mechanisms, increased health expenditures and lower productivity capacity can cause or accentuate household poverty and reduce the amount of resources available for non-health consumption, food-items in particular. In order to assess the overall impact on households, several studies have examined the degree of household consumption insurance (the extent to

which households are able to smooth the path of consumption of non-health goods and services following negative health shocks). As discussed above, households resort to a number of different coping strategies to deal with the negative consequences of health shocks, including labour supply adjustments, drawing on savings, borrowing, remittances and social assistance. In resource-poor settings, however, these adjustments are not always able to accommodate the impact of disease and injury, and households are forced to cut back the consumption of non-health goods and services.

The degree of consumption insurance provided by private and informal institutions has clear policy implications. If households are able to smooth consumption using private channels, then the public provision of health care and social security might simply crowd-out the private efforts, with limited gains for the society (Morduch, 1995, 1999; Gertler et al., 2008). Nevertheless, there is ample evidence from developing countries indicating that household consumption is negatively affected by health shocks. Gertler and Gruber (2002), for example, report that there is far from perfect consumption insurance in the case of severe illnesses in Indonesia. The authors estimate that a 10,000 rupiah reduction in income resulting from a health shock decreases non-health consumption by 10%, meaning that households are able to insure less than 40% of the income loss due to severe illness. Similarly, Wagstaff (2007) evaluates the degree of consumption insurance following health shocks among Vietnamese households. Although households benefit from increased flows of unearned income (social security, informal solidarity and remittances), they are not able to smooth food consumption in the face of health shocks.

These results, together with those separately produced by Abegunde and Stanciole (2008), indicate that the consumption of non-health items such as electricity and housing can also increase as a result of illness, adding further strain on the ability to maintain current levels of food consumption. Mendola et al (2007) also provide evidence from the Western Balkans to suggest that high levels of out-of-pocket expenditures can increase the incidence of poverty and push households into poverty. This evidence suggests that in many developing countries the degree of consumption insurance against illness falls short of full insurance and that there is scope for expanding systems of social protection and heath care coverage.

Many of the most important potential effects on households have not been adequately measured in economic impact studies to date. Few have measured the impact of ill-health on levels of household productivity, which is likely to be lower for people with chronic conditions (Ramu et al., 1996), or the impact on long-term production through reductions in savings, changes in household activity patterns, or reduced educational investment (e.g. Kochar, 2004). The end result is that many studies produce estimates of lost time and productivity costs that have little to do with the impact of illness on the discounted present value of household consumption.

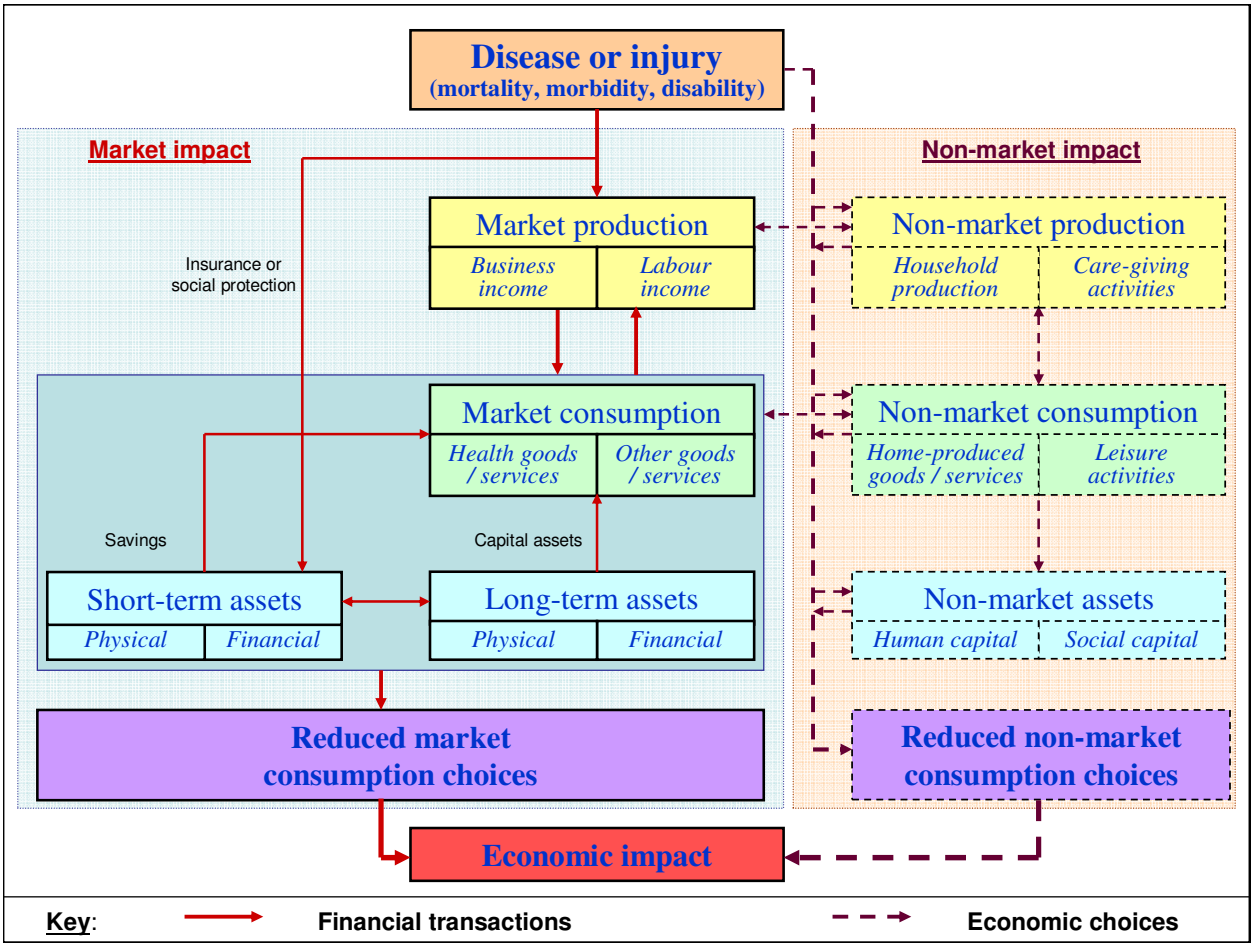
4.1.1. The impact of disease and injury on the household

As discussed in Section 2.1, households seek to optimize their opportunities to consume the goods and services that they most value, both in the form of market goods and 'non-market' goods. In addition, they derive utility from leisure time, and from being healthy. The main constraint on their consumption of market goods and services is the amount of income that they have at their disposal (from earned or business income, derived from assets that they own, or from borrowings). Households also need to take into account future consumption needs (e.g. for health care or children's education), so may decide to withhold some of their current income or consumption opportunities for the future (i.e. savings, investments). In this sense, they strive to maximize the utility derived from future consumption. The object of choice in this case is the present discounted value of alternative consumption flows.

Disease or injury may interfere with these economic objectives and choices in a number of ways (see **Figure 3**). In the simplest case, the incidence of a disease or injury event has two immediate potential effects. Firstly, the diseased or injured person may have to reduce their normal level of productive activity (whether paid or unpaid) and, secondly, the household may need to increase its consumption of health services or goods (at the expense of other goods and services). In terms of market impacts, reduced earnings from business and/or labour income plus additional expenditure on health services must reduce non-health consumption opportunities. Households may reduce their consumption of these non-health goods and services (for example, by cutting back expenditure on clothes, social activities or durable goods), or they may try to maintain current levels of non-health consumption by liquidating household assets (such as cash savings

held in a bank account) or resorting to loans. They may also have to cut back on other, non-market activities (e.g. including household production and subsequent consumption of home-produced goods and services) or their investment in people, e.g., education, health and social capital formation (Steinberg et al., 2002).

Figure 3 Financial and economic impacts of disease or injury on households
(single period case)



Households make these consumption choices in the form of trade-offs between consumption in current and future time periods, and in their time allocations to market production, non-market production, health improvement and leisure. Ideally, these trade-offs should reflect both *time preferences* (consumption today versus consumption next period) and *risk preferences* (certain consumption versus uncertain consumption). Following Deaton (1992), we can more formally articulate such household inter-temporal choices over alternative consumption plans. In the

simple two-period case, and abstracting from leisure and the production of health, one could have $u = v(c_1, c_2)$, where u is the level of utility derived from consumption, v is the utility function, and c_1, c_2 are the levels of consumption of goods and services in periods 1 and 2, respectively. The household maximizes utility subject to constraints on income and wealth. There are various approaches to solving this optimization problem, depending on what is assumed about the trade-offs that households make between consumption levels in different periods (that is, the degree of substitutability of consumption). A fairly common assumption is to assume that these preferences are inter-temporally additive, i.e., $u = v_1(c_1) + v_2(c_2)$. This expression for household utility could be modified to account for other characteristics (z) that shape consumption preferences, such as income and age. In addition, a time preference parameter (δ) could be incorporated that is used to discount the future consumption bundles to a single present value.[4] In a more general specification and given the appropriate discount rate, the optimal path of lifetime consumption could be determined. One issue to face in making use of such lifetime calculations, however, is that consumption choices are assumed to be inter-temporally consistent.

To incorporate uncertainty into this framework, utility in the preceding model is replaced with expected utility (EU). Thus, $u = E_t [v_1(c_1) + v_2(c_2)]$, where E_t is the conditional expectation of the utility derived from consumption (conditional in the sense that the expectation is based on information available to the household at time t). In order to make uncertainty more explicit in this model, this expression can be rewritten in terms of the probabilities of these events (π). Thus, $u = \pi_s v_1(c_{1s}) + \pi_s v_2(c_{2s}) + \pi_w v_1(c_{1w}) + \pi_w v_2(c_{2w})$, where there are two states (s and w) that represent uncertainty. In this way it is possible to see how risk preferences enter into consumption choices. The EU model explicitly assumes that utility from consumption is also additive over these states. In the EU model, risk-averse households will be less responsive to inter-temporal incentives, while households who are less averse to risk will be more willing to reallocate their consumption in response to these incentives (Deaton, 1992). In other words, the degree of inter-temporal substitutability is inversely related to the degree of risk aversion.[5]

[4] Some studies have shown that time preference rate is related to health status (Lammers and Wijnbergen, 2007).

[5] The close relationship between risk aversion and inter-temporal substitution is averted in the generalized expected utility (GEU) model, which was developed independently by Epstein and Zin (1989) and Weil (1990).

The EU model is useful for analyzing the microeconomic implications of a health shock for household consumption. The empirical challenge is to assess the impact of an adverse health shock at the household level in terms of the present value of consumption foregone, which requires the expected incremental effects of the health shock on consumption to be monetized in each period. These incremental effects can be defined as the value of consumption with the disease/injury, *less* the corresponding value of consumption without the disease/injury (the counterfactual), i.e. the net changes in utility derived from the consumption directly and/or indirectly attributable to the adverse health event. As discussed earlier, we can seek to derive the present value of the reduced flows of consumption over time, or the welfare value of this reduction. The complexity of these estimations are discussed in the following sections.

This approach also requires that an appropriate time preference rate be derived in order to discount the future consumption foregone. Following the expected utility approach, the appropriate rate of time preference is conditional on risk preferences. The elicitation of these preferences is discussed in Lammers and Wijnbergen (2007). The expected utility approach they use is summarized in Appendix D. For example, assuming that households have constant relative risk aversion utility functions, the time preference rate can be estimated and applied to the projected stream of incremental consumption foregone in order to assess the impact of the disease or injury. Given that there are multiple persons in a household, this may require that the time and risk preferences of the head of household be used as a proxy for those of the household unit. The practicality of trying to do this is discussed subsequently.

4.1.2. Expenditures on health by households

At the household level, costs incurred in the acquisition of health services should represent the resources that could have been used for other types of consumption had the disease or illness not occurred, taking into account the fact that the impact on the present value of future consumption depends partly on whether the costs are funded from current income, savings, sale of assets or borrowing. Absolute monetary levels of household out-of-pocket expenditure on health generally or on a specific disease or injury can be linked to household spending for all goods and services in order to derive a relative measure of financial burden (Xu et al., 2003).

Household out-of-pocket spending can be defined as "the direct outlays of resident households, including gratuities and payments in-kind, made to health practitioners and to suppliers of pharmaceuticals, therapeutic appliances, and other goods and services, whose primary intent is to enhance and restore the health of individuals or population groups" (WHO, 2003). For economic burden studies, where the counterfactual is framed in terms of the absence of disease, it is also relevant to include the costs of health insurance borne by households (if disease did not exist, there would be no need to pay insurance premiums).

The question of insurance raises an issue concerning whether the focus of economic impact studies is on a particular disease / injury, a group of related disease entities (around cardiovascular disease, for instance) or on ill-health generally. Households may choose or be required to pay insurance premiums as a form of pre-payment in the event of a household member falling ill and requiring health services. By definition, this is non condition-specific since the eventuality of ill-health is not known with any certainty, whereas other household health expenditures (over and above what is covered by the terms of the insurance) tend to be specific to a particular health condition. These costs include fees - or if insured, copayments - for medical consultations or diagnostic tests for a given health condition, expenditures on medications, plus other 'non-medical' costs such as travel expenses. Since most economic burden studies in health have the objective of demonstrating the magnitude of losses due to a particular condition, this is the approach taken here.

Adoption of such an approach, however, requires the apportionment of insurance premiums to specific disease or injury conditions, and this in turn is likely to require an actuarial exercise to determine the amount that premiums would fall by in the absence of any risk of a specific disease or any resulting insurance claims resulting from it. While this is theoretically a suitable approach to take, it may be decided on practical grounds to use cruder approximations of the fraction of insurance premiums that are associated with a particular disease (informed by national health sub-accounts if available, for example) or even ignore this category of household expenditure on health, particularly in settings where health insurance payments represent a modest component of

total health-related expenditures. It should be borne in mind that underestimating health-related cost in this way will result in an overestimation of non-health consumption flows.

Financial payments towards the partial or full cost of health care services or technologies (e.g. hospital inpatient and outpatient care, primary health care, diagnostic tests and drugs) can be determined from household surveys, either in relation to overall income and expenditure patterns, or in specific relation to health (such as the World Health Survey, which has been conducted in more than 70 countries worldwide; www.who.int/healthinfo/survey). There are a number of problems, however, with household surveys. Firstly, questions asked in different ways produce different estimates of expenditures, with no clear patterns of 'bias'. For example, the World Health Survey asked two questions at different times, covering the same period. The first requested simply total health expenditures. The second asked for the components separately. In some countries, the sum of the components exceeded the response to total health expenditures based on the single question, while in other countries the opposite occurred. Secondly, recall biases are likely to be present (the longer the recall period, the larger the bias). It is likely that the responses to expenditure questions will be more accurate for large, unusual items such as hospitals, or for recurrent items such as chronic diseases, than for small, occasional expenses. Thirdly, many questionnaires ask for expenditure incurred by someone with a condition, and ascribe all of this to the condition. This may not be appropriate since some of these expenses may well have been incurred in the absence of the disease. This requires a control to identify the counterfactual, and estimates based on responses where there is no control group will generally be overestimates. In order to avoid such overestimation, one needs to establish the net impact (which adjusts for the fact that some expenses are incurred anyway, as for example shown by Mills (1994) in her study of the economic impact of malaria on households in Nepal, or by Wang et al (2006) in their study of the impact of chronic disease in rural China.

Non-medical intervention costs include financial payments for (public or private) transport to health care facilities; subsistence costs while attending a hospitalized household member; special food needs; asset modifications resulting from ill-health / disability (e.g. wheelchair access to one's home); repairs to damaged property (e.g. to vehicles as a result of substance abuse); and any other household financial resources that are consumed as a result of the health condition. The

measurement of these financial outlays can be obtained by surveying a representative sample of households with the condition in question, again with suitable controls in place to avoid overestimation of these costs by making reference to a counterfactual.

4.1.3. Labour and productivity losses

An important economic consequence of disease or injury at the household level is that, through its impact on functioning, individuals are unable to perform their usual day-to-day activities. Just as being fit and healthy tends to boost productivity (particularly in lower-income countries where employment opportunities tend to depend more on physical strength and endurance; Thomas and Strauss, 1998), so poor levels of health restrict productivity. In the macroeconomic context, the question of whether a sick worker will be replaced from the pool of under- or un-employed workers will need to be considered, but in the microeconomic context, this is a clear loss to the particular household concerned. Prolonged absence from paid work due to illness can be expected to increase the need to pay for health services and goods out of other household resources (particularly savings), while the inability to carry out unpaid but productive household activities because of ill-health (including child care, meal preparation etc.) may lead to time being taken off from paid work by another household member or necessitate the purchase of these services or goods on the open market. Non-market consequences, which contribute towards the total impact of disease or injury if an economic welfare perspective is taken, are dealt with in a subsequent sub-section (4.1.5); here the focus is on marketed production and income.

In principle it would be possible to distinguish separate health effects on labour supply (labour force participation, number of days or hours worked, early retirement for health reasons) and on labour productivity (the quality of work or the amount of output produced per unit of work)[6]. However, as noted by Suhrcke et al. (2005), in practice it is common that studies focus on the impact on earnings or wages, which contains elements of both labour supply and labour productivity effects and therefore prevents the identification of them separately. The potential endogeneity between health and economic outcomes is another important issue that has to be

[6] See Suhrcke et al.. (2005) for a comprehensive review of studies dealing with the labour market impacts of health, particularly in European countries.

taken into account. As discussed in **Box 2**, this is a major concern in macroeconomic studies that can also be observed at the household level and which if ignored can yield inconsistent estimates of the effect of health on labour productivity and other household outcomes. Econometric techniques are available that can help to control for the endogenous determination of these variables.

