WHO GUIDE TO COST-EFFECTIVENESS ANALYSIS

CONTENTS

The graphic on the cover is a stylized representation of Figure 5 in Background Paper 7 showing uncertainty analysis.

Making Choices in Health:

WHO guide to cost-effectiveness analysis

Edited by

T. Tan-Torres Edejer, R. Baltussen, T. Adam,
R. Hutubessy, A. Acharya, D.B. Evans and
C.J.L. Murray

World Health Organization
Geneva

© World Health Organization 2003

Printed in Switzerland

Contents

Acknowledgements

Many individuals have contributed to the refinement of the framework, methods and tools for generalized cost-effectiveness analysis. Their contributions are recognized collectively under the name WHO-CHOICE Collaboration. In addition to them, a number of experts provided valuable input in anonymous reviews for the published papers, during the consultative meeting in Geneva in 2001, and during other scientific meetings where WHO-CHOICE or generalized cost-effectiveness analysis was presented. Their contributions are gratefully acknowledged. Another important source of feedback has been the policy-makers and analysts from many different countries who have attended the workshops on generalized cost-effectiveness analysis that we have conducted over the past years. Close interaction with them as they applied the tools and techniques gave us insights into a range of concerns from a user's perspective, and made us aware of the need to be responsive to different audiences.

Lastly, we would like to give special thanks to Margaret Squadrani, Marilyn Vogel, Kai Lashley and Keith Wynn who, with their dedication, patience and painstaking attention to detail, made the actual production of the book possible.

This guide benefited from the input of many experts, including those participating in a meeting on Methods for Cost-Effectiveness Analysis held in Geneva in January 2002. Those who participated were:

Dr Arnab Acharya, The Institute of Development Studies, UK
Dr Jan Barendregt, Erasmus University, The Netherlands
Dr Stephen Birch, McMaster University, Canada
Dr Dan Brock, National Institutes of Health, USA
Dr Werner Brouwer, Erasmus University Rotterdam, The Netherlands
Dr Vilma Carande-Kulis, Centers for Disease Control & Prevention, USA
Dr Michael Drummond, University of York, UK
Dr Alastair Gray, University of Oxford, UK
Dr Paul Kind, University of York, UK
Dr Martin Meltzer, Centers for Disease Control and Prevention, USA
Dr Mark Miller, National Institutes of Health, USA
Professor Anne Mills, London School of Hygiene & Tropical Medicine, UK
Dr Raul Molina, Universidad Autonoma Metropolitana-Iztapalapa, Mexico
Dr Benjamin Nganda, University of Nairobi, Kenya
Dr Louis Niessen, Erasmus University, The Netherlands
Dr Max Price, University of the Witwatersrand, South Africa
Professor Jeff Richardson, Centre for Health Programme Evaluation, Australia
Dr Juan Rovira, The World Bank, USA
Dr Orville Solon, University of the Philippines, Philippines
Dr Siripen Supakankunti, Chulalongkorn University, Thailand
Dr Ben van Hout, Utrecht University, The Netherlands
Dr Leena K. Eklund, WHO Regional Office for Europe, Denmark
Dr Hossein Salehi, WHO Regional Office for the Eastern Mediterranean, Egypt

The views expressed in these guidelines are those of the authors and not necessarily those of the participants.

Preface

A number of guidelines on cost-effectiveness analysis (CEA) already exist. There are two reasons for producing another set. The first is that traditional analysis has focused on assessing new or additional interventions in comparison with current practice in that area. It is difficult to use this type of "incremental" analysis to determine if the current mix of interventions represents an efficient use of resources. Secondly, for all but the richest countries, the cost and time required to evaluate the large number of interventions and identify opportunities to enhance efficiency are prohibitive. It is important to maximize the possibility of generalizing results from one setting to another. The approach of generalized CEA (GCEA) proposed in this Guide seeks to provide analysts with a method of assessing whether the current mix of interventions is efficient as well as whether a proposed new technology or intervention is appropriate. It also seeks to maximize the generalizability of results across settings.