Because of the sensitivity and complexity of questions around income and lost production, many economic impact studies of disease and injury to date have used time measures such as lost days of work as a proxy measure for lost output, to which a monetary value is then attached. A number of approaches have been used for attaching a monetary value to lost days of work, based either on the value of lost income level (such as a daily wage rate) or on the value of lost production (such as GDP per capita). Most simply, and for that reason the most common too, an estimate of lost income is made for the sampled or target population as a whole (e.g. average wage rate for working adult males), while more data-intensive studies attempt to capture the economic loss of each individual (via specific wage rates or the market value of reduced units of output). Perhaps unsurprisingly, different valuation methods can produce widely diverging final estimates of economic burden (Attanayake et al., 2000; Verstappen et al., 2005). For example, an output-related cost study of the household burden of malaria in Sri Lanka found that costs varied from three times more to five times less than their baseline estimate, depending on the approach and underlying assumptions used (Attanayake et al., 2000).

A central limitation of the input-based approach is that it assumes that every economically active person is a wage earner, when in reality many individuals or households derive income from small businesses that they own and operate. In addition, studies that use an input-based approach to calculate lost household production are most likely overestimated because they simply assume that the duration of an individual's absence from work fully corresponds to the market value of those lost days. However, such an assumption overlooks the so-called 'coping strategies' used by households to mitigate the adverse circumstances of one of their member's being ill.

Coping strategies are of particular significance in predominantly informal or subsistence economies, where work output may be just as much a collective household effort as it is an

individual one (Sauerborn et al., 1996; Russell, 2004). In these economies, days of work lost due to ill-health in one family member may be made up by or substituted with the work inputs of another, thereby mitigating losses in household income. However, it is important to note that coping mechanisms are not costless and can also entail costly consequences for the household, even if these consequences are deferred and impact only on future consumption possibilities (Ahlburg, 2000). A further source of potential overestimation is that, particularly in (mostly higher-income) countries, sickness absences from work are typically paid for by employers or the state, meaning that from a household perspective the loss is again largely if not wholly mitigated.

For measuring the value of lost market production, as with the additional health expenses discussed earlier, it is important to assess the value of the production of the sick person *and* his/her family *compared* to the counterfactual of what would have happened in the absence of the illness. This can be achieved via what has been termed the output-related approach (Goldschmidt-Clermont, 1987), which focuses on measurable changes in income or product rather than time inputs; only a few studies have used this approach to date (see Attanayake et al., 2000). A central requirement for estimating the actual (rather than potential) losses in production is the comparison of households with and without the health condition in question, so that it is the net effect that is captured and attributed to the condition in question. This is particularly important in agricultural societies or for people engaged in informal labour where there are seasons in which work intensity is high and others in which work intensity is low. In these situations, days ill do not necessarily translate neatly into days of lost work.

We conclude that studies which take the reported or observed days of illness and simply attach an average wage rate to each day - whether adjusted or not for age and gender - stand a high chance of overstating the economic impact of lost market production, particularly in settings where this is a significant degree of substitutability in labour inputs. We therefore recommend that an output-related approach should be used wherever possible to identify total net losses in production or income, since it is better equipped to isolate only the fraction of market production of a household that is actually (rather than potentially) lost, partly as a result of taking into account household coping strategies that may mitigate these potential losses. The potential adverse consequences of these short-term coping strategies are discussed in the following section.

4.1.4. Effects on human, physical and financial capital formation

As discussed earlier, a proportion of a household's health expenditure may be paid from savings, loans or the sale of assets, particularly if they are poor. This, combined with reduced savings associated with lowers income, has potentially damaging consequences, not only in terms of the elevated risk of impoverishment or indebtedness, but also in terms of reduced opportunities to generate the stock of financial and physical capital that will enable it to maintain or increase its consumption possibilities in the future. For example, a study of the economic impact of tuberculosis in Thailand found that an estimated 15% of poor households sold property and 10% took out loans in order to meet the costs of treatment (Kamolratanakul et al., 1999). The economic value of these diminished savings or losses in capital stock clearly represents part of the overall economic impact of ill-health on households - foregone consumption opportunities in the future - and should consequently be included in the estimation process.

One distinction that needs to be drawn here relates to assets that can and cannot be liquidated. Land, buildings, livestock and financial products can and do get liquidated due to household illness. A second category of assets include those that cannot be readily liquidated or sold on the market, such as social networks and human life itself. These forms of (human and social) capital and their potential depletion as a result of disease or injury are discussed below in Section 4.1.5, but we note here that in the case of human capital and its formation, the picture is quite complex because there are both market and non-market aspects to consider. For example, the decision to reduce expenditure on school education because of illness can be interpreted as a loss in non-health market consumption (which may lead to diminished production or income levels as an adult), whereas the foregone nurturing and knowledge transfer that follows from the death of a child's parent is a non-market loss in human capital formation.

Physical and financial capital formation

Very few microeconomic impact studies in health have assessed depleted investments in physical and financial capital formation. This is partly because it is not a significant issue in (mainly high-

income) countries with robust health insurance and welfare systems (which protect households from excessive health spending; Smith, 1999), but is also due to the inherent difficulties of measuring or predicting these longer-term consequences of disease or injury. Measurement and valuation of these foregone savings or liquid assets at the microeconomic level of the household could potentially be informed by additional survey questions to be administered alongside those relating to health and non-medical costs. As before, such information needs to take account of the expected losses that occur *without* disease or injury, which points to the need for suitable control households to be included in the survey. But there are a number of measurement difficulties related to the validity and reliability of self-report responses to questions around income and wealth that need to be carefully addressed and controlled for (Turrell, 2000). For example, use of retrospective study designs and data is complicated by the possibility that unobserved factors may have determined the outcome of interest, while randomized controlled trials pose a number of more practical challenges.

Nevertheless, cross-sectional or prospective household surveys have addressed these issues in a measurable way. One study from rural China, for instance, utilized the baseline survey of a community-based rural health insurance study to estimate via logistic regression the impact of ill-health and associated medical expenditures on household investments in human and physical capital (Wang et al., 2006). The authors found that households with a member hospitalized in the last year spent three times as much on medical expenses as those that did not, and this expenditure had the effect of 'crowding out' other categories of consumption, including education (26% reduction), farming expenses (15%) and savings (47%).

Human capital formation

Over and above reduced expenditures on education in the short-term, there are potentially significant longer-term consequences to a child being out of school (whether forced to by their own ill-health or by the economic situation that a household finds itself in). Education is widely agreed to constitute a vital investment for economic development (both at the household and societal level), but stands to be compromised by ill-health in a number of ways. Most obviously, a child may miss days or weeks of schooling because of ill-health in the same way that adults

miss work, but with the difference that the economic value of this absence needs to be measured in terms of future rather than current earnings.

More insidiously, parasitic or other childhood diseases may impinge on the physical or cognitive capabilities of children, similarly resulting in diminished educational attainment levels and consequent earnings (Alderman, Behrman and Hoddinott, 2003). In a worst case scenario where average life expectancy in a population is severely impeded by an epidemic disease such as HIV/AIDS, educational investments may be seen as a luxury or high-risk good that is likely to offer only a modest return on the household's investment. In high child mortality settings, households often compensate for the elevated risk of a child's death by having many children, thereby reducing the amount available to invest in each surviving child (WHO, 2001).

Although there is a sizeable literature on the association between nutritional status, ill-health and subsequent education, the longer-term value of these depleted investments in terms of lost earnings is hard to quantify, and depends on a number of assumptions concerning observed versus expected earnings for different levels of educational attainment. Many of the studies of impact of diseases on education look at absenteeism rather than years of schooling, while the studies measuring impact of education on income look at years of schooling. There does not appear to be any straightforward way of moving from absenteeism to years of schooling. Similarly, studies looking at the link between cognitive function and disease largely measure impact with respect to particular tests rather than to years of schooling (which is what would be required in order to arrive at a reasonable estimation of the impact of disease on future earnings). One or two studies, however, have succeeded in making the connection between health status, school attendance and lifetime earnings. For example in Zimbabwe, stunting, via its association with a 7-month delay in school completion and a 0.7-year loss in grade attainment, has been estimated to reduce lifetime income by 7%-12% (Alderman, Hoddinott, and Kinsey, 2003), while iron deficiency in adults has been estimated to decrease productivity between 5%-17%, depending on the nature of the work performed (Horton, 1999).

Conceptually, therefore, it is clear that future earnings as an adult foregone as a result of disease or sub-optimal health as a child/adolescent represent an integral component of financial (market-

based) loss. What is less clear is how to go about the measurement and valuation of this loss, due to the long lag-period concerned, the influence of confounding factors and the inter-connected pathways of causality between health, income and wealth. Evidently, a longitudinal or life-cycle approach to the modeling of these relationships or effects is needed; one-off cross-sectional surveys cannot adequately capture the changes in time and resource allocation that occur over time following a health shock. But mounting new multi-wave longitudinal surveys that incorporate both health and wealth information require intensive and lengthy research and data collection resources that are rarely available outside the wealthiest countries.

In practical terms, therefore, a first analytical step would be to ascertain the relevance of this dimension of economic loss; for example, tobacco use is an important risk factor for disease, but does not impact on school attendance or educational outcomes (Leslie and Jamison, 1990). Where such losses are likely to be significant, as with nutritional deficiencies for example, a second option would be to utilize or adapt existing evidence on the links between school participation, educational outcomes and adult earnings. In the absence of such data availability, the remaining options would be to build on / repeat existing households surveys (with a view to deriving prospectively-derived estimates of impact) or to omit this component from quantitative estimation and instead describe the expected extent of the losses in qualitative terms.

4.1.5. Non-market impacts

Lost non-market production and consumption

This section addresses the question of to what extent ill-health also impacts on the opportunity to produce or consume non-market goods and services. Time spent by a household member providing informal (and financially uncompensated) care to a sick person could have been spent on other productive activities or in leisure, while time spent accessing and receiving health care could have been alternatively spent on non-health related activities (consumption of home-produced goods or leisure). In addition, ill-health can impact on the non-marketable assets of a household, including its repository of knowledge, experiences, and social networks, plus its stock of health.

As with market production, the relevance of non-market production to ill-health needs to be understood in terms of its potential contribution to changes in consumption possibilities at the household level. Since non-market production is not financially compensated, however, it does not effect the allocation of financial resources, and therefore only enters into consideration if the scope of the economic impact extends beyond these strictly financial flows to a broader framework - i.e. if the quantity of reference is social product as defined earlier. Possible ways in which ill-health might change consumption opportunities via its effect on non-market production include:

- reduction in the time spent in non-market production by a person because of illness, including travel time to seek or use services;

- time spent by a household member caring for a sick person (within or outside the household), at the expense of other productive activities (marketed or non-marketed);

- substitute non-market production (possibly at the expense of market production) by a household member on behalf of another (unpaid) household member who is sick. This is the non-market side of coping strategies that are used by households to minimize the impact of ill-health, and as such the actual economic consequences are likely to be small.

Unlike marketed production, where the underlying quantity of interest is income or money, measurement and valuation of non-market production is concerned with arriving at an estimate of the economic value of any loss (which must be computed indirectly since household production is a non-market activity). Studies in the past have typically tackled the valuation of these losses via an input-based approach that first assesses time loss and then applies an economic value to it. Two main approaches have been used:

- *Opportunity cost method*: The first approach focuses on the opportunity cost of devoting time inputs to household production and sets a value for it as the income the individual could earn in the market. The rationale behind this approach, which stems from Becker's

(1965) theory of the allocation of time, is that individuals allocate their time in much the same way as they allocate their income (that is, the margin benefit of an hour spent on work, leisure or household production is the same). The drawbacks of this 'foregone earnings' approach are that a) many individuals engaged in household production are involuntarily unemployed (which suggests a potential earnings value of close to zero) and b) it equates the value of time spent on household activities with the production of potentially very different types of marketed goods and services (Chadeau, 1992).

- *Market cost method*: The second approach proposes that household production can be measured on the basis of services purchased through the market and can be valued as the cost of hiring someone to do the household tasks. Unlike the opportunity cost approach which uses time to value the inputs of household production, the market cost approach requires measurement of output, namely the quantities of goods and services produced within households. This output-oriented approach has been used to generate so-called household satellite accounts, which are designed to supplement market accounting measures such as GDP by placing monetary values on the 'productive activities' that households perform (defined as all goods and services that could have been produced by another economic unit). The main limitation of this approach is that it does not easily extend to important productive household activities that cannot be contracted out to a third person, such as studying or health-seeking (Abraham and Mackie, 2004).

Although the two methods *should* give identical valuations, disaggregated analysis taking into account worker and family characteristics suggests that the market cost method may be preferable if the purpose of valuation is to incorporate estimates into national accounts (i.e. account for services produced), while the opportunity cost method is preferable if the purpose is to determine the value of welfare losses to individual household members (James, 1996). It might even be argued that any sort of valuation is to be avoided, for example because of the difficulty in separating out productive versus leisure elements of household work. To leave non-market production or consumption expressed as an assortment of non-monetized physical quantities, however, runs the risk of removing consideration of these quantifiable losses from policy dialogue, which is tantamount to placing a value of zero on these valued activities.

From a conceptual (as well as consistency) point of view, there is greatest appeal for a market-oriented output approach to valuing non-market losses. Again, this is because such an approach remains closest to the principle of valuing actual (rather than potential) foregone economic activity or production. Use of an input-based approach may exaggerate the negative consequences of disease or injury; for example, days of household work lost due to ill-health in one family member may be made up by or substituted with the unpaid inputs of another, thereby mitigating losses in household production and consumption as a result of ill-health. As with market losses, the appropriate measurement of any change in output resulting from disease or injury would require the use of a matched set of control households, so that only the net effects attributable to a health shock are counted.

Adoption of an output-based approach to the valuation of non-market production is not without its challenges. Certain non-marketed household outputs such as home study are hard to quantify in economic terms (compared to others which have a readily measurable economic value, for example home-grown vegetables). The challenge of appropriately measuring and valuing these economic consequences suggests that a careful decision needs to be made about whether to include or exclude these potential losses in a study. For some specific diseases, however, such as Alzheimer's disease, these economic losses are known to be quite substantial and would therefore merit inclusion. In these cases, a set of household-level productive activities that are adversely affected by the disease would need to be defined, and then compared to matched households without the disease in question. Any net difference in household output can subsequently be multiplied by its market value in order to derive an estimate of foregone household production. The simplest and most conservative market price to use for valuing such productive activities is the wage rate of a person employed to undertake housekeeping duties.

Lost leisure

The economic value or opportunity cost of leisure time can theoretically be identified as the point at which an individual is indifferent to either an increase in income or an increase in leisure time (Becker, 1965). Under this approach, individuals derive utility by combining market goods with

time to produce 'final commodities' (such as a meal prepared at home with market produce). In this sense, time can be viewed as an input into - as well as a constraint to - individual or household utility, and its value can be determined empirically via stated preference (e.g. the amount that an individual is prepared to pay in order to give up one hour's paid work) or via revealed preference (e.g. the observed relationship between hours worked and income received)

For paid workers in a perfectly competitive market (with no taxation), it can be shown that the price of an hour of leisure at the margin is the money income that is given up by not working that hour; therefore, the price of the last unit of leisure can be taken to be equivalent to the hourly wage (Posnett and Jan, 1996). In the real world, however, there is commonly an excess supply of labour (due to unemployment), markets are not perfectly competitive (due to monopoly, for example) and workers forego a proportion of gross income in the form of taxes [7]. Under these prevailing market conditions, a closer estimate of the opportunity cost of leisure for paid workers - although strictly true only for the last unit of leisure - is given by their net (as opposed to gross) wage rate (Posnett and Jan, 1996).

For unemployed persons, one would theoretically need to establish the personal qualities and experiences of those affected by ill-health in order to know the value of what has been foregone, but in practice a crude indicator such as the average market wage for the relevant occupational group might be used. For ascertaining the pure consumption value of leisure time (e.g. for the voluntarily unemployed), studies have generated estimates ranging between 25%-75% of gross hourly wages (Lam and Small, 2001; Posnett and Jan, 1996), based on discrete choice experiments of the trade-offs that individuals (such as commuters) are prepared to make between time (saved traveling) and money (spent on road tolls). Such estimates could be used as one basis for indicating the possible range of these economic losses, although analysts should be aware of the inherent uncertainties and limitations associated with the valuation of this foregone time using stated preference techniques. As implied above, however, these methods apply at the margin, and can be used to approximate the value of small losses of leisure time only. They overstate the value of large losses in leisure time.

[7] In addition, time spent working might be considered by individuals to be pleasant or unpleasant, suggesting that people's e the value or utility of work might not be wholly captured by the wage rate.

4.1.6. Economic welfare losses

As discussed earlier, the ultimate quantity of interest in economic impact studies is economic welfare, made up of the market and non-market components described above but also non-market assets, which extend from a household's accumulated knowledge and social networks, through to its stock of health and well-being. Disease or injury that results in death, disability or suffering robs individuals and households of these valued non-market assets, each of which contributes to the intrinsic value of human life and happiness. Consequently, estimation of these utility losses represents an integral part of a full economic welfare analysis of the consequences of disease and injury to households.

Since the value of changes in overall economic welfare cannot be valued with reference to observed market transactions, measurement necessarily requires indirect imputation based on the stated or revealed preferences that household members have for maximizing their utility. As discussed in Section 3.3, stated preference techniques ask people the maximum amount they would hypothetically be willing to pay for a particular outcome or commodity, whereas the revealed preference approach uses market observations to value the trade-offs that individuals are prepared to make in order to satisfy their preferences (from which estimates of the value of a statistical life have been derived; for a detailed review, see Viscusi and Aldy, 2003).

In theory at least, the monetary equivalent of these preferences - expressed in terms of people's willingness to pay - incorporates the *total* market and non-market utility associated with a particular outcome; for example, estimates of the statistical value of a life year theoretically include not only the intrinsic value of life itself but also the consumption opportunities that go with it. Since estimates of the statistical value of life relate to the trade-off between income and the risk of *death*, however, they do not encapsulate the value associated with avoiding non-fatal outcomes. This implies that value of life estimates will *under*-estimate the true cost of diminished health for most diseases. Yet on other grounds there are reasons for thinking that WTP-based estimates provide a top-end valuation of lost life (Olsen and Smith, 2001; Smith and Richardson, 2005).

There are a number of other important concerns around the validity and accuracy of WTP-based measures of welfare loss or gain, which are described in Section 3.3 and elsewhere (e.g. Gyrd-Hansen , 2005; Smith and Richardson, 2005; Olsen and Smith, 2001; Viscusi and Aldy, 2003). The way WTP questions are framed, for example, is expected to have an influence on the extent to which a complete valuation of welfare change - which should capture not only the intrinsic value of health but also the discounted present value of future non-health consumption opportunities - is actually obtained. As a result, there remains considerable uncertainty around the appropriate value to apply to these health-related losses. Risky behaviours such as smoking, excess drinking or drug-taking are a case in point. Such actions are a source of harm and increased health risks, but may also generate direct utility or satisfaction for users, posing additional challenges on how to measure the net contribution to welfare.

Given these concerns we would propose at a minimum that where non-market consequences are included in an economic impact analysis, they be reported separately from market losses, and that where WTP is used, it is restricted to the valuation of non-market losses not captured by empirically-based methods. This can be arrived at by first adopting the measurement approaches advocated above for the valuation of marketed losses and also for specified non-market production losses, and then subtracting these elements from WTP-based estimates of the value associated with the change in welfare. Such a hybrid approach comes closest to measuring the full economic welfare losses of disease or injury while maintaining the distinction between market and non-market losses and minimizing the reliance on hypothetically-based values.

4.2. Microeconomic impact of disease and injury at the level of the firm

A primary reason for studying the microeconomic impacts of disease or injury on firms is to measure their expenditures on health inputs and how changes in labour productivity may result in financial (monetary) losses. Indirect productivity losses are added to direct costs at the firm level to estimate the total economic losses to the firm.

Although productivity effects can be dealt with at the macroeconomic level, there is a plausible case for looking at their impacts at the firm level too. The motivation for a microeconomic approach is that the effects of an adverse health shock are not likely to be uniformly distributed across firms or industries. For example, the costs of offsetting the shocks may be relatively higher for firms that have higher labour intensity and/or lower capital endowments. Also, firms may vary in their previous efforts to mitigate potential health shocks, so that the magnitude of the impacts can vary across firms due to their optimizing behavior. The level of such previous efforts could vary between firms according to the desire to avoid health risks or the associated health costs. In addition, the effects on investment, rates of firm growth, and the flow of income to investors may also vary across firms because of their differential ability to adjust their operating and investment activities to respond to the health shock and maintain efficiency.

4.2.1. The impact of disease and injury on the firm

Just as firms differ in their organizational forms (e.g., private or publicly-traded, non-profit or for-profit, corporate or non-corporate partnerships and sole proprietorships), they may attempt to achieve different underlying objectives. For example, they may attempt to: achieve a profit objective; maintain or increase market share; or maximize shareholder wealth through a combination of business management strategies. This being the case, the financial and economic effects of health shocks should be approximated with reference to the appropriate firm objective. To simplify the discussion we will focus on for-profit corporate firms and suggest that firms attempt to maximize shareholder wealth. This is frequently cited as the most inclusive firm-level objective, since it incorporates the fundamental investor trade-off between risk and return (Ross, Westerfield and Jordan, 2006).

Thus, we assume that shareholders invest in companies to obtain the future dividends and market value appreciation that stocks provide. The implied cash flows and increases in wealth to investors/owners represent future consumption possibilities. Investors buy corporate debt instruments for approximately the same reason. Generally, we assume that firm managers and investors make business and financial decisions depending on their underlying time and risk preferences. Since shareholders invest for the purpose of generating a return, the economic

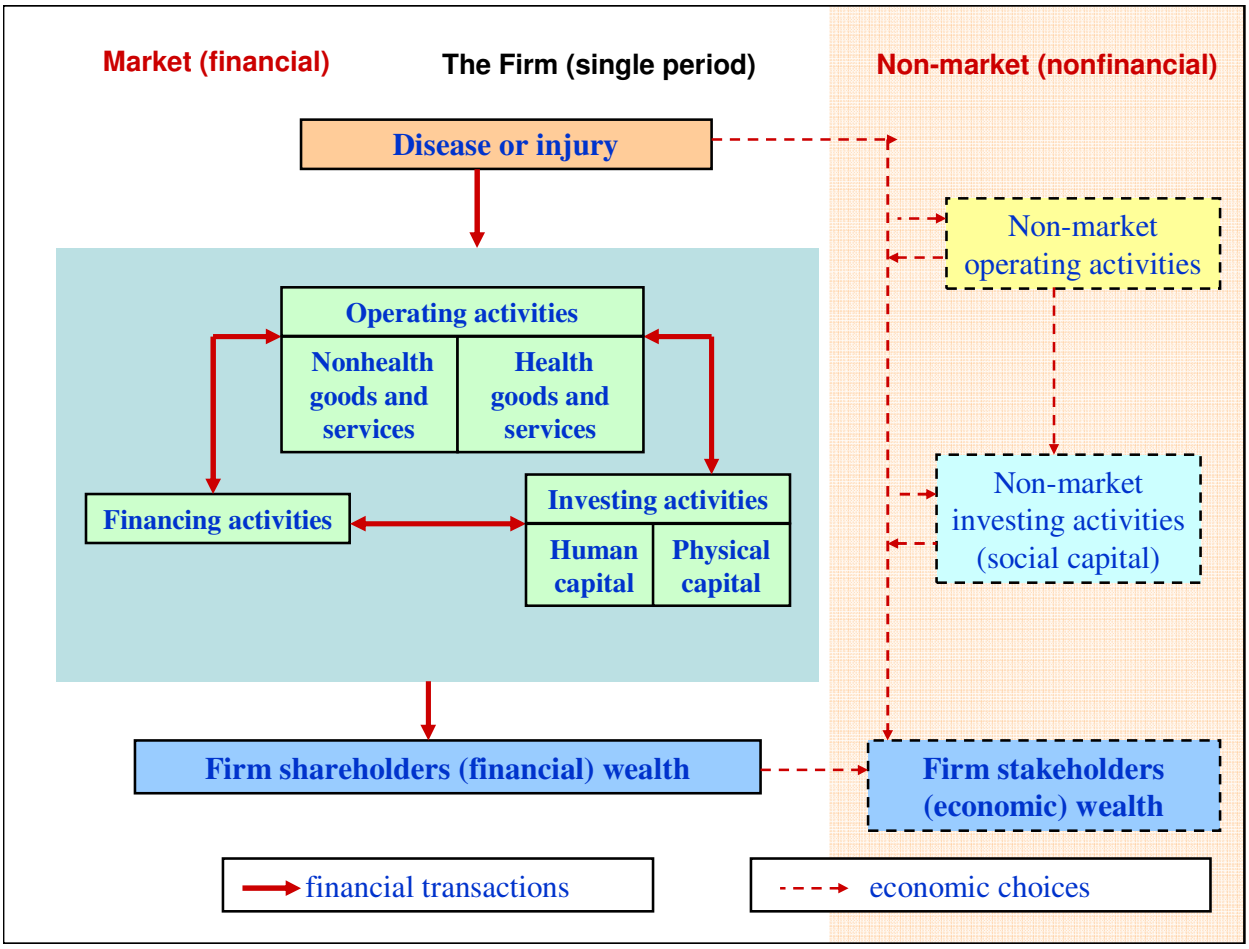
quantity of interest for measuring the financial and economic impacts of adverse health shocks is the lost returns and reduced wealth and, consequently, lower future consumption opportunities. These losses occur when health shocks are sufficiently large to actually raise the direct and indirect costs of the firm and reduce earnings. The reduction of earnings translates into lower future dividends paid and reduced shareholder wealth and consumption. In addition, these losses might force reductions in corporate profit retention and investment activities of firms.

The various effects of health shocks on firms can be illustrated as in **Figure 4**. Health shocks impact the firm through their financial effects on firm operating, investing and financing activities. Also, the figure illustrates non-market (non-financial) impacts of health shocks on other "stakeholders" of the firm (e.g., labour, consumer, and environmental groups).

The primary focus is on the *operating activities* of the firm where the purpose is to produce and market goods and services. This includes the purchase of variable inputs (labour services and labour-related inputs such as health insurance and the provision of health services for workers and other purchased inputs) and various fixed inputs. Variable and fixed health-related expenditures are most directly related to the costs of acquiring and maintaining labour. Operating activities generate net income for the firm which, in turn, is either paid out to shareholders as dividends or retained as shareholder equity capital in the firm.

Investing activities are primarily associated with the purchase and sale of various long-term capital assets (primarily physical capital). Investments in human capital might be related to improving the health of the company's labour force. *Financing activities* are associated with the acquisition of funds to support both operating activities and investing activities of the firm. Financing decisions are those that respond to health shocks and we assume that the financing decision is generally "separable" from the health investment decision. However, under conditions of severe health shocks (covariant health shocks) where the health shock is correlated across many workers and even across firms and their subsidiaries, where the firm's internal financial resources are insufficient and the firm is unable to avoid the health costs, the firm's choices may be restricted by a capital constraint. In this special case the firm's cost of capital may be affected by the financing of the health-related expenditures and the value of the firm may change.

Figure 4 The firm and the impacts of health shocks



While these financial and economic transactions help us to conceptualize the firm-level flows, the impacts might be classified into: out-of-pocket expenditures on health services; productivity losses; the opportunity cost of reduced investment; and the economic value of non-market activities. The first three effects are financial costs and can be expressed in aggregate form (as a present value) in order to determine the overall market-based impacts of health shocks on firms. By adding in the non-market impacts we could approximate the total economic impacts of health shocks.

In formal terms, it is assumed that firms attempt to maximize the present value of shareholder wealth (V), subject to an income constraint and a production technology constraint. For simplicity it is assumed that income is static and that the infinite annuity (R) that

investors/owners receive from the firm is discounted at the fixed market interest rate (i) over an infinite time horizon, thus: Max V = R/i. The firm selects the amount of labour (L) and capital (K) to employ and how much annual earnings it will retain. Thus, ignoring revenue growth, the expected annual net income paid to investors (R) is: $R = pY – wL – rK - \Delta E$, where Y is output, p is its market price, w is the effective wage rate, r is the rate of return on capital, and ΔE is the amount of annual net income that is retained by the firm as equity capital. The production technology constraint is given by $Y = Af(K, hL)$, where output is a function of: A, an efficiency parameter; K, the level of physical capital; L, the quantity of labour; and h, the quality of labour or human capital which might depend on the level of health expenditures (Sala-i-Martin, 2005).

A change in health expenditures is represented by a change in the effective wage rate, holding the level of L constant. Productivity losses due to changes in the efficiency of production may be associated with any of the terms in the production technology function (decreases in h, L, and A). A negative health shock might reduce labour through a deterioration of the quality of each unit of labour or human capital (h) or a decline in the quantity of available labour (L). If there is a decline in overall productive efficiency of the firm, there would also be a decrease in the efficiency parameter. Thus, through these various channels a health shock could lead to a decline in both the level of output and the rate of growth of firm output, and an increase in total costs.

In addition to the health-related out-of-pocket costs and productivity losses, a firm will incur an opportunity cost in the form of reduced productive investment elsewhere in the business. An opportunity cost (loss of a benefit) is incurred if the firm is forced to forego an investment in profitable new projects as a consequence of making a health project investment due to an adverse health shock. In this framework an increase in opportunity cost due to a health shock would be represented by a change in the required rate of return on capital (r) and potentially the rate (i) at which investors discount future dividend income. Due to greater uncertainty imposed by the health shock in the form of a risk premium, the firm would need to apply a higher discount rate to all future health and non-health investment projects and investors would also require a higher rate of return.

4.2.2. Expenditures on health by firms

Due to the negative effects on productivity resulting from disease or injury, firms may choose to protect the health and safety of their workers through investments in physical and human capital. Thus, health-related out-of-pocket expenditures of firms can be an important component of the overall financial and economic costs of disease and injury. In certain sectors (e.g., manufacturing and construction) companies may spend a proportion of their earnings to invest in safer technologies and equipment, thereby enhancing the quality of the work environment. Such investment may not only increase a firm's productive capacity, it may also reduce accidents and subsequent health care expenditures by the firm.

Also, private firms often provide a range of health-related benefits for their employees (and families) in the form of health insurance, sickness benefits, payment towards the cost of medical care, in-company healthcare providers, and contributions to pension funds. These benefit packages are paid for out of corporate revenues and they have the potential to significantly reduce a firm's profits.[8] As with household-level health insurance costs, these employer expenditures may be seen as a general cost of illness - in the sense that if there were no illness, there would be no need to pay for health insurance - rather than as a condition-specific cost. To ascertain the specific costs attributable to a particular disease or injury category one would need detailed firm-level medical expenditure records and an apportionment of insurance and other benefits according to the risk probabilities of disease or injury.[9]

Although firm-level surveys provide an obvious source of information relating to these health-related expenditures, a number of potential problems should be noted. While there are some health-related costs that are fixed (do not vary with the incidence or prevalence of disease or injury), if these are sunk costs they are not included. Only health-related expenditures that vary with the level of incidence or prevalence are included. Some health-related expenditures may be

[8] General Motors spent over $5.5 billion in 2005 on health benefits for its more than one million workers, retirees and family members (equivalent to $1,500 in employee medical expenses for each new car it sells), an important contributor to corporate losses for that year. Source: http://www.washingtonpost.com/wp-srv/nation/documents/wagoner_feb_10.pdf

[9] For example, the economic cost of obesity to U. S. business has been estimated to be $12.7 billion (in 1994 prices) using standard epidemiological methods for risk attribution (Thompson et al., 1998).

associated with co-morbidities, and they may need to be allocated to specific disease or injury events. For example, a firm may protect the health and safety of its workers through investments in physical and human capital. Some of these are sunk costs, and some may be fixed costs. Also, a firm may provide health-related benefits for its employees (and their families) in the form of health insurance, sickness benefits, payment towards the cost of medical care, in-company healthcare providers, and contributions to pension funds. These benefit packages may or may not vary with the level of incidence or prevalence of an specific disease or injury. The fundamental principle is that to be included the costs should be incremental and, therefore, due to the increased prevalence or incidence of disease or injury.

4.2.3. Labour and productivity losses

In line with measurement and valuation of production losses at the household level, we advocate the use of an output-based approach to lost production at the level of the firm. The output approach is designed to capture the change in labour productivity as part of the change in earned income of the firm, and can address several problems that arise when employing alternative methods of valuation. The output approach recognizes that the decline in productivity due to a health shock is equivalent to an increase in inefficiency.

The impaired health of workers (reduced quantity and quality of labour) reduces the productivity of labour and may reduce the productivity of other firm assets, specifically those which are related to the use of labour through the underlying technology of production. Thus, there is the potential for induced *technical inefficiency* where the firm operates inside the efficiency frontier and consequently output from a given input mix is less than the maximum possible level.

At the firm-level there may also exist *allocative inefficiency*, when the input mix is not consistent with cost minimization for a given level of output. This would be the case if labour availability is reduced due to a large health shock, causing the firm to use labour and non-labour inputs in a sub-optimal way. Whether allocative inefficiency occurs may depend on whether the relative prices of labour and non-labour inputs change due to the health shock. Thus, induced technical inefficiency would result in a decrease in the level of firm output (when compared to the

counterfactual of no health shock). However, allocative inefficiency may or may not result in reduced firm output. Appendix E describes and illustrates these efficiency concepts and corresponding productivity losses in more detail.

In terms of how to actually measure the productivity losses that are attributable to the disease/injury event, a study by Audibert et al (2003) provides an example of how changes in farm-level efficiency due to malaria can be estimated using non-parametric analysis. They report that an increasing severity of malaria infection results in a significant reduction of technical efficiency. However, this is a data-intensive methodology and is unlikely to be practical for general use across different diseases and when firms differ in their production technologies. The complexity involved in estimating productivity losses increases when we consider the range of possible cost impacts at the individual and organizational levels of a firm, together with the evolution of these costs over time. Firms also incur costs associated with recruiting workers who are replacing (permanently or temporarily) sick or deceased employees.

In an output-based approach, these costs can be calculated by multiplying the number of workers leaving the firm by the average administrative cost of recruitment. Also, firms may employ different strategies to control the losses in productivity such as: trying to prevent new cases, avoiding or reducing the costs associated with existing and future cases, and/or providing health treatment and support for affected employees to extend their productive working lives and, thus, postponing the costs (Rosen et al., 2000). Therefore, alternative survey-based methods need to be developed to better elicit output-based estimates of lost productivity due to disease/injury at the firm level.

4.2.4. Opportunity cost of reduced investment

Actions taken by the firm in anticipation of, or in response to, a health shock could have long-term financial consequences for earnings and investor consumption possibilities. In addition to the health-related out-of-pocket costs and productivity losses, a firm will incur an opportunity cost in the form of reduced productive investment elsewhere in the business, given that it cannot undertake all of its investment projects. An opportunity cost (loss of a benefit) is incurred if the

firm is forced to forego an investment in profitable new projects (e.g. revenue-enhancing projects or efficiency-improving technologies) as a consequence of being forced to make a health-related investment due to a health shock. The opportunity cost is associated with the return that would be realized in the next best alternative use of funds. We suggest that this opportunity cost will increase the firm's cost of capital and, thus, reduce the expected present value of the projects that it is evaluating when a health shock occurs.

4.2.5. Non-market impacts

Non-market costs are those that would be incurred by the firm if a significant adverse health shock were to occur and the firm's non-market operating and investment activities were impaired. These are not financial costs, yet represent - in theory at least - part of the economic impact of disease or injury at the firm level. Two such non-market costs might include the disruption of morale and discipline of other employees and the deterioration of labour relations. Berger et al (2001) also indicate that consideration of these economic impacts of disease on firms is necessary if a broader economic welfare framework is used. They suggest that since productivity is determined in part by good health, these considerations might extend to measurement by firms of the value that their employees place on health in the workplace, or ascertainment of the amount firms are willing to pay for what employees and other stakeholders value. Since there is no market for these workplace attributes, it is difficult to establish a price for them and, obviously, there is no observed transaction. These effects clearly fall within the non-market operating activities of the firm.