The WHO Guide to Cost-Effectiveness Analysis should be considered as complementary to existing guidelines on CEA. GCEA proposes the evaluation of interventions against the counterfactual of "doing nothing", thereby providing decision-makers with information on what could be achieved if they could start again to build the health system, i.e. reallocate all health resources. As will be shown in the Guide, this information is a prerequisite to the definition of an efficient mix of interventions, achievable in the long run. This specific feature—not addressed in traditional CEA which typically evaluates new interventions in comparison with the current mix (intervention mix constrained CEA or IMC-CEA)—categorizes GCEA as a different, more fundamental, type of economic analysis. For many narrower applications of CEA, such as the appraisal of a new version of an existing drug in a specific country, the currently practised CEA remains an appropriate method although it should be realized that this does not inform decision-makers on the best use of health resources in general.[i] It is also possible to undertake a traditional analysis as part of a GCEA.

The main objective of this Guide is to provide policy-makers and researchers with a clear understanding of the concepts and benefits of GCEA. It provides guidance on how to undertake studies using this form of analysis and how to interpret the results. The main focus is on those

methodological issues which make GCEA different from traditional CEA, such as the definition of the counterfactual for analysis. In addition, attention is paid to controversial issues in CEA where choices are required, such as the inclusion or exclusion of productivity costs. Furthermore, the Guide provides some detailed discussions on issues which are little debated in the literature but nevertheless important, for example, the technical approach to the transferability of cost estimates across settings. On all these matters, the Guide has benefited from a meeting of experts in cost-effectiveness analysis convened by WHO in Geneva in early 2002. In that meeting, the first version of the Guide was presented, and this published version builds on some of the discussion in that meeting.

Since GCEA and IMC-CEA are both embedded in the same economic framework, they share many of the same techniques which are discussed in detail elsewhere, such as the methods for the allocation of hospital costs. In those instances, the reader is referred to the other literature. This Guide proposes a standard set of methodological choices on how to perform GCEA to enhance the comparability and generalizability of results. The intended audience are those analysts with some background in CEA.

The Guide, in Part One, begins with a brief description of GCEA and how it relates to the two questions raised above. It then considers issues relating to study design, estimating costs, assessing health effects, discounting, uncertainty and sensitivity analysis, and reporting results. Detailed discussions of selected technical issues and applications are provided in a series of background papers, originally published in journals, but included in this book for easy reference in Part Two.

The first paper by Murray et al., on "Development of WHO Guidelines in Generalized Cost-Effectiveness Analysis" formally lays out the motivation and framework for GCEA. It highlights the use of GCEA for improving sectoral efficiency, based on the comparative analysis of current as well as proposed new interventions against a common counterfactual.

The second paper, "PopMod: A longitudinal population model with two interacting disease states", is a detailed technical description of the multi-state dynamic life table that calculates the health and mortality experience of a population with two interacting conditions or disease states, as well as other causes of mortality and morbidity. It was developed by WHO in a spreadsheet format, and subsequent collaboration with a scientific consultancy group and Statistics Canada allowed the model to be transferred into various programming environments including a microsimulation version.

The next two papers, "Programme costs in the economic evaluation of health interventions" and "Econometric estimation of country-specific hospital costs", describe how cost estimates can be derived for different subregions. The programme cost paper describes how one category of costs, those which represent resources consumed at all levels aside from

the actual delivery of the intervention to an individual recipient, is estimated through quantification of resource inputs, choice of resource prices, and accounting for different levels of coverage. Sources of data included figures from the literature and expert opinion. On the other hand, the hospital cost paper provides details on how multiple datasets collectively containing 2054 country years of observation from 49 countries were inputted into a model which can be used to produce estimates for hospital costs in countries where data are not available, with appropriate confidence intervals.

Uncertainty surrounds estimates of intervention effectiveness, costs and the resulting cost-effectiveness ratios. Two papers on uncertainty analysis are included. These are "Uncertainty in cost-effectiveness analysis: probabilistic uncertainty analysis and stochastic league tables" and "Stochastic league tables: communicating cost-effectiveness results to decision-makers". The first paper shows how to construct uncertainty intervals for the estimated cost-effectiveness ratio of a single intervention, using Monte Carlo simulations to sample from distributions of key variables. The second paper takes a sectoral perspective and describes how uncertainty intervals of multiple interventions, even if overlapping, can still provide decision-makers with useful information through a stochastic league table showing the probabilities that a specific intervention will be included in the optimal mix for different resource levels.

"Effectiveness and costs of interventions to lower systolic blood pressure and cholesterol: a global and regional analysis on reduction of cardiovascular risk" is an application of the methods of generalized cost-effectiveness analysis in the field of cardiovascular risk factors. The analysis, using the methods described in the Guide and the papers, presents the cost-effectiveness of different kinds of interventions versus cardiovascular risk factors. It demonstrates the versatility of the method in dealing with non-personal or population-based (e.g. decreasing salt content of processed foods) and personal interventions (e.g. intake of medicines for individuals with a predicted risk of a cardiovascular event in 10 years above a pre-specified level). It also shows how GCEA can be used to calculate the cost-effectiveness of different combinations of interventions at different coverage thresholds.

The paper on the policy uses of generalized cost-effectiveness analysis, "Generalized cost-effectiveness analysis; an aid to decision-making in health", describes the past experiences in sector-wide cost-effectiveness analysis and its apparent limitations that have restricted the results of CEA from being routinely incorporated into policy. This paper suggests that generalized cost-effectiveness analysis can be used to identify existing inefficiencies, if any, with current choices, and also to propose more cost-effective alternatives.

The final paper, "Ethical issues in the use of cost-effectiveness analysis for the prioritization of health care resources", goes through the different ethical choices being made when undertaking cost-effectiveness analyses

and applying the results in policy development. Unlike the other papers which deal exclusively with generalized cost-effectiveness analysis, this is a generic exposition which applies to all cost-effectiveness analyses. It is included because of the need to make all analysts aware of the value choices inherent in work which is usually considered "technical" in nature.

The Guide and these papers are written in the context of the work of WHO-CHOICE: **CHO**osing **I**nterventions that are **C**ost-**E**ffective. WHO-CHOICE is assembling regional databases on the costs, impact on population health and cost-effectiveness of key health interventions. This work started in 1998 with the development of standard tools and methods and is now collecting, analysing, and disseminating for policy purposes data on costs and outcomes of a variety of interventions.

In addition to the present guidelines, WHO-CHOICE has developed tools to allow data to be collected and reported in a standardized way. They include a costing template (CostIt) to record cost data in a way that is of most use to analysts and policy-makers. This template assembles information on the quantities of inputs used in an intervention and their unit prices. Tools to collect primary data in a form consistent with the template have also been developed for particular interventions. Furthermore, a population model (PopMod) has been developed which will automatically calculate the effectiveness of interventions for a standardized population, in terms of outcome indicators such as disability-adjusted life years (DALYs) averted. "Monte Carlo League" (MCLeague), a programme that presents uncertainty around costs and effects to decision-makers in the form of stochastic league tables is also available. It provides additional information beyond that offered by the traditional treatment of uncertainty in cost-effectiveness analysis, presenting the probability that each intervention is included in the optimal intervention mix for given levels of resource availability. Versions of these tools are provided in the accompanying compact disc and any updates can be obtained from www.who.int/evidence/cea.

The first results of WHO-CHOICE were reported in the *World Health Report 2002*, which included estimates of the cost-effectiveness of a large number of interventions to reduce risks to health in 14 epidemiological subregions of the world. They are reported in such a way at www.who.int/evidence/cea that analysts in different settings can modify the assumptions, if that is necessary, to adapt the results to local conditions. In the long term, WHO-CHOICE will make available and update regularly estimates of the cost-effectiveness of a larger number of interventions in these epidemiological subregions. Refinement of the methodology is ongoing with collection and re-analysis of more cost data to improve estimates, and experiments with new modelling tools including microsimulation.

Making tools that are user-friendly to facilitate contextualizing results to local settings is a continuing priority. As more country analysts use the

tools and the contextualised results, their feedback will be valuable in
further improving WHO-CHOICE and making it an indispensable part
of the evidence base to support decisions in resource allocation. That
being said, it is important to emphasize that evidence of the costs and
effects of interventions is only one input to the priority-setting process. It
is a key input, showing how resources could be used to maximize
population health. But this evidence needs to be evaluated alongside
information on how each possible mix of interventions influence other
health system goals such as reducing inequalities, or being responsive to
the population they serve. WHO-CHOICE is beginning to develop tools
to facilitate this type of policy dialogue by providing, as a first step,
evidence of the cost-effectiveness of interventions in the poor.