A second category of these non-market costs relates to the social networks or capital that firms establish with their stakeholders (e.g., consumer and environmental groups). Although either of the above two categories of non-market costs are theoretically valid, current shortcomings in the willingness to pay approach jeopardize such valuation efforts and we recommend that if this approach is attempted to deal with valuation of these non-market effects, the estimates should be kept separate from the valuation of the financial costs to the firm.

4.3. Microeconomic impact of disease and injury at the governmental level

A further microeconomic agent that is potentially affected by the incidence of disease or injury is the government. Governments are producers of goods and services in the same way as a firm - though mostly governments have focused on public goods. Government hire labour from households and, along with capital, transform this into goods and services that are either provided or sold to households and firms. Their other role has been to redistribute income through taxes and subsidies. The role of government at the macroeconomic level was discussed in Section 3. From a microeconomic perspective, it is worth considering what types of policy questions can be addressed by studies of the economic burden on households. As described in **Box 1**, we see two specific policy questions:

- What proportion of government expenditure could have been saved and directed to an alternative use in the absence of illness?

- What impact does ill-health have on the government workforce and on the government's ability to provide services?

The first of these - estimation of public sector care and prevention expenditures attributable to a particular disease - have typically made up a large proportion of the 'direct cost' component of cost-of-illness studies, while the second - estimation of lost productivity in public agencies - represents a part of the 'indirect cost' component of such studies.

When considering these costs from the microeconomic perspective, what might represent the underlying quantity of interest? Consistent with the general framework outlined earlier, and assuming a fixed budget, government expenditure on health services and goods reduces spending opportunities for other services and goods, such as education or transport, so the underlying quantity of interest can readily be understood in terms of the opportunity cost of reduced non-health consumption. As a large-scale employer, public agencies face the same prospect of production losses from ill-health as private corporations do, but unlike the private sector, economic losses cannot generally be gauged in terms of profit maximization or market share

(although this might apply to state-owned enterprises). Rather, the underlying economic objectives of most public agencies revolve around regulation or governance and the delivery of public services, so it is reductions in the output of these functions that are of concern. As pointed out above, though, government is also concerned with the redistribution of resources (equity).

4.3.1. Expenditures on health by government

Public expenditure is the value of goods and services bought by the State, including any provision of health services and goods, so at least within a fixed budget scenario any increase in health spending (as a result of a new disease outbreak, for example) suggests a reduction in current or future spending opportunities on non-health services or goods. Accordingly, government expenditures on a particular disease or on health generally have a clear opportunity cost with respect to non-health consumption possibilities. In this respect, there is little methodological controversy over the measurement of these expenditures, although care must obviously be taken in the appropriate collection and analysis of data in order to ensure comprehensive and consistent cost estimation.

Estimation of these expenditures is facilitated by the availability of an internationally standardized system of health accounts (see www.who.int/nha), which attempt to classify and record not only the outputs of a health system - with reference to source (e.g. central government, private households), function (e.g. nursing care, health promotion) and provider (e.g. hospital or ambulatory care services) - but also their respective inputs (e.g. wages, supplies, building costs). Such information is currently of most use where the focus is on ill-health generally, since disease-specific estimates are often not available. However, systems of sub-accounts for some diseases and in some countries are now available or being developed, which will provide a core information resource for financial impact studies of a particular disease entity.

For identification and measurement of government health expenditures attributable to a disease, it is important to include not only the main categories of health services and goods consumed by patients (including inpatient and outpatient hospital care, primary health care, ancillary care, medical equipment, devices and consumables, diagnostic tests, prescription drugs etc.), but also

non-patient cost components, which range from the planning and administration of health programmes, to training and health education, and to health prevention and promotion activities (although one could argue that the latter might continue to be incurred even in the counterfactual situation of no disease). Such costs are regularly overlooked, most often because of measurement difficulties with respect to attribution of joint or overhead costs, but potentially constitute a substantial component of total government expenditure for a given health condition. Developments have recently made in the appropriate measurement of health programme-level costs as part of the WHO's CHOICE project (Johns et al., 2003).

The range of government expenditures for a particular disease or injury may also extend beyond the conventionally agreed boundaries of the health system *per se*, to include related welfare costs such as social services (e.g. for elderly people disabled by disease), education (e.g. for 'special needs' children) or criminal justice services (e.g. for people with substance use disorders). Similar procedures to health service costing apply to their appropriate measurement, which are based on estimation of the number of persons with the disease who use the service, the intensity of service utilization over a defined period of time and the cost per unit of service use. It is nevertheless acknowledged that certain activities pose particular challenges to costing due to attribution problems, such as establishment of the time and cost of a police arrest or judicial court appearance. Existing guidelines provide recommendations concerning the likely set of welfare services that may be included in economic impact studies of mental disorders, substance abuse or violence, as well as suggestions for their valuation (WHO, 2003).

As mentioned above in Section 2, it is important to take into account the future expected stream of government expenditures on health and related welfare services in order to avoid over-estimation of the economic losses due to a disease. Specifically, the present value of resources needed in the future to treat *additional* morbidity and mortality relating to the disease in question (due to the increased rate of 'survival' in diseased persons) needs to be deducted from the savings in health intervention costs in the index year.

A final category of government expenditure relates to social security payments, which are a large and growing proportion of total government expenditure in many countries. Generally, social

security expenditures attributable to disease or injury are considered as transfer payments in the economy (from one group in society made up of taxpayers, to another group made up of state benefit recipients), and are not included as a societal cost of disease since no economic loss is actually recorded. From the specific perspective of the government, however, there may be an interest in determining the extent of social security payments made in relation to a disease, in which case the opportunity cost of these payments can be seen in terms of the forgone opportunity to direct public resources to alternative uses.

4.3.2. Labour and productivity losses

Disease and injury may also reduce the productivity and efficiency of the government workforce. As with firms, these effects can be picked up via changes in consumption flows at the aggregate macroeconomic level of society itself where the interest is in establishing the overall societal impact of disease. Equally, however, the interest may be confined to the particular impact of one or more disease entities on the productive capacity of the government workforce. Where this is the case, the limitations that we already described in relation to an input- or wage-based approach to measurement of these production losses hold true; that is to say, it is not safe to assume that the imputed lost earnings of a government worker due to ill-health can be taken as a good indication of the actual loss in production, whether that be measured in terms of patients consulted, classes taught or reports compiled. This is because, just like households or firms, public agencies have in place a number of compensating mechanisms which restrict the disruption to the normal expected flow of output, such as a doctor or teacher covering for a sick colleague. Accordingly, where there is interest in documenting the economic value of losses in government productivity, our recommended strategy would be to pursue an output-based approach which seeks to compare execution of a public agency's core function(s) with and without disease. In this case, however, and unlike households, the probability is that there will be insufficient units of analysis to allow for a controlled comparison, so this would imply the need for an alternative analytical design, such as comparison of work teams with/without the index health condition within the same public agency, or 'before and after' comparisons that adjust for baseline productivity levels.

4.4. Microeconomic impact: key points

Households

- The economic impact of disease or injury at the household level should be measured and valued with reference to overall changes in welfare or its essential components - the consumption of non-health goods and services, leisure and health status.

- The trade-offs that households make between consumption in current and future time periods, and in their time allocations to production, health improvement and leisure should reflect both *time preferences* (consumption today versus consumption next period) and *risk preferences* (certain consumption versus uncertain consumption).

- Use of an input-based approach to the measurement and valuation of lost production - which assumes that the duration of an individual's absence from work fully corresponds to the market value of those lost days - is likely to overestimate economic losses, because it overlooks the so-called 'coping strategies' used by households to mitigate the adverse circumstances of ill-health. An output-based approach that measures actual (rather than potential) net losses in income or production provides a more robust basis for estimation.

- Valuation of the full economic welfare impact of illness on households - including the value of leisure and health itself - requires the application of willingness-to-pay measures. Since there remains considerable uncertainty around the appropriate social value to apply to these health-related welfare losses, we recommend separate reporting of market and non-market losses in economic impact studies and advise caution in the use and interpretation of WTP-based measures.

Firms

- The economic quantity of interest for measuring the financial and economic impacts of health shocks on firms is lost returns and reduced wealth and consequently lower

consumption opportunities of investors. Losses occur when health shocks are sufficiently large to actually raise the direct and indirect costs of the firm and reduce earnings. The reduction of earnings translates into lower future dividends paid and reduced shareholder wealth and consumption. Losses might also force reductions in profit retention and investment activities.

- An incidence-based approach that uses a standard discounted cash flow structure provides an appropriate measurement framework for estimating the financial impact of disease or injury on the firm.

- We advocate the use of an output-based approach to lost production at the level of the firm, which is able to capture the change in labour productivity as part of the change in earned income of the firm, and can be used to show how a decline in productivity due to a health shock is equivalent to an increase in inefficiency.

Government

- The two main impacts of disease or injury of interest relate to the provision of health and related services (at the expense of non-health services and goods) and the loss of government workforce productivity.

- Social security expenditures attributable to disease or injury are considered as transfer payments (and therefore not included as a societal cost of disease since no economic loss is actually recorded). From the specific perspective of the government, however, there may be an interest in determining the extent of social security payments made in relation to a disease, in which case the opportunity cost of these payments can be seen in terms of the forgone opportunity to direct public resources to alternative uses.

5. CONCLUSION

In view of the multifaceted economic consequences of ill-health and an enduring interest in measuring these effects, this guide set out to provide policymakers and researchers with an accessible review of - and primer for - methodological issues and current practice relating to economic impact studies in health. The motivation behind the production of such a guide was a concern around the methodological integrity and consequent policy value of many economic impact studies in the past and a desire to promote greater rigor and consistency in the future.

In order to place the future conduct of economic impact studies on firmer theoretical foundations, we emphasized in Section 2 the need to address a number of critical conceptual issues or questions, which we anticipate will encourage analysts to be more precise and insightful about what an economic impact study aims to do and what it can achieve. We repeat these again here (**Box 3**) in a bid to reinforce their importance, and urge their repeated use to anyone planning or indeed evaluating an economic impact study in health. One further question not listed below but certainly worth asking is whether an economic impact study is really needed in the first place.

Box 3 Key questions for planning or evaluating an economic impact study in health

- *What is the **perspective** or level of aggregation for the study?* (e.g. microeconomic level of households/firms/government, or aggregate impact at macroeconomic/societal level?)
- *What is the **scope** of the study?* (e.g. overall economic welfare or one/some of its constituent elements, defined in this guide as the consumption of non-health goods and services, leisure and health itself?)
- *What is the defined **quantity of interest** for the study?* (e.g. if non-market losses such as unpaid care-giving by family members at the household level are measured, is it meaningful to combine these with market losses elements into a single estimate of economic loss?)
- *What is the **counterfactual** / comparator situation against which economic losses are to be assessed?* (e.g. prevalence-based or incidence-based approach?)
- *What if any account is being taken of **dynamic effects** that may occur beyond the current period?* (e.g. the impact of depleted capital accumulation on future economic growth)

One important conclusion arising from discussion of these conceptual issues is that a number of legitimate questions, quantities of interest and counterfactuals can be specified when undertaking an economic impact study. So although a prescriptive set of recommendations would have the attraction of promoting greater comparability between studies, there is in fact no single perspective or measurement approach that trumps all others - because the appropriate choices will depend on the nature of the policy question being addressed. This will also need to take into account the availability of resources, time constraints, data and technical expertise. Instead, we advocate the adoption of a menu-based or algorithmic set of considerations, which between them will determine the analytical approach that is ultimately selected. We present such an algorithm below in an attempt to help analysts make these choices in a series of logical steps (**Figure 5**).

In the following two sections of the guide, we then considered the various channels via which adverse economic effects might be transmitted, and how their impacts might be measured and valued at the macroeconomic or microeconomic levels. By applying the conceptual questions above to the existing body of literature, a better understanding of the merits and shortcomings of different measurement approaches has emerged; these are discussed in detail in the relevant sections of Section 3 (see also Appendix C), and briefly synthesized again in **Box 4**. In short, we conclude from this methodological review that:

- A conventional cost-of-illness approach offers an inadequate and ill-defined representation of the macroeconomic consequences of disease and injury;
- CGE models offer the most complete assessment of market-based disease consequences, but would only be practicable to construct/apply in special cases;
- Of the remaining two (market-based) approaches identified, regression-based models can produce valid if small estimates of the impact of disease on economic growth and are therefore likely to be of most relevance to measuring large health shocks (where there is sufficient data), while calibration-based simulation models offer a tool that is pragmatic and flexible but depends more on assumptions and extrapolated data.
- Only one other approach - the 'full-income' model - goes beyond market-based losses, but in so doing introduces considerable uncertainty into the valuation process;

Box 4 Summary of measurement approaches used to assess the societal impact of ill-health

Cost-of-illness: A large proportion of economic studies in health completed to date employ some variant of the 'cost-of-illness' methodology (which combines the 'direct' costs of medical care, travel costs etc. with the 'indirect' cost of lost production because of reduced working time). Our primary concerns with this widely-used approach are: 1) no consistent meaning can be given to the aggregate losses that are conventionally included (undefined quantity of interest); 2) no consistent meaning can be given to the counterfactual that is conventionally adopted; 3) no consideration is given to dynamic effects such as the contribution of depleted capital accumulation or human capital investment to diminished economic growth; and 4) reliance on a 'human capital' approach to valuing production losses due to ill-health or premature death is unrealistic in most settings (where a pool of underemployed or unemployed labour exists). Although computationally the most straightforward approach (a probable reason for its popularity), the critical conceptual shortcomings listed above mean that we conclude that it is an inadequate model for capturing the economic impact of disease or injury at the societal level, and do not recommend its use for that purpose. However, elements of the cost-of-illness calculus - in particular, those relating to direct costs - could still be used to address specific questions, such as the level of health consumption or expenditure for a given disease entity or injury category (not only at the aggregate level of the market economy but also at the microeconomic level of households, firms or government).

Regression-based estimation models: Most economic growth studies to date have used regression analysis to estimate whether cross-country differences in health indicators such as overall life expectancy or disease-specific mortality have an impact on economic growth or GDP (the key determinants of which are taken to be human capital plus physical capital and labour). Resulting estimates have tended to show that ill-health only has a modest impact on economic growth. A prominent concern with such studies relates to how well they can accurately attribute changes in wealth to health (as opposed to the other way round).

Simulation-based calibration models: Rather than directly estimating the relationship between health status and economic growth (via regression), calibration models use previously established values for key model parameters - such as the supply of labour or the aggregate savings rate - to simulate how changes in life expectancy or mortality rates affect these parameters, and thereby economic growth. A limitation is that parameter values often have to be taken from different settings or contexts, although this can be seen as an advantage for countries with limited data availability.

Computable general equilibrium models: As the name suggests, CGE models depart from the more commonly-used partial equilibrium models and set out to show the economic impact of health shocks across all sectors of the market economy. These models offer the most complete assessment of market-based consequences and would be particularly suitable for assessing the impact of health conditions with multi-sectoral impacts (e.g. avian flu, SARS, HIV). The main drawback is that CGE models are complicated and costly to build. They also do not attempt to capture any non-market effects such as the value of lost household production or diminished health itself.

Full-income models: By attaching the estimated value of a statistical life (VSL) to years lost to disease or injury, this approach evidently goes beyond purely market-based losses. However, the conceptual basis for calculating VSL remains contested, and despite its name, the 'full-income' approach still represents only a partial estimate of total lost economic welfare due to disease or injury. Accordingly, caution should be exercised when employing willingness-to-pay (WTP)-based techniques such as VSL, and resulting estimates of economic welfare loss should be reported separately from estimates of market losses.

Concerning the microeconomic impact of disease or injury, we have not so much tried to distinguish between a series of competing analytical *models* as recommend a number of analytical *principles* that should guide the measurement and valuation process. In broad terms, however, we recommend the adoption of dynamic modeling approaches that take account of inter-temporal choice (i.e. the trade-offs that households or firms make between consumption now versus the future), and also risk preferences (i.e. how households or firms react differentially to a health shock). We anticipate the need for more developed or formalized models of inter-temporal choice, both at the household and firm level.

A common thread underpinning our proposed analytical principles is an interest in shifting away from imputation- or input-based estimation towards the assessment of actual net losses. Studies which take the duration of illness (in days) and simply attach an average wage rate, for example, are highly likely to be overstating the true economic impact of lost market production. Accordingly throughout Section 4, we advocate the pursuit of an output-based approach towards measurement and valuation, and propose the inclusion of controls wherever feasible as one means of adjusting for the coping strategies that households or firms make when confronted with sickness, serious injury or death. Some of the proposed principles, and how they differ from those employed in many previous studies, are summarized below in **Table 1**.

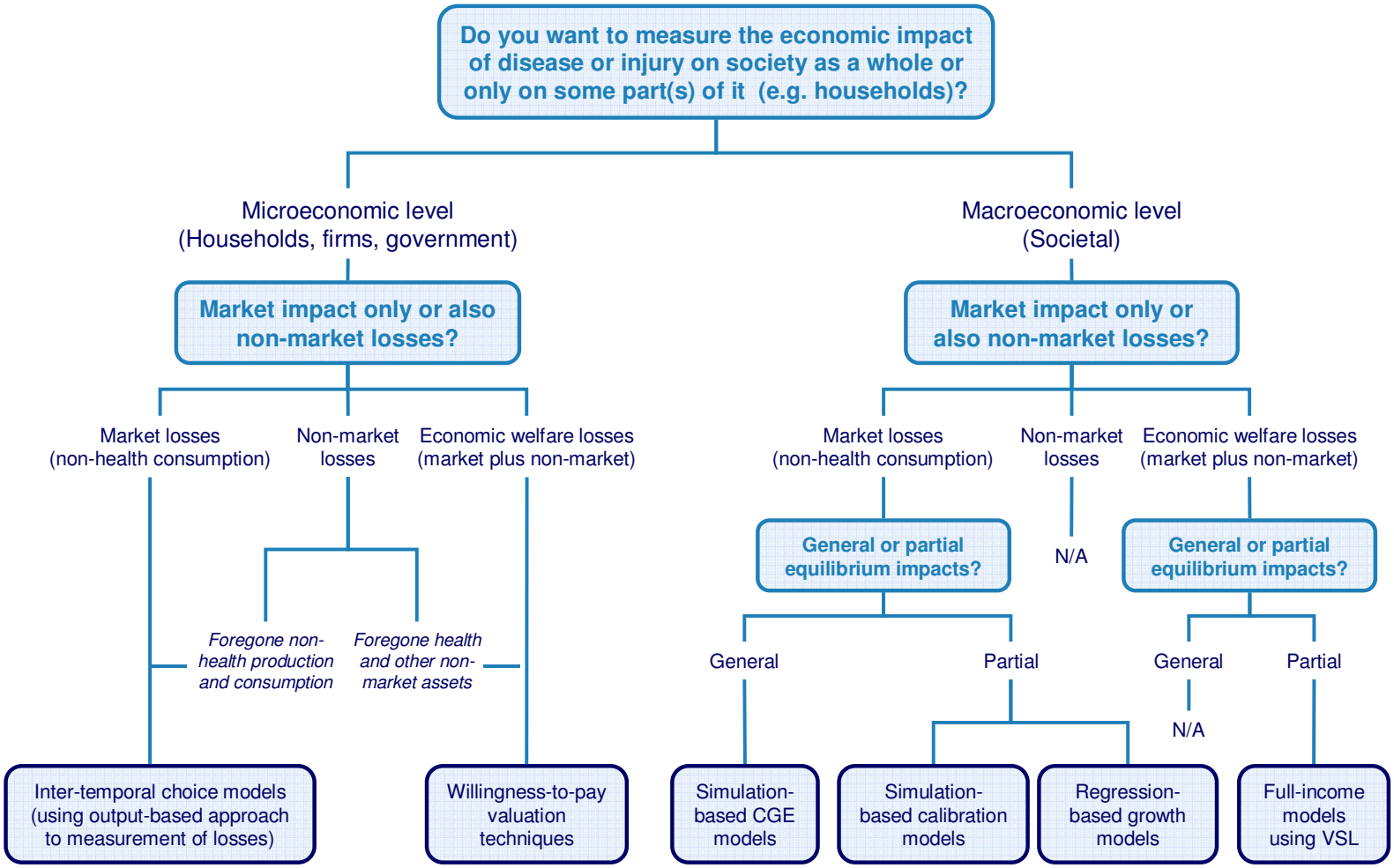
Table 1 Measurement of production losses at the microeconomic level

Measurement issue	Common problems	Proposed solutions
Production losses (market and non-market)	Input-based approach (time * wage)	Output-based approach
	Imputed estimate of potential losses (unrealistic)	Observed estimate of actual losses (realistic)
Controls	None (i.e. gross impact)	Include (i.e. net impact)
Coping strategies	Not accounted for	Incorporate (current and future impacts)

Presenting all analytical approaches together into a simplified framework - as **Figure 5** tries to do - runs the risk of masking or missing other defining characteristics of economic impact studies in health, but captures what we consider to be the core features that differentiate study types. It is hoped that by following this logical sequence of steps, analysts will be clear as to what they can (and cannot) expect or achieve when looking into the economic consequences of disease and injury.

We finish by recalling that whatever analytical approach is selected, the primary purpose of economic burden studies is to shed light on the magnitude and distribution of costs falling on different agencies in society as a result of diminished health status, rather than to deal with the question of what to do about these attributable costs. Ultimately we regard this as a more important question for health policy makers and analysts to address. For these resource allocation questions, additional information is required concerning the efficiency or cost-effectiveness of a range of possible intervention strategies capable of making a meaningful dent in the current or projected burden of disease or injury.

Figure 5 Algorithm for determining what methodological approach to use for economic impact studies in health



Abbreviations used: N/A: Not available; CGE: Computable general equilibrium; VSL: Value of statistical life

REFERENCES

1. Abegunde D, Mathers C, Adam T, Ortegon M and Strong K (2007). The burden and costs of chronic diseases in low-income and middle-income countries. *The Lancet*, 370(9603): 1929-1938.
2. Abegunde D and Stanciole A (2008). The economic impact of chronic diseases: How do households respond to shocks? Evidence from Russia. *Social Science and Medicine*, 66: 2296-2307.
3. Abraham KG, Mackie C (2005). *Beyond The Market: Designing Non-market Accounts For The United States*. Washington, DC, USA: National Academy Press.
4. Acemoglu D and Johnson S (2006). Disease and development: The effect of life expectancy on economic growth. *NBER Working Paper 12269*.
5. Agénor P and Montiel P (1996). *Development Macroeconomics*. Princeton, NJ: Princeton University Press.
6. Agénor P and Moreno-Dodson B (2006). Public infrastructure and growth: New channels and policy implications. *World Bank Policy Research Working Paper 4064*.
7. Aghion P and Durlauf S (2008). Growth regressions and policy evaluation. *Background paper for the Commission on Growth and Development*. Washington, DC, USA: World Bank.
8. Ahlburg D (2000). The Economic Impacts of Tuberculosis. *The Stop TB Initiative 2000 series*. Geneva: WHO.
9. Aísa R and Pueyo F (2004). Endogenous longevity, health and economic growth: A slow growth for a longer life? *Economics Bulletin*, 9: 1-10.
10. Alderman H, Behrman JR, Hoddinott J (2003). Health, human capital and economic growth. In López-Casasnovas G, Rivera B and Currais L (eds.) *Health and Economic Growth: Findings and Policy Implications*. Cambridge, MA, USA: MIT Press.
11. Alsan M, Bloom D and Canning D (2006). The effect of population health on foreign direct investment inflows to low- and middle-income countries. *World Development*, 34: 613-630.
12. Arndt C and Lewis J (2000). The macro implications of HIV/AIDS in South Africa: A preliminary assessment. *South African Journal of Economics*, 68: 380-392.
13. Artadi E and Sala-I-Martin X (2003). The Economic Tragedy of the XXth Century: Growth in Africa. NBER Working Paper 9865, National Bureau of Economic Research. Cambridge, Massachusetts; USA.
14. Arthur W (1981). The economics of risks to life. *American Economic Review*, 71: 54-64.
15. Ashraf Q, Lester A and Weil D (2008). When does improving health raise GDP?. *Department of Economics Working Paper 2008-7*. Providence, RI: Brown University.

16. Atkinson AB (1975). *The Economics of Inequality*. London, UK: Oxford University Press.

17. Attanayake N, Fox-Rushby J, Mills A (2000). Household cost of 'malaria' morbidity: a study in Matale district, Sri Lanka. *Tropical Medicine and International Health*, 5: 595-606.

18. Audibert M, Mathonnat, Henry M (2003). Social and health determinants of the efficiency of cotton farmers in northern Cote d'Ivoire. *Social Science and Medicine,* 56: 1705-1717.

19. Bardhan P and Udry C (1999). *Development microeconomics*. New York, NY, USA: Oxford University Press.

20. Barendregt JJ, Bonneux L, van der Maas PJ (1997).The health care costs of smoking. New England Journal of Medicine, 337: 1052-7.

21. Barro R, Sala-I-Martin X (1995). *Economic Growth*. New York, NY, USA: McGraw-Hill, Inc.

22. Barro RJ (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106: 407-443.

23. Becker G (1965). A theory of the allocation of time. *The Economic Journal*, 75: 493-517.

24. Becker G, Philipson T, Soares R (2005). The quantity and quality of life and the evolution of world inequality. *American Economic Review*, 95: 277-291.

25. Bell C, Shantayanan D, Gersbach H (2003). *The Long-Run Economic Costs of AIDS: Theory and an Application to South Africa*. World Bank, Washington D.C. and Süd-Asien Institut, Heidelberg.

26. Berger M, Murray J, Xu J and Pauly M (2001). Alternative valuations of work loss and productivity. *Journal of Occupational and Environmental Medicine*, 43:18-24.

27. Bergson A (1938). A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics*, 52: 310-34.

28. Beutels P, Edmunds WJ and Smith R (2008). Partially wrong? Partial equilibrium and the economic analysis of public health emergencies of international concern. *Health Economics*, 17: 1317-22.

29. Berk A, Paringer L, Mushkin SJ (1978). The economic cost of illness: fiscal 1975. *Medical Care*, 16: 785–790.

30. Bhargava A, Jamison D, Lau L, Murray C (2001). Modeling the effects of health on economic growth. *Journal of Health Economics*, 20: 423-440.

31. Bloom D and Canning D (2000). The health and wealth of nations. *Science*, 287(5456): 1207-1209.

32. Bloom D and Canning D (2008). Population health and economic growth. *Background paper for the Commission on Growth and Development*. Washington, DC, USA: World Bank.

33. Bloom D and Malaney P (1998). Macroeconomic consequences of the Russian mortality crisis. *World Development*, 26: 2073-2085.

34. Bloom D and Williamson J (1998). Demographic transitions and economic miracles in emerging Asia. *World Bank Economic Review*, 12: 419–455.

35. Bloom D, Canning D and Graham B (2003). Longevity and Life-cycle Savings. *Scandinavian Journal of Economics*, 105: 319-338.

36. Bloom D, Canning D and Sevilla J (2004). The effect of health on economic growth: A production function approach. *World Development*, 32: 1-13.

37. Bloom D, Canning D, Hu L, Liu Y, Mahal A and Yip W (2007). The contribution of population health and demographic change to economic growth in China and India. *Program on the Global Demography of Aging Working Paper 28*, Harvard Initiative for Global Health.

38. Bloom D, Canning D, Jamison DT (2004). Health, wealth and welfare. *Finance and Development*, 41: 10-15.

39. Bonnel R (2000). HIV/AIDS and economic growth: a global perspective. *South African Journal of Economics*, 68: 820-55.

40. Borghouts JA, Koes BW, Vondeling H, Bouter LM (1999). Cost-of-illness of neck pain in The Netherlands in 1996. *Pain*, 80: 629-36.

41. Brouwer W, Koopmanschap M, Rutten F (1997). Productivity costs measurement through quality of life? A response to the recommendation of the Washington panel. *Health Economics*, 6: 253-259.

42. Brouwer W, Koopmanschap M, Rutten F (1999). Productivity losses without absence: measurement validation and empirical evidence. *Health Policy*, 48: 13-27.

43. Castelló-Climent A and Doménech R (2008). Human capital inequality, life expectancy and economic growth. *The Economic Journal*, 118(528): 653-677.

44. Chadeau A (1992). *What is households' non-market production worth?* OECD Economic studies No. 18: OECD; Paris, France.

45. Chakraborty S (2004). Endogenous lifetime and economic growth. *Journal of Economic Theory*, 116: 119-137

46. Chima R, Goodman C, Mills A (2003). The economic impact of malaria in Africa: a critical review of the evidence. *Health Policy*, 63: 17-36.

47. Collins D, Lapsley H, Brochu S, Easton B, Perez-Gomez A, Rehm J, Single E (2006). *International guidelines for the estimation of the avoidable costs of substance abuse*. Ottawa, Canada: Health Canada.

48. Cooper BS, Rice DP (1976). The economic cost of illness revisited. *Social Security Bulletin*, 39: 21-36.

49. Cuddington J (1993). Modeling the macroeconomic effects of AIDS, with an application to Tanzania. *The World Bank Economic Review*, 7: 173-189.

50. Cuddington J and Hancock J (1994). Assessing the impact of AIDS on the growth path of the Malawian economy. *Journal of Development Economics*, 43: 363-368.

51. Cutler D, Richardson E (1997). Measuring the health of the US population. *Brookings Papers on Economic Activity, Microeconomics*, 29: 519–39.

52. Davey PJ, Leeder SR (1993). The cost of cost of illness studies. *Medical Journal of Australia*, 158: 583–584.

53. Deaton A (1992). *Understanding Consumption*. Oxford, UK: Oxford University Press.

54. Dolan P (1998). The measurement of individual utility and social welfare. *Journal of Health Economics*, 17: 39–52.

55. Drummond M (1992). Cost-of-illness studies: a major headache? *Pharmacoeconomics,* 2: 1-4.

56. Drummond ME, Sculpher MJ, Torrance GW, et al (2005). *Methods for the Economic Evaluation of Health Care Programmes* (3rd edtition). Oxford, UK: Oxford University Press.

57. Epstein LG, Zin SE (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework." *Econometrica*, 57: 937-969.

58. Epstein LG, Zin SE (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: an empirical analysis. *Journal of Political Economy*, 99: 263-286.

59. Ettaro L, Songer TJ, Zhang P, Engelgau MM (2004). Cost-of-illness studies in diabetes mellitus. *Pharmacoeconomics*, 22: 149-164.

60. Finley J (2006). Endogenous longevity and economic growth. *Program on the Global Demography of Aging Working Paper 7*, Harvard Initiative for Global Health.

61. Fogel R (1994). Economic growth, population theory, and physiology: The bearing of long-term processes on the making of economic policy. *American Economic Review*, 84: 369-395.

62. Gallup JL, Sachs JD (2001). The economic burden of malaria. *American Journal of Tropical Medicine and Hygiene*, 64: 85-96.

63. Gallup J, Sachs J and Mellinger A (1999). Geography and Economic Development. In: Plsekovic, Boris and Joseph Stiglitz (eds.). *Annual World Bank Conference on Development Economics*, 1998/1999. Washington D.C., World Bank, pp.127-78.

64. Gertler P and Gruber J (2002). Insuring consumption against illness. *American Economic Review*, 92: 51-70.

65. Gertler P, Levine D and Moretti E (2008). Do microfinance programs help families insure consumption against illness?. *Health Economics*, forthcoming.

66. Gold MR, Siegel JE, Russel LB, et al (1996). *Cost-effectiveness in health and medicine*. New York, NY, USA: Oxford University Press.

67. Goldschmidt-Clermont L (1987). *Economic evaluation of unpaid household work: Africa, Asia, Latin America and Oceania*. Women, Work and Development, no. 14. Geneva, Switzerland: International Labour Organization.

68. Greener R (2002). AIDS and the macroeconomic impact. In: *State of the art: AIDS and economics*, edited by Steven Forsythe. Washington, DC , USA: Futures Group International:.

69. Greenland S (2005). Epidemiologic measures and policy formulation: lessons from potential outcomes. *Emerging Themes in Epidemiology*, 2 :5.

70. Gyimah-Brempong K, Wilson M (2004). Health human capital and economic growth in Sub-Saharan African and OECD countries. *The Quarterly Review of Economics and Finance*, 44: 296-320.

71. Gyrd-Hansen D (2005).Willingness to pay for a QALY: theoretical and methodological issues. *Pharmacoeconomics,* 23: 423-32.

72. Haacker (2002). The economic consequences of HIV/AIDS in Southern Africa. *IMF Working Paper No. 02/38*.

73. Hartunian NS, Smart CN, Thompson MS (1980). The incidence and economic costs of cancer, motor vehicle injuries, coronary heart disease, and stroke: a comparative analysis. *American Journal of Public Health*, 70: 1249–1260.

74. Hazan M and Zoabi H (2006). Does longevity cause growth? A theoretical critique. *Journal of Economic Growth*, 11: 363-376.

75. Hodgson TA, Kopstein AN (1984). Health care expenditures for major diseases in 1980. *Health Care Financing Review*, 5: 1-12.

76. Hodgson TA, Meiners M. (1982). Cost-of-Illness Methodology: A Guide to Current Practices and Procedures. *Milbank Memorial Fund Quarterly* 60:429-462.

77. Hodgson TA, Meiners MR (1979). *Guidelines for Cost-of-Illness Studies in the Public Health Service.* Bethesda, MD, USA: Public Health Service Task Force on Cost-of-Illness Studies.

78. Hu TW (2006). Perspectives: an international review of the national cost estimates of mental illness, 1990-2003. *Journal of Mental Health Policy and Economics*, 9: 3-13.

79. Hu TW, Sandifer FH (1981). *Synthesis of Cost-of-Illness Methodology: Part I.* Report to the National Center for Health Services Research, U.S. Department of Health and Human Services.

80. Jack W and Lewis M (2008). Health improvements and economic growth. *Background paper for the Commission on Growth and Development*. Washington, DC, USA: World Bank.

81. James HS (1996). The valuation of household production: how different are the opportunity cost and market price valuation methods? Unpublished paper (available on-line at http://econwpa.wustl.edu/eps/mic/papers/9612/9612001.pdf)

82. Jamison D, Lau L and Wang J (2005). Health's contribution to economic growth in an environment of partially endogenous technical progress. In López-Casasnovas G,

Rivera B and Currais L (eds.) *Health and Economic Growth: Findings and Policy Implications*. Cambridge, MA, USA: MIT Press.

83. Jefferis K, Kinghorn A, Siphambe H and Thurlow J (2008). Macroeconomic and household-level impacts of HIV/AIDS in Botswana. *AIDS*, 22: S113-S119.

84. Johannesson M (1996). The willingness to pay for health changes, the human-capital approach and the external costs. *Health Policy*, 36: 231-244.

85. Johansson P-O (1995). *Evaluating Health Risks: An Economic Approach*. Cambridge, UK: Cambridge University Press.

86. Johansson L (2007). Fiscal implications of AIDS in South Africa. *European Economic Review*, 51: 1614-1640.

87. Kalemli-Ozcan S (2006). AIDS, reversal of the demographic transition and economic development: Evidence from Africa. *NBER Working Paper 12181*.

88. Kalemli-Ozcan S, Ryder H and Weil D (2000). Mortality decline, human capital investment and economic growth. *Journal of Development Economics*, 62: 1-23.

89. Kambou G, Devarajan S, Over M (1993). The economic effects of the AIDS epidemic in Sub-Saharan Africa: a general equilibrium analysis. *Revue d'économie du développement,* 1: 37-62.