---

[i] On a different basis, the traditional approach to CEA is conceptually important as an additional
analysis to generalized CEA, once the efficient mix of interventions has been identified. Such
"incremental" analysis—evaluating interventions in comparison to the current mix of interventions—is
then used to provide information on how this efficient mix of interventions can best be achieved, starting
from the current mix. The current guidelines addresses the use of generalized CEA only, i.e. on how to
identify the efficient mix of interventions.

LIST OF AUTHORS

Arnab Acharya
The Institute of Development Studies
University of Sussex
Brighton, BN1 9RE, UK

Taghreed Adam
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Rob M.P.M. Baltussen
Institute for Medical Technology Assessment (iMTA)
Erasmus University
Rotterdam, The Netherlands

Jan J. Barendregt
Department of Public Health
Erasmus University
Rotterdam, The Netherlands

Dan Brock
Department of Clinical Bioethics
Warren G. Magnuson Clinical Center
National Institutes of Health
Bethesda, MD, USA

Claude Charette
Statistics Canada
R.H. Coats Building
Holland Avenue
Ottawa, Ontario K1A OT6, Canada

Dan H. Chisholm
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

David B. Evans
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Steve Gribble
Statistics Canada
R.H. Coats Building
Holland Avenue
Ottawa, Ontario K1A OT6, Canada

Raymond C.W. Hutubessy
Stop TB Department
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Benjamin Johns
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Jeremy A. Lauer
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Carlene M.M. Lawes
Clinical Trials Research Unit
University of Auckland
Auckland, New Zealand

Christopher J.L. Murray
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Louis Niessen
Institute of Medical Technology Assessment and
Institute of Health Policy Management
Erasmus University Medical Centre
Rotterdam, The Netherlands

Anthony Rodgers
Clinical Trials Research Unit
University of Auckland
Auckland, New Zealand

Klaus Röhrich
Creative Services
CS Science & Technologies
Technoparc Pays de Gex
55 rue Auguste Piccard
01630 Saint Genis-Pouilly, France

Niels Tomijima
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Tessa Tan-Torres Edejer
Global Programme on Evidence for Health Policy
World Health Organization
20 Avenue Appia
1211 Geneva 27, Switzerland

Harald Wirth
Creative Services
CS Science & Technologies
Technoparc Pays de Gex
55 rue Auguste Piccard
01630 Saint Genis-Pouilly, France

LIST OF ACRONYMS AND ABBREVIATIONS

BCG	Bacille Calmette-Guérin vaccine (for tuberculosis)
BOD	burden of disease
CAB	Commonwealth Agricultural Bureau
CBA	cost-benefit analysis
CE	cost-effectiveness
CEA	cost-effectiveness analysis
CER	cost-effectiveness ratio
CIF	cost, insurance and freight (price of imports)
CostIt	the WHO-CHOICE template for costing interventions
CPI	consumer price index
CVD	cardiovascular disease
DALY	disability-adjusted life year
DisMod	DisMod is a WHO software tool that may be used to check the internal consistency of epidemiological estimates of incidence, prevalence, duration and case fatality for diseases
DOTS	directly observed therapy, short course (for tuberculosis)
FOB	free on board (price of exports)
GBD	global burden of disease
GCEA	generalized cost-effectiveness analysis
GDP	gross domestic product
GIS	geographical information system
HALE	healthy life expectancy
HRAM	health resource allocation model
HRQL	health-related quality of life
HSPR	health sector priorities review
HSV	health state valuation
HUI	health utilities index
HYE	healthy year equivalent
HYL	healthy years lived
ICER	incremental cost-effectiveness ratio
ILO	International Labour Office
IMC-CEA	intervention-mix constrained CEA
IPM	incidence, prevalence and mortality