90. Kamolratanakul P, Sawert H, Kongsin S et al (1999). Economic impact of tuberculosis at the household level. *International Journal of Tuberculosis and Lung Disease*, 3: 596-602.

91. Kang HY, Kim HJ, Park TK, Jee SH, Nam CM, Park HW (2003). Economic burden of smoking in Korea. *Tobacco Control*, 12: 37-44.

92. Knowles S and Owen P (1995). Health capital and cross-country variation in income per capita in the Mankiw-Romer-Weil model. *Economics Letters*, 48: 99-106.

93. Kochar A (2004). Ill-health, savings and portfolio choices in developing countries. *Journal of Development Economics*, 73: 257-285.

94. Koopmanschap A, Rutten F, van Ineveld B and van Roijen L (1995). The friction cost method for measuring indirect costs of disease. *Journal of Health Economics*, 14: 171-189.

95. Kuhn A, Ruf O (2008). *The value of a statistical injury: new evidence from the Swiss labor market.* Working paper 367, Institute for Empirical Research in Economics, University of Zurich; Switzerland.

96. Lam TC, Small K (2001). The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review*, 37: 231-251.

97. Lammers J, Wijnbergen S (2007). *HIV/AIDS, risk aversion and inter-temporal choice.* Tinbergen Institute Discussion Paper, TI 2007-098/1.

98. Landefeld JS, McCulla SH (2000). Accounting for non-market household production within a national accounts framework. *Review of Income and Wealth*, 46: 289–307.

99. Landefeld JS and Seskin E (1982). The economic value of life: Linking theory to practice. *American Journal of Public Health*, 72: 555-566.

100. Laxminarayan R, Klein E, Dye C, Floyd K, Darley S and Adeyi O (2007). Economic Benefit of Tuberculosis Control. *World Bank Policy Research Working Paper 4295*.

101. Leardini G, Salaffi F, Caporali R, Canesi B, Rovati L, Montanelli R; Italian Group for Study of the Costs of Arthritis (2004). Direct and indirect costs of osteoarthritis of the knee. *Clinical and Experimental Rheumatology*, 22: 699-706.

102. Lee R (2003). The demographic transition: Three centuries of fundamental change. *Journal of Economic Perspectives*, 17: 167-190.

103. Lence SH (2000). Using consumption and asset return data to estimate farmers' time preferences and risk attitudes. *American Journal of Agricultural Economics*, 82: 934-947.

104. Liu JLY, Maniadakis N, Gray A, Rayner M (2002). The economic burden of coronary heart disease in the UK. *Heart,* 88: 597-603.

105. Lorentzen P, McMillan J and Wacziarg R (2005). Death and development. *NBER Working Paper 11620*.

106. Mankiw G, Romer D and Weil D (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107: 407-437.

107. McCarthy D, Holger W, Wu Y (2000). *The growth costs of malaria*. *NBER Working Paper 7541*.

108. McDonald S and Roberts J (2002). Growth and multiple forms of human capital in an augmented Solow model: A panel data investigation. *Economics Letters*, 74: 271-276.

109. Mendola M, Bredenkamp C and Gragnolati M (2007). The impoverishing effect of adverse health events: Evidence from the Western Balkans. *World Bank Policy Research Working Paper 4444*.

110. Miguel E and Kremer M (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72: 159-217.

111. Miller T, Mrozek J, Calhoun C and Arthur W (2002). Accounting frameworks for societal cost-outcome analyses of morbidity reduction. Papers, AERE Workshop, "Economic Valuation of Health for Environmental Policy: Assessing Alternative Approaches", March 18-19, 2002 www.bus.ucf.edu/mdickie/Health%20Workshop/Papers/Miller.pdf

112. Mills A(1994). The economic consequences of malaria for households: a case-study in Nepal. *Health Policy*, 29: 207-229.

113. Morduch J (1995). Income smoothing and consumption smoothing. *Journal of Economic Perspectives*, 9: 103-114.

114. Morduch J (1999). Between the state and the market: Can informal insurance patch the safety net?. *The World Bank Research Observer*, 14: 187-207.

115. Murphy K and Topel R (2006). The value of health and longevity. *Journal of Political Economy*, 114: 871-904.

116. Murray CJ, Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S (2003). Comparative quantification of health risks conceptual framework and methodological issues. *Population Health Metrics*, 1:1.

117. Noorbakhsh F and Paloni A (2001). Human capital and FDI inflows to developing countries: New empirical evidence. *World Development*, 29: 1593-1610.

118. Nordhaus W (2002). The Health of Nations: The Contribution of Improved Health to Living Standards. *NBER Working Paper 8818*.

119. Nordhaus W (2005). Irving Fischer and the contribution of improved longevity to living standards. *American Journal of Economics and Sociology*, 64: 367-392.

120. Olsen JA, Donaldson C (1998). Helicopters, hearts and hips: Using willingness to pay to set priorities for public sector health care programmes. *Social Science & Medicine*, 46: 1-12.

121. Olsen J, Smith R (2001).Theory versus practice: a review of 'willingness-to-pay' in health and health care. *Health Economics*, 10: 39-52.

122. Oster G, Colditz GA, Kelly NL (1984).The economic costs of smoking and benefits of quitting for individual smokers. *Preventive Medicine*, 13: 377-89.

123. Over M (1992). The macroeconomic impact of AIDS in sub-Saharan Africa. Washington, DC, USA: World Bank.

124. Piermartini R and Teh R (2005). Desmystifying modeling methods for trade policy. *WTO Discussion Paper no. 10*. Geneva, Switzerland: World Trade Organization.

125. Prskawetz A, Kögel T, Sanderson W and Scherbov S (2007). The effects of age structure on economic growth: An application of probabilistic forecasting to India. *International Journal of Forecasting*, 23: 587-602.

126. Quattek K and Fourie T (2000). Economic impact of AIDS in South Africa: a dark cloud on the horizon. Johannesburg, South Africa: ING Barings (South African Research),

127. Ramu K, Ramaiah KD, Guyatt H, Evans D (1996). Impact of lymphatic filariasis on the productivity of male weavers in a south Indian village. *Transactions of the Royal Society of Tropical Medicien and Hygiene*, 90: 669-70.

128. Reinhart V (199-9). Death and taxes: Their implications for endogenous growth. *Economics Letters*, 62: 339-345.

129. Rice DP (1966). *Estimating the Cost of Illness*. Health Economics Series, No. 6. DHEW Pub. No. (PHS) 947-6. Rockville, MD: U.S. Department of Health, Education and Welfare.

130. Rice DP (1967). Estimating the Cost of Illness. *American Journal of Public Health*, 57: 424-440.

131. Rice DP (2000). Cost of illness studies: what is good about them? *Injury Prevention*, 6: 177-179.

132. Rice DP, Hodgson TA, Kopstein AN (1985). The economic costs of illness: a replication and update. *Health Care Fin Rev* 7: 61-80.

133. Romer P (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94: 1002-1037.

134. Romer P (1990). Endogenous Technological Change. *Journal of Political Economy*, 98: S71-S102.

135. Rosen S (1988). The value of changes in life expectancy. *Journal of Risk and Insurance*, 1: 285-304.

136. Rosen S, Simon J, Thea D, Vincent J (2000). Care and Treatment to Extend the Working Lives of HIV-Positive Employees: Calculating the Benefits to Business. *South African Journal of Science*, 96: 300-305.

137. Rosen S, Simon J, Vincent J, MacLeod W, Fox M, Thea D (2003). AIDS is your business. *Harvard Business Review*, February 2003: 81-87.

138. Ruger J, Jamison D, Bloom D, Canning D (2006). Health and the Economy. In: MH Merson, RE Black, AJ Mills (Eds) *International Public Health: Diseases, Programs, Systems and Policies*. Sudbury, MA, USA: Jones and Bartlett Publishers.

139. Russell S (2004). The economic burden of illness for households in developing countries: a review of studies focusing on malaria, TB and HIV/AIDS. *American Journal of Tropical Medicine and Hygien*, 71 (suppl 2): 147-155.

140. Sala-I-Martin X, Doppelhofer G and Miller R (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review*, 94: 813-835.

141. Sala-i-Martin X (2005). On the Health Poverty Trap. In López-Casasnovas G, Rivera B and Currais L (eds.) *Health and Economic Growth: Findings and Policy Implications*. Cambridge, MA, USA: MIT Press.

142. Samuelson P (1947). *Foundations of Economic Analysis*. Boston, MA, USA: Harvard University Press.

143. Sanso M and Aísa R (2006). Endogenous longevity, biological deterioration and economic growth. *Journal of Health Economics*, 25: 555-578.

144. Sauerborn R, Adams A, Hien M (1996). Household strategies to cope with the economic costs of illness. *Social Science and Medicine*, 43: 291-301.

145. Segel J (2006). *Cost of illness studies - a primer*. RTI-UNC Centre of Excellence in Health Promotion Economics. RTI International, North Carolina, USA. Available on-line at: www.rti.org/pubs/COI_Primer.pdf

146. Sen A (1973). *On Economic Inequality*. New York, NY, USA: Norton.

147. Sen A (1985). *Commodities and Capabilities*. Oxford, UK: Oxford University Press.

148. Shastry G and Weil D (2003). How much of cross-country income variation is explained by health? *Department of Economics Working Paper 2003-8*. Providence, RI, USA: Brown University.

149. Shepard D, Ettling M, Brinkmann U and Sauerborn R (1991). The economic cost of malaria in Africa. *Tropical Medicine and Parasitology* 42(3, Suppl I):199-203.

150. Shiell A, Gerard K, Donaldson C (1987). Cost of illness studies: an aid to decision-making? *Health Policy*, 8: 317-323.

151. Smith J (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives*, 13: 145-166.

152. Smith R (2006). Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management. *Social Science and Medicine*, 63: 3113-3123.

153. Smith R, Yago M, Millar M and Coast J (2005). Assessing the macroeconomic impact of a healthcare problem: The application of computable general equilibrium analysis to antimicrobial resistance. *Journal of Health Economics*, 24: 1055-1075.

154. Smith R, Richardson J (2005). Can we estimate the 'social' value of a QALY?: Four core issues to resolve. *Health Policy*, 74: 77-84.

155. Steinberg M, Johnson S, Schierhout G, Ndegwa D (2002). *Hitting Home: how households cope with the impact of HIV/AIDS epidemic. A Survey of Households Affected by HIV/AIFDS in South Africa.* Washington, DC, USA: The Henry J. Kaiser Family Foundation.

156. Strauss J and Thomas D (1998). Health, Nutrition, and Economic Development. *Journal of Economic Literature*, 36: 766-817.

157. Suhrcke M, McKee M, Arce RS, Tsolova S and Mortensen J (2005). The contribution of health to the economy in the European Union. Brussels, Belgium: European Commission, DG Health and Consumer Protection.

158. Suhrcke M, Nugent R, Stuckler D and Rocco L (2006). Chronic Disease: An Economic Perspective. London, UK: Oxford Health Alliance.

159. Thomas D, Strauss J (1998). The micro-foundations of the links between health, nutrition and development. *Journal of Economic Literature*, 36: 766-817.

160. Thompson D, Edelsberg J, Kinsey K and Oster G (1998). Estimated economic costs of obesity to U.S. business. *American Journal of Health Promotion*, 13:120-127.

161. Turrell G (2000). Income non-reporting: implications for health inequalities research. *Journal of Epidemiology and Community Health*, 54: 207-214.

162. United Nations Department of Economics and Social Affairs / Population Division (2004). *The Impact of AIDS*. New York, NY, USA: United Nations.

163. Usher D (1973). An imputation to the measure of economic growth for changes in life expectancy. IN: M. Moss (ed.) *The measurement of economic and social performance* (pp. 193–226). New York, NY, USA: Columbia University Press.

164. Verstappen SM, Boonen A, Verkleij H, Bijlsma JW, Buskens E, Jacobs JW; Utrecht Rheumatoid Arthritis Cohort Study Group (2005). Productivity costs among patients with rheumatoid arthritis: the influence of methods and sources to value loss of productivity. *Annals of the Rheumatic Diseases*, 64: 1754-1760.

165. Viscusi K, Aldy JE (2003). The value of a statistical life: a critical review of market estimates from around the world. *Journal of Risk and Uncertainty*, 27: 5–76.

166. Wagstaff A (2007). The economic consequences of health shocks: Evidence from Vietnam. *Journal of Health Economics*, 26: 82-100.

167. Wang L, Zhang L, Hsiao W (2006). Ill health and its potential influence on household consumptions in rural China. *Health Policy*, 78: 167-177.

168. Weil D (2007). Accounting for the effect of health on economic growth. *Quarterly Journal of Economics*, 122: 1265-1306.

169. Weil P (1990). Non-expected utility in macroeconomics. *Quarterly Journal of Economics*, 105: 29-42.

170. WHO (1999). *The World Health Report, 1999: Making a difference.* Geneva, Switzerland: World Health Organization.

171. WHO (2001). *Macroeconomics and health: Investing in health for economic development*. Final Report of the Commission on Macroeconomics and Health. Geneva, Switzerland: World Health Organization.

172. WHO (2003). *Guide to Producing Health Accounts.* Geneva, Switzerland: World Health Organization.

173. Wilhelmson K and Gerdtham UG (2006). Impacts on economic growth of investing in maternal-newborn health. *Moving towards universal coverage WHO series*. Geneva, Switzerland: World Health Organization.

174. Winslow CE (1951). The cost of sickness and the price of health. *WHO Monograph Series* No. 7. Geneva, Switzerland: World Health Organization. [Chapter 4 on the 'Interrelationships of poverty and disease' is reprinted in: *Bulletin of the World Health Organization*, 84: 153-158 (2006)].

175. Wiseman V and Mooney G (1998). Burden of illness estimates for priority setting: A debate revisited. *Health Policy*, 43: 243-251.

176. Wiseman V and Mooney G (2000). Burden of disease and priority setting. *Health Economics*, 9: 369-372.

177. Xu K, Evans DB, Carrin G et al (2007). Protecting households from catastrophic health expenditures. *Health Affairs*, 26: 972-983.

178. Xu K, Evans DB, Kawabata K, Zeramdini R, Klavus J, Murray CJ (2003). Household catastrophic health expenditure: a multi-country analysis. *Lancet*, 362:111-7.

179. Yaari ME (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32: 137-150.

180. Young A (2005). The gift of the dying: The tragedy of AIDS and the welfare of future African generations. *Quarterly Journal of Economics*, 120: 423-466.

181. Zhang J and Zhang J (2005). The Effect of Life Expectancy on Fertility, Saving, Schooling and Economic Growth: Theory and Evidence. *Scandinavian Journal of Economics*, 107: 45-66.

182. Zhang J, Zhang J and Lee R (2001). Mortality decline and long-run economic growth. *Journal of Public Economics,* 80: 485-507.

Appendix A Ill-health and its economic consequences for the household

To illustrate the different possibilities for measuring the economic consequences of disease from a household's perspective, let us take the case of an agricultural household producing goods for market consumption and for home consumption. The household might hire labour to work on the farm, hire out their own labour on the market, or do both either simultaneously or at different times of the year. Following Bardhan and Udry (1999), a household in the canonical agricultural model can be assumed to seek to maximize its utility (U) described by the function:

$$U = U(L, C, M) \tag{1}$$

where
L = leisure time
C = consumption of home produced goods
M = consumption of market goods

First partial derivatives are all positive, and second partial derivatives are negative, representing diminishing marginal utility with increasing consumption. First order cross-partial derivatives are also positive, suggesting that the marginal utility associated with an increase in any of the arguments in U is positively associated with the level of the other arguments.

A household seeks to maximize this subject to three constraints:

$$C = C(L_c, B, y_c) - \sum p_c y_c \tag{2}$$

$$M = wL_w + (p_o F(L_o, B, y_i) - \sum p_i y_i) - \sum p_c y_c + E \tag{3}$$

and

$$L = T - L_w - L_o - L_c \tag{4}$$

In equation (2), $C(\bullet)$ is the production function for home consumed goods here assumed to require household labour (L_c), land (B) and other inputs (y_c). The latter set of goods is purchased at price p_c (although p_c could be zero in some cases).

Equation (3) describes the purchase of market goods, determined by household labour sold on the open wage market (L_w); the net value of the crops sold; the cash outlays to produce C ($\sum p_c y_c$); and non-labour income (E) which could be positive or negative. The net value of the crops sold is determined by the price of the output (p_o), the quantity produced, and the price of inputs (p_i). Output is determined by the production function $F(\bullet)$, in which L_o is household labour inputs, B is land and y_i are the purchased inputs.

Finally, equation (4) explains the time constraint. The finite amount of time available to the household (T), must be allocated between leisure and the production of marketed output, non-marketed consumption goods, and any wage earning activities.

This simple framework can be used to identify three possible ways of thinking about the economic consequences of ill health. Assume for simplicity that illness reduces the available time. This can influence all components of equation (4), and through it, all components of household utility or welfare in equation (1). We can measure economic consequences purely in terms of the lost consumption of market goods M. This is what is typically the focus of much of the macroeconomic literature on the economic consequences of health, which focus on some proxy of market consumption opportunities such as GDP.

However, this does not fully describe the consequences of illness on welfare. A slightly broader version would be to look at C and M, total lost consumption (market and non-market goods). Again, this is only part of the economic impact, so a third component is L, or the lost leisure. Between them, C, M and L are responsible for the overall impact on welfare. They can be measured independently and directly, at least in theory. The alternative is to try to measure the change in U that results from illness, which is mediated by the changes in C, M and L. This is the most complete concept of economic welfare from the perspective of an economist.

Now, let us introduce a further complication. Health (H) is generally considered to be a direct argument in the utility function as well. There is a health production function:

$$H = H(L_h, C_h) \qquad (5),$$

which we assume to be simply a function of time inputed by the household (L_h) and expenditure on health goods and services (C_h).

The equations above can now be modified in the following way. The household now maximizes a utility function:

$$U = U(L, C, M_n, H) \qquad (1a)$$

where
L = leisure time
C = consumption of home produced goods
M_n = consumption of non-health market goods
H = health status.

The constraints are now:

$$C = C(L_c, B, y_c) - \sum p_c y_c \qquad (2a)$$

$$M_n = wL_w + (p_o F(L_o, B, y_i) - \sum p_i y_i) - \sum p_c y_c - C_h + E \qquad (3a)$$

and

$$L = T - L_w - L_o - L_c - L_h \qquad (4a)$$

Here we assume that consumption of health goods and services increases utility indirectly through its impact on H, and not directly. The household must now decide how much time to devote to leisure, production of C and O, to wage labour and to health. It must also decide what inputs of health goods and services, and of goods and services to purchase to help in the production of M and C. In a more complex multi-period framework it would need to decide whether to save for future consumption, or to borrow.

The important point now is that the economic impact of disease can be measured in yet another way. The full welfare approach would be to measure the change in U again. But now the direct determinants of U are defined as leisure, non-market consumption, non-market non-health consumption and health status. Again, each can be measured in turn as separate components.

Although this model refers to an agricultural household, it was used only for heuristic purposes. Non-agricultural households produce non-market consumption, although not quite in the same way. This would include cooking, house cleaning and child minding, for example. If they might not produce a product for sale, this simply means that the production function for O disappears from equation (3a). We can generalize this to say that the economic impact of ill health could be measured by measuring each of the individual components of equation (1a), or the change in utility itself.

Appendix B Critical overview of the cost of illness methodology

The bulk of economic impact studies in health use some version of the cost of illness (CoI) approach, which was formalized by Rice and colleagues in the late 1960s and subsequently revised on several occasions (Rice 1966, 1967; Cooper and Rice, 1976; Rice, Hodgson and Kopstein, 1985). This was the first codified method for estimating the societal or population-level burden of disease, and remains by far the most common measurement approach. The possible economic consequences of specific illnesses are divided into 'direct costs', the expenses incurred because of the illness (including medical care, travel costs, etc.), and 'indirect costs', the value of lost production because of reduced working time. They did not try to measure the costs of pain and suffering, describing them as 'intangible costs'. Direct and indirect costs are then summed to provide the overall cost the illness imposes on society, often expressed as a percentage of current period Gross Domestic Product (GDP).

All costs incurred over a single index year, whether by new or pre-existing cases, are counted. By definition, intangible costs are assumed to be unquantifiable. Direct costs include resources used for prevention, treatment and rehabilitation for each disease (in a later update of their 1967 study, Rice and Cooper (1976) expanded their definition of direct costs to include expenditure on medical research, training, and capital investment in medical facilities). To the extent that they could be measured, costs incurred outside the health sector – e.g. travel to seek care; special food, equipment and clothing; modifications to houses and cars etc. – are also included.

Indirect costs or lost production are associated with reduced work time due to morbidity and mortality. The lost time of all cases in the index year due to morbidity in that year are added to the potential working years lost by people who died from the particular disease in that year. Sex- and age-specific labour force participation rates have been used to estimate the proportion of the lost time that would have been spent in the work force (assuming a constant rate of unemployment), and this lost time can then be multiplied by respective gross wage rates in order to give a total (discounted) estimate of the indirect costs of illness for the index year. Average earnings have also been used, because this was believed to be more equitable by not implying, for example, that the potential gains from eliminating diseases principally affecting white males were higher than those affecting black males (Rice, Hodgson & Kopstein, 1985).

Indirect costs have also been imputed for women not in paid employment, on the grounds that, if only the production of the formal labour force was included, comparisons of the economic costs across different diseases would be biased against those largely affecting women. Some studies have valued the time of women not in waged labour in terms of what they could have earned in the workforce (using the actual wages of women with similar characteristics), while others have used the average wages of domestic workers. Cooper and Rice (1976) have criticized the former approach on the grounds that it values time based on what a woman could be doing rather than on what she actually did. If this were used to value the time of women working in the home, they argued that it should also be used for people in the workforce who may be underemployed or seasonally unemployed.

A notable methodological development relates to the addition of an incidence-based approach as an alternative to prevalence-based studies of the cost of illness (e.g. Hartunian, Smart, and

Thompson 1980; Cooper and Rice, 1976; Berk, Paringer and Mushkin, 1978; Hodgson and Kopstein 1984). The definition of indirect and direct costs did not change, but they were estimated only for new cases of the disease occurring in the index year. Since the direct and indirect costs of chronic diseases such as HIV/AIDS or mental disorders may be incurred for many years after onset, the incidence approach requires the ability to predict natural history and treatment patterns over time, plus the time and age of death for all incident cases. The advantage of the incidence-based approach is that it is more appropriate for considering the potential benefits of prevention (for acute diseases with short durations, the prevalence and incidence approaches are identical).

A number of critiques of the COI approach have been published. For example, the variety of different approaches taken to estimating direct and indirect costs has been acknowledged to limit the comparability of results across studies (Ettaro et al., 2004; Hu, 2006; Segel, 2006), while the use of a 'human capital' approach to the measurement of lost production is said to lack a theoretical foundation and to overestimate actual economic losses (Hodgson and Meiners, 1982; Koopmanschap et al., 1995). Furthermore, the CoI approach assumes that the value to society of an individual's life is measured solely in terms of future production potential, thereby ignoring other dimensions of illness and death - as well as all non-market activities except 'housekeeping' time - that may be more important to an individual than lost economic output.

Human capital approach

Much of the methodological debate surrounding cost-of-illness studies has focused on the human capital approach, which values total production losses due to illness, disability or premature death by calculating the total period of absence and multiplying this by the wage rate of the absent worker. This would be consistent with neo-classical theory where the firm employs labour to the point where the value of the marginal product of a worker is equated to the wage rate, but it raises several problems in implementation.

One of the main shortcomings of the human capital approach is to only take into account the production and consumption costs of disease and injury (Johanesson, 1996). This ignores other benefits of improved health status, which might include for instance changes in leisure time and the intrinsic value of health, both for the individual and for others (altruistic motive). In the extreme case, by viewing the value of an individual as being equal to the value of his/her contribution to total production, the human capital approach implies that the statistical life of retired people has no value (Johansson, 1995; Landefeld and Seskin, 1982).

More generally, some authors argue that the human capital approach fails to take into account some of the general-equilibrium impacts resulting from changes in health associated with disease and injury (Arthur, 1981; Miller et al., 2002). This corresponds to a chain of wider economic impacts which might include, for instance, changes in fertility decisions. On the other hand, this suggests that the prolongation of life also brings an additional burden for the society in the form of extra claims on future consumption. The studies by Arthur (1981), Johanesson (1996) and Miller et al (2002) provide a theoretical welfare economics foundation for the human capital approach, provided that one recognizes that the human capital has a limited scope in terms of the components of welfare that are included in the valuation of economic consequences of disease and injury.

Another major criticism of the human capital approach is that it assumes the presence of a "full employment" economy, in which case the loss of each affected person cannot be offset by another worker (since labour markets are assumed to have cleared). This assumption is unrealistic in settings where a pool of underemployed or unemployed labour exists. Excess labour supply would allow the time lost to ill-health by a worker to be covered by others and would, therefore, mitigate the actual loss to the firm (Drummond, 1992; Koopmanschap et al., 1995). Just like households, firms do in fact pursue a number of coping strategies to mitigate economic losses due to ill-health, such as postponing or canceling non-urgent work, or having colleagues of the sick worker take on his/her duties within normal working hours. A Dutch study found that only 25-30% of conventionally calculated productivity costs remain after accounting for these compensating mechanisms (Jakob-Tacken et al., 2005). However, high rates of absenteeism may lead to coping strategies that adversely affect average levels of productivity, such as limiting staff specialisation or maintaining labour reserves to reduce the risk of staff shortages.

Friction cost approach

The friction cost approach has been proposed as a means of estimating the true production loss (Koopmanschap and van Inveld, 1992), since it restricts itself to the short-term impact of illness at the level of the firm by counting only the production lost while a replacement worker is found (i.e. it depends on the time span that organizations require to restore initial production levels). However, it should be noted that the extent of over-estimation depends on the nature of the job (whether the work must be performed on a schedule or can be made up later) and the type and amount of the labour pool that is available to the hiring firm (Hodgson, 1994). For economic effects beyond the short-term, such as changes in the supply of labour in the medium-term, it is proposed that impacts be evaluated via a macro-econometric model.

Application of this method in Holland suggests that CoI estimates based on a human capital approach significantly overestimates the impact of illness on firms and, by extension, the economy (Koopmanschap et al., 1995).To its advocates, the friction cost method is preferable because it allows for disequilibria in economies due to rising and falling unemployment through the business cycle.[10]

The friction cost approach has been criticised on several grounds (Liljas, 1998). Again, the implicit assumption is that sick workers are being replaced by those that are unemployed. In addition, the friction cost approach implies that increased sickness and absence from work would decrease the unemployment rate, but Johannesson and Karlsson (1997) argue that this is not supported by empirical evidence. More important for the application of the technique to general COI studies is that the increased costs of compensating for work absence are difficult to attribute to a particular disease as opposed to illness *per se*. Moreover, the technique is difficult to apply in the non-formal sector, which accounts for a considerable proportion of economic activity in developing countries (this limitation is true for most approaches, however). Further, the friction cost approach only captures the costs of replacing sick workers, not the costs of diminished performance while still at work (presenteeism). Comparison of different methods shows that,

[10] Koopmanschap et al. (1995) account for diminishing returns to labour by adjusting friction costs downwards by 20% to reflect an elasticity of annual labour time with respect to productivity of 0.8.

while illness at work can be quite common, the total impact on overall productivity losses at the firm level is generally expected to be quite small because employees continue carrying out most of their work (Brouwer et al., 1999).[11] In a more recent study, Collins et al (2005) find that the cost associated with presenteeism greatly exceeded the combined costs of absenteeism and medical treatment for a number of primary chronic health conditions.

[11] For some conditions (e.g., flu, cold or neck-ache) impaired performance occurring before or after absence from work itself may lead to an increase in estimated production losses by about 10% (Brouwer et al., 2002).

Appendix C Macroeconomic approaches to measuring the impact of disease and injury

Type of study	Definition and aims	Empirical approach	Implied quantity of interest	Data requirements	Advantages	Limitations	Data sources
'Cost-of-illness'	Estimates direct costs (medical care, travel costs, etc) and indirect costs (value of lost production from lower labour inputs) due to disease and injury	Direct and indirect cost are calculated to obtain monetary cost per patient, which is then multiplied by number of cases to obtain aggregate CoI. Traditionally done for life time costs of all new cases this year, or all new and pre-existing cases this year. This is often expressed as % of this year's GDP	Implied quantity of reference cannot be well defined: direct costs are foregone non-health market consumption, while indirect costs are forgone health and non-health production and consumption, often including both market and non-market costs.	• Disease-specific epidemiological data (e.g. prevalence, incidence, mortality); • Direct costs (treatment, diagnostics, drugs, medical equipment, supplies, transportation, special food and other); • Indirect costs (work days lost, absenteeism, unemployment, wages or some other indicator of marginal productivity)	• Simple implementation with relatively modest data and technical requirements • Data available free of charge in some cases • For advocates, usually results in a large number • Elements of the method can be used to address specific questions (eg. amount of direct costs due to a given disease at aggregate, household, firm or government level)	• Quantity of interest undefined; • Effects on demographic changes and depleted investment in physical and human capital (HK) excluded • HK approach lacks theoretical foundation and likely to overestimate actual losses; • HK raises equity concerns by assigning higher values for some groups than others (often misleadingly addressed by assuming same labour productivity for everyone)	• Several large nationally representative household surveys are available free of charge covering a number of developing countries: World Health Survey (who.int/healthinfo/survey), Living Standards Measurement Study (worldbank.org/lsms), Demographic and Health Surveys (measuredhs.com) • National and international agencies also collect panel survey data, particularly in developed countries: Panel Study of Income Dynamics (psidonline.isr.umich.edu), Medical Expenditure Panel Survey (meps.ahrq.gov), British Household Panel Survey (iser.essex.ac.uk/ulsc/bhps), European Union Statistics on Income and Living Conditions (ec.europa.eu/eurostat), Household, Income and Labour Dynamics in Australia (melbourneinstitute.com/hilda) • Specific diseases might require collecting primary survey data

Type of study	Definition and aims	Empirical approach	Implied quantity of interest	Data requirements	Advantages	Limitations	Data sources
Growth regressions	Estimate the impact of health indicators on GDP growth	Assumes a production function that uses physical capital, human capital and labour as inputs for producing GDP. Parameters of the production function are estimated using panel data regressions which calculate the contribution of health indicators to GDP growth	Monetary value of market production (health and non-health) foregone because of illness	• Economic data (GDP, capital stock, investment, natural resources); • Demographic indicators (population by age and gender, size of workforce); • Human capital (education, experience); • Epidemiological data (life expectancy, adult survival rates, prevalence, incidence, mortality)	• Established methodology; results can be compared with other studies • Data available free of charge in some cases	• Many studies use imperfect measures of population health (eg. life expectancy) • Endogeneity bias might require the use of econometric techniques to control for the simultaneous determination of health and income • Econometric estimation requires sample with large number of countries • Results sensitive to specification of production function	• Penn World Tables (pwt.econ.upenn.edu) compile international panel data on a range of macroeconomic indicators, including GDP growth, population, investment, consumption, government expenditures, openness and prices. Other sources include the CIA World Factbook (cia.gov/library/publications/the-world-factbook) and Thomson Datastream (datastream.com) • World Bank provides several datasets with macroeconomic indicators covering physical capital stock, human capital, education, political institutions, etc (econ.worldbank.org/resource.php?type=18) and the World Development Indicators • Epidemiological and mortality data are provided by WHO (who.int/topics/global_burden_of_disease), World Bank (devdata.worldbank.org/hnpstats), the Human Mortality Database (mortality.org) and other sources

Type of study	Definition and aims	Empirical approach	Implied quantity of interest	Data requirements	Advantages	Limitations	Data sources
Calibration models	Estimate the impact of health indicators on GDP or GDP growth	Micro estimates of exogenous impact of health on income are combined with country-level health indicators within an structural macro framework. This can be used to simulate the direct impact of health on income and to decompose differences across countries. Parameters estimates of the impact of health on economic outcomes are taken from (micro) studies in the literature	Monetary value of market production (health and non-health) foregone because of illness	• Economic data (GDP, capital stock, investment, natural resources); • Demographic indicators (population by age and gender, size of workforce); • Human capital (education, experience); • Epidemiological data (life expectancy, adult survival rates, prevalence, incidence, mortality)	• Based on structural modeling, therefore avoids endogeneity linked to empirical growth model • Flexible and suitable for applications with limited data • Data available free of charge in some cases	• More sensitive to mortality-related changes in workforce, but less so to capturing morbidity impacts • A possible way to capture morbidity impacts could model the marginal productivity or labour as function of health as well; however, this has seldom been done • Data limitations imply that parameters often have to be taken from other countries or settings	• Penn World Tables (pwt.econ.upenn.edu) compile international panel data on a range of macroeconomic indicators, including GDP growth, population, investment, consumption, government expenditures, openness and prices. Other sources include the CIA World Factbook (cia.gov/library/publications/the-world-factbook) and Thomson Datastream (datastream.com) • World Bank provides several datasets with macroeconomic indicators covering physical capital stock, human capital, education, political institutions and other (econ.worldbank.org/resource.php?type=18) and the World Development Indicators • Epidemiological and mortality data are provided by WHO (who.int/topics/global_burden_of_disease), World Bank (devdata.worldbank.org/hnpstats), the Human Mortality Database (mortality.org) and other sources

Type of study	Definition and aims	Empirical approach	Implied quantity of interest	Data requirements	Advantages	Limitations	Data sources
Computable general equilibrium (CGE) models	Estimates the impact of health indicators on: • GDP or GDP growth • Sectoral and distributio n impacts	Models start from preferences and constraints of individual agents (households, firms, government, etc) and build up to the simultaneous macroeconomic market-clearing equilibrium (set of prices that guarantees that demand equals supply across all sectors). Disease impact introduced as shocks to the supply of labour and the demand for services	• Monetary value of market production (health and non-health) foregone because of illness • Models used to estimate impact on different components as well, such as labour time, fertility, etc	• Disaggregated input-output matrices describing the overall economic structure; • Disease-specific epidemiological data (prevalence, incidence, mortality, duration of illness spells) • Parameters characterizing scenarios with and without the disease	Focus on general equilibrium effects captures intersectoral and dynamic adjustments which can be ignored by other models (eg ripple effects across sectors)	• Results sensitive to choice of parameters / specification • Considerable technical expertise required to obtain results • High computational costs (though falling now) • Severe data requirements: difficult to access complete input-output matrices; available data can be very retrospective • Data limitations imply that parameters often have to be taken from other settings / countries • In many cases data is not available free of charge	• The International Food Policy Research Institute provides a collection of Input-output social accounting matrices covering several developing countries (ifpri.org/data/data_menu.asp). The Global Trade Analysis Project (gtap.agecon.purdue.edu) also provides more elaborate data but at substantial cost • The Generalized Algebraic Modeling System (gams.com) is a popular software package with algorithms that can be used to solve models • Epidemiological and mortality data are provided by WHO (who.int/topics/global_burden_of_disease), World Bank (devdata.worldbank.org/hnpstats), the Human Mortality Database (mortality.org) and other sources

Type of study	Definition and aims	Empirical approach	Implied quantity of interest	Data requirements	Advantages	Limitations	Data sources
Full income models	Seeks to estimate the impact of health indicators on economic welfare	• Estimates of mortality and value of statistical life (VSL) are combined to convert the value of lives lost to disease or saved by intervention into monetary measure. This is then added to the estimates of lost market production and interpreted as lost welfare. Often this is compared with current year's GDP • Either revealed preferences or stated preferences techniques might be used as alternatives to VSL to estimate the willingness to pay (WTP) for health improvements	Monetary value of market and non-market production plus an estimate of the value of the lives lost, interpreted as social welfare foregone because of illness (death, most frequently)	• Economic data (GDP, real discount rates); • Value of statistical life, ideally by age, sex, and many other key factors	• Allows the estimation of the "full" economic welfare impact, which the other methods do not seek to capture • Simple implementation with relatively modest data and technical requirements, once VSL is available	• Controversial parameters, such as VSL and discount rates, are not observable and subject to extensive debate • Estimates are not bounded, can generate extreme results with difficult interpretation • Information failures pose threats to willingness to pay estimates • More readily applicable to mortality as opposed to morbidity impacts	• Sources described above can be used to obtain macroeconomic and epidemiological data • The value of statistical life, main parameter of analysis, can be obtained either from national estimates based on survey data or from literature reviews

Appendix D Time Preferences in an Expected Utility Model

In this note, the presentation in Lammers and Wijnbergen (2007) is followed. Generally, it is assumed that the expected utility (EU) model holds for economic agents (households) who have power utility functions - that is, one in which utility is equal to wealth raised to a constant that is less than 1 - and thus exhibit constant relative risk aversion.