MCLeague	the WHO-CHOICE software—Monte Carlo League table—used to show the impact of uncertainty on CERs
mm Hg	millimeters of mercury (for measuring blood pressure)
OER	official exchange rate
OLS	ordinary least squares regression
ORT	oral rehydration therapy
PopMod	the WHO-CHOICE model for estimating the population health impact of interventions
PPP	purchasing power parity
QALY	quality-adjusted life year
QWB	quality of well-being scale
SBP	systolic blood pressure
SCF	standard conversion factor
SD	standard deviation
SE	standard error
TB	tuberculosis
VIF	variance inflation factors
WDR	World Development Report
WHO	World Health Organization
WHO-CHOICE	WHO's project on providing cost-effectiveness information: **CHO**osing Interventions that are **C**ost-**E**ffective
WTP	willingness to pay
YLD	years of healthy life lost to non-fatal health outcomes
YLL	years of life lost to premature mortality

Part One: Methods for generalized cost-effectiveness analysis

Rob Baltussen, Taghreed Adam,
Tessa Tan-Torres Edejer, Raymond Hutubessy,
Arnab Acharya, David B. Evans,
Christopher J.L. Murray

1 WHAT IS GENERALIZED COST-EFFECTIVENESS ANALYSIS?

1.1 INTRODUCTION

The growing use of cost-effectiveness analysis (CEA) to evaluate the costs and health effects of specific interventions is dominated by studies of prospective new interventions compared to current practice *(1;2)*. This type of analysis does not explicitly take a sectoral perspective where the costs and effectiveness of all possible interventions are compared in order to select the mix that maximizes health for a given set of resource constraints. The estimated cost-effectiveness of a single proposed new intervention is compared either with the cost-effectiveness of a set of existing interventions reported in the literature or with a fixed price cut-off point representing the assumed social willingness to pay for an additional unit of health. The implicit assumption that the required additional resources would need to be transferred from another health intervention or from another sector is rarely discussed.

Much of the theoretical literature has taken a broader view of cost-effectiveness, exploring its use in allocating a fixed health budget between interventions in such a way as to maximize health in a society *(3;4)*. We refer to this as sectoral CEA. Only a few applications of this broader use—in which a wide range of preventive, curative and rehabilitative interventions that benefit different groups within a population are compared in order to derive implications for the optimal mix of interventions—can be found. Examples include the work of the Oregon Health Services Commission *(5)*, the World Bank Health Sector Priorities Review *(6)* and the Harvard Life Saving Project *(7)*. Of these, only the World Bank attempted to make international or global comparisons of sectoral cost-effectiveness.

At the heart of this broadened policy use is the notion that health resources should be allocated across interventions and population groups to generate the highest possible overall level of population health. If

the calculations show that some current interventions are relatively cost-ineffective, and that some which are not undertaken fully are relatively cost-effective, resources could be reallocated across interventions to improve population health. In other words, moving resources from cost-ineffective interventions to cost-effective ones could enhance the allocative efficiency of the health sector.[1] Interest in the promise of enhancing allocative efficiency of health systems has led to analytical efforts to study the cost-effectiveness of a broad range of interventions in a number of countries *(8;9)*.

Several challenges have emerged to this wider use of CEA. First, analysts and decision-makers have correctly noted that resource allocation decisions affecting the entire health sector must also take into account social concerns such as prioritizing the sick, reducing inequalities in health, or addressing the well-being of future generations. Vociferous debate on the use of CEA to prioritize the use of Medicaid resources in Oregon State in the United States is one indication of these concerns in the political arena *(5;10;11)*. So far there have been two proposed responses to this challenge: abandon the practice of using CEA to inform resource allocation decisions entirely or progressively incorporate more of these social concerns into the methods of CEA. Second, current CEA practice often fails to identify existing misallocation of resources by focusing on the evaluation of new technologies or strategies. The very wide range of cost-effectiveness ratios found in the compendia of CEA listed above suggest that addressing current allocative inefficiencies in many countries may yield substantial health gains, possibly more than identifying new technologies that will make small improvements in health.

Third, for all but the richest societies, the cost and time required to evaluate the large set of interventions needed to use CEA to identify opportunities to enhance allocative efficiency may be prohibitive. The results of many, if not most, CEA studies are so context-specific that they cannot be used to inform policy debate