Agents are assumed to make choices in order to maximize their utility from expected monetary outcomes (EM),

$$U(M_t) = \frac{E(M_t)^{1-\gamma}}{1-\gamma}, \qquad (1)$$

where $U(M_t)$ is the utility derived from monetary outcome M in period t, γ is the coefficient of relative risk aversion, and E is the expectations operator (which is applied to the distribution of the uncertain outcome).

Risk preferences are defined according to the value of γ. For example, when $\gamma = 0$ the agent is risk neutral, when $\gamma > 0$ the agent is risk averse (when $\gamma = 1$, we get the special case of the log utility function), and when $\gamma < 0$ the agent is risk seeking.

Now assume there are two monetary outcomes at different points in time, M_t and M_{t+k}. An agent would be indifferent between these two outcomes if they provide the same utility,

$$U(M_t) = \frac{1}{(1+\rho)^k} U(M_{t+k}), \qquad (2)$$

where $1/(1+\rho)^k$ is the discount factor and ρ is the rate of pure time preference for utility of M in period t versus utility of M in period t+k.

If agents are risk neutral, so that $\gamma = 0$ in (1), there is no curvature to the utility function and (2) can be rewritten in terms of just the monetary outcomes,

$$M_t = \frac{1}{(1+\rho)^k} M_{t+k}. \qquad (3)$$

In this case the agents will value the two monetary outcomes equally even though they occur in different time periods. Based on (3), one can derive the implied time preference parameter, ρ, so that the equality in (3) still holds.

The rate of time preference can be corrected for mortality risk. For example, the uncertainty of survival due to a health shock may lead households to discount the future more heavily (Yaari, 1965). Then, suppose the annual survival probability is p ($0<p<1$), and the survival function is $S_{t+k} = p^{k-t}$. Then the mortality risk-adjusted discount rate (ρ^m) can be defined as,

$$\left[\frac{1}{(1+\rho^{m})}\right]^{k} = \frac{S_{t+k}}{S_{t}(1+\rho)^{k}}, \tag{4}$$

where ρ is still the pure rate of time preference. To solve for the pure rate of time preference as a function of the probability of survival, one can substitute the survival function into (4) and solve for ρ to get (5).

$$\rho = (1+\rho^{m})p - 1. \tag{5}$$

Lammers and Wijnbergen discuss how an estimate of the probability of survival (p) might be derived from survey responses.

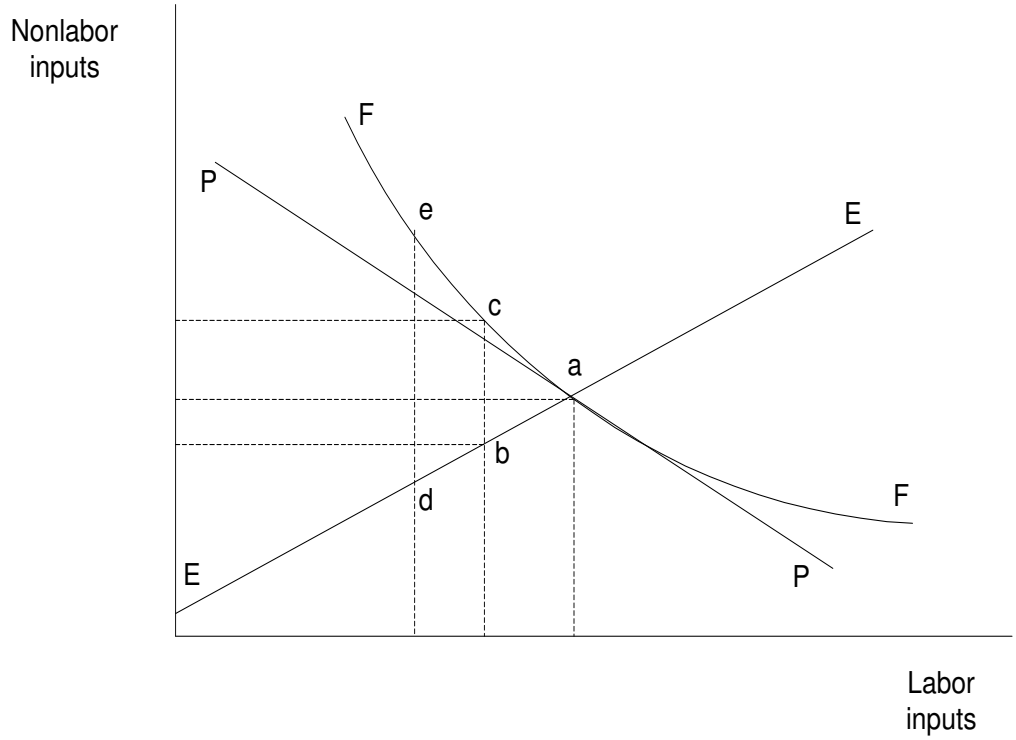
The assumption of risk neutrality is restrictive, yet this assumption can be relaxed. Dropping the assumption that agents are risk neutral in their inter-temporal choices and assuming that U(M) is an increasing function of γ, one can substitute (2) into (4) and derive the rate of time preference corrected for mortality risk and curvature of the utility function,

$$\rho = \left[\frac{S_{t+k}}{S_{t}}\left(\frac{M_{t+k}}{M_{t}}\right)^{1-\gamma}\right]^{\frac{1}{k}} - 1. \tag{6}$$

Using (6), the value of ρ can be calculated for agents who exhibit risk-averse preferences.

Lammers and Wijnbergen also extend the EU model to consider the case where relatively low discount rates might be due to diminished marginal utility of consumption due to poor health. This might be characteristic of HIV^{+} cases, whose expected future budget constraints are relatively tight due to large expenditures related to their illness and loss of income due to their inability to work. Due to anticipated differences in future marginal utility, the imputed discount rate will be lower than the true degree of time preference. They show that the implied discount rate can be obtained from the power utility function by deriving and inserting a consumption discount factor (the marginal utility of consumption in period t+k divided by the marginal utility of consumption in period t) into (6).

Appendix E Technical and allocative efficiency effects of adverse health shocks for firms



(with no technical change)

The upper curve (FF) reflects the level of output that can be produced with various combinations of labour and non-labour inputs for a given technology. The curvature of this iso-product frontier reflects the current state of technology. Point (a) represents a level of output that is on the frontier, it is both technically efficient (it lies on the expansion path coming out of the origin, EE, and is on the frontier) and it is allocatively efficient (it lies on the relative price line PP). Now assume there is an adverse health shock which reduces the availability of the labour input. Initially, the firm is forced to use less labour and moves back from point (a) to point (b), which is a technically inefficient level of output (it is inside the production frontier, FF). If the firm decides to replace the lost labour with more non-labour inputs it can move to point (c) which is technically efficient, but allocatively inefficient (it is not on the relative price line, PP, which is assumed to stay constant in this case).

To expand on these productivity effects and relate them to firm-level economic burden, consider two scenarios: a *static case* where the impacts are confined to increases in health-related expenditures and reductions in the level of single-period output, and a *dynamic case* where we add the effects of technological change which occur over time. In the *static case* we assume that the health shock and the costs to the firm are all confined to one period. In the simplest case we might assume the health shock is not too large and healthy workers temporarily fill-in for the unhealthy (absent) ones. If the level of firm production and profit are not reduced and no sickness benefits are afforded to workers, then the financial losses to the firm could be quite small or

negligible. However, given a sufficiently large health shock (i.e., the health shock is sufficiently large such that the slack labour in the firm is not able to absorb the loss of absent workers) there are two possible outcomes, depending on how the firm responds to the potential drop in production. As illustrated in the figure above, the firm may decide to incur additional labour and health-related expenses to maintain production (e.g., the firm remains on the efficiency frontier). This is point (c) on the graph. Then, firm earnings are reduced (e.g., the firm becomes allocatively inefficient because costs are not minimized) and the dividend payout to shareholders is potentially reduced (reducing investor consumption possibilities). However, the firm could elect to hold the payout to shareholders constant (in the static case the firm concerns itself only with the current period) by paying the dividend out of retained earnings, or it could reduce proportionately the payment to shareholders.

Alternatively, the firm could decide not to hire additional labour or make the added health expenditures and allow production to fall. The firm becomes technically inefficient, but maintains the mix of production inputs at their allocatively efficient levels (as at point (b) in the figure). Firm earnings would decline and again the firm could decide to either maintain the payout to shareholders or allow it to fall proportionately. Thus, in the static case, and given a sufficiently large health shock, the loss of investor consumption possibilities is not a deterministic result, since the firm may decide to maintain the consumption possibilities of shareholders in the short run, even at the expense of reducing firm equity capital.

Finally, in the static case we could hypothesize that firms exhibit different risk preferences. In this situation the magnitude of the potential loss due to a health shock (the size of the gamble) might induce firms to respond differently. Risk-averse firms might undertake larger health expenditures and/or efforts to mitigate the adverse productivity effects in order to maintain production and earnings, while risk neutral firms would not take these precautions.

In the *dynamic case* we assume that the health shock and the costs to the firm are realized over some relevant planning horizon and the firm's managers form expectations about the health shocks and learn how to mitigate the adverse productivity effects over time. The figure above illustrates the impact of a health shock in a static framework, where there is no technological change. Now assume that technological change is shifting the iso-product frontier (FF) inward toward the origin over time. Thus, technological progress allows the firm to produce a given level of output with less of each input. Then a health shock may have more complex effects on efficiency. The inward shift of the FF curve may be a parallel or nonparallel shift. For example, a non-parallel shift may imply that technological change is labour-saving. A health shock might not have any significant effect on output if the pace of technological change is rapid enough to offset the adverse health shock and the reduction of labour has no effect on output or productivity. This is not likely to be the case if the health shock evolves at a more rapid pace that exceeds the rate of technological change. Thus, the dynamics of these two factors may determine the path of both firm-level output and aggregate output.

Appendix F Glossary of main terms and definitions

Adult survival rate (ASR): A measure of adult population health defined as the probability that a 15 year-old individual will survive his/her 60th birthday if prevailing current patterns of mortality were to stay the same throughout the 45 years period.

Calibration models: Empirical approach to measuring the macroeconomic impact of health indicators on GDP growth. Within a structural macroeconomic framework, the approach combines microeconomic estimates of the exogenous impact of health on income with country-level health indicators. Simulation techniques are used to estimate the direct impact of health on income, to forecast the economic burden in future years and to decompose differences across countries.

Computable general equilibrium (CGE) models: Empirical approach to measuring the macroeconomic impact of health indicators on GDP growth, including sectoral and distributional impacts. Models start from preferences and constraints of individual agents (households, firms, government, etc.) and build up to the simultaneous macroeconomic market-clearing equilibrium. The equilibrium is characterized as the simultaneous price-guided clearing of all sectors of the economy (i.e. the set of prices that guarantees that demand equals supply across all sectors). The impact of disease is introduced as a shock to the supply of labour and the demand for services.

Consumption insurance: The extent to which households, following a negative health shock, are able to equalize the rate of marginal utility of consumption over time by smoothing out the path of consumption of non-health goods and services. This is an important measure of household economic welfare that depends on access to formal and informal mechanisms of social security, which in turn affects the household's ability to maintain stable levels of non-health consumption.

Coping strategies: Strategies used by households or firms to limit the adverse effects of disease and injury, including substitution of labour within the household (to preserve production and income flows) and the disposal of assets or savings (to pay for health care).

Cost-of-illness (CoI): Empirical approach to estimating the societal impact of disease and injury which combines 'direct costs' (medical care, travel costs, etc) and 'indirect costs' (the value of lost production because of reduced working time) into an overall estimate of economic impact on society, often expressed as a percentage of current GDP. As discussed in Section 3 of this guide, a number of methodological limitations restrict its applicability to estimate the macroeconomic impact of disease and injury.

Counterfactual: The comparator situation against which the current or future economic impact of a disease is determined. A number of counterfactuals are possible, for instance: a) the total elimination of all the cases of the disease; b) the elimination of all new cases in a given year; c) a reduction in the number of new cases in a given year, etc.

Endogeneity: Endogeneity or reverse causality is defined as the simultaneous determination of variables within an economic system. Endogeneity is an important issue in cross-country growth models that aim to estimate the economic impact of disease and injury in economic growth. While better health will lead to higher income through various specified channels, it is also generally the case that higher income will lead to better health, as more money is available to prevent and cure diseases and people have access to better quality health goods and services. Moreover, it is also possible that unobservable factors affect both health and income simultaneously, causing further difficulties for the identification of the effect of health on economic growth.

Friction cost approach: A refinement of the human capital approach that proposes to estimate the true level of foregone production by restricting itself to the short-term impact of illness at the level of the firm; it dies this by counting only the production lost while a replacement worker is found (i.e. it depends on the time that organizations require to restore initial production levels).

Full income approach: Analytical approach aiming to provide a more comprehensive measure of the total economic welfare consequences of disease and injury. Full income models incorporate both market and non-market consequences, including estimates of the economic value of diminished or lost health.

Gross domestic product (GDP): A measure of national income and output for a given country's economy. GDP is defined as the total market value of all final goods and services produced within the country in a given period of time (usually a calendar year).

Growth regressions models: Empirical approach to measuring the macroeconomic impact of health indicators on GDP growth. This approach assumes that GDP is produced according to a production function that uses as inputs physical capital, labour, human capital and technology. The parameters of the production are then recovered using panel data regressions which estimate the contribution of health measures to GDP growth over time across different countries.

Human capital approach: Measurement approach to estimate the value of production losses due to illness, disability or premature death, achieved by multiplying the total period of absence by the wage rate of the absent worker. This would be consistent with neo-classical theory where the firm employs labour to the point where the value of the marginal product of a worker is equated to the wage rate. The main limitation of the approach is that it (unrealistically) assumes the presence of full employment in the economy, and by focusing only on the productive capacity of individuals, ignores other benefits of improved health status.

Input-output matrix (I/O): A matrix representation of the structure of the economy, which provides a description of the sources (input) and uses (output) of monetary resources across aggregate sectors, industries and economic agents (e.g. government, external sector, etc). I/O matrices form the basis of most CGE models, which use the information in the I/O matrix to calibrate the parameters of the model to fit the current state of the economy. Parameters are later changed in order to simulate the impact of policy or institutional changes, for example, the likely economic impact of the eradication of a given disease or injury. I/O matrices are also known as social accounting matrices.

Output-based approach: Measurement approach that seeks to estimate actual economic losses resulting from disease and injury, which it does by evaluating changes in economic output resulting from a productive activity, either market or non-market based. In contrast, an input-based approach predicts potential economic losses by measuring changes in the level of inputs to production.

Quantity of interest: An expression of the underlying entity that an economic impact study is intending to capture, such as the impact of disease on firm production or national income. A number of possible quantities of interest can be identified, each of which has a clear economic meaning. However, great care should be taken when aggregating the various costs of disease or injury, since more than one quantity of interest may be implied by doing so (thereby rendering results potentially meaningless from a conceptual point of view).

Revealed preference: Technique to elicit individual preferences based on the observation of market trade-offs that individuals are prepared to make in order to satisfy their preferences. This technique has been specifically used to place a value on human life by observing the trade-offs that individuals make between income and the risk of death. The revealed amount of money needed to accept a specified increase in the chance of death is used to estimate a (market-based) estimate of the value of a statistical life.

Stated preference: Technique to elicit individual preferences based on the application of survey questionnaires where individuals are asked directly about their hypothetical willingness to pay for specific health improvements (or some other desired outcome).

Willingness to pay (WTP): Technique to elicit the value that individuals place on an economic resource or change in welfare by observing how much a person is willing to pay in order to obtain it. In the case of market transactions, WTP is observed directly and amounts to the price that is paid, while the valuation of non-market services and goods (such as the value of human life or the value of pain/suffering) might require the use of indirect measures, such as revealed choices or stated preferences.

What impact does ill-health have on the rate of economic growth in countries? How much do households pay for medical or other expenses because of illness? What is the impact of ill-health on productivity in the work place? What proportion of government resources could have been directed to alternative uses in the absence of illness? These are some of the questions that can contribute to our understanding of the full impact of disease and ill health in different contexts, and which can be answered by quantifying the economic consequences of disease and injury.

Interest in measuring the economic consequences of illness has a long history and shows no sign of abating. But although there is general agreement about the channels via which poor health can have economic effects, there have been many problems with the methods used to obtain empirical estimates. In light of these methodological shortcomings, as well as the strong continuing demand for economic impact studies in health, WHO is proposing a defined conceptual framework within which the economic impact of disease or injury can be considered and appropriately estimated, with a view to enhancing the consistency and coherence of economic impact studies in health. Specifically, this guide addresses:

- ***the conceptual foundations of economic impact studies, which covers a number of core issues such as study perspective, timescale and scope; and***
- ***the implications that choices around key conceptual issues have for the measurement and valuation of costs at the macroeconomic and microeconomic level.***

It is hoped that this guide will contribute to improve standards of practice and, by increasing the degree of comparability and uniformity between studies, promote more informed discussion in this field.

9 789241 598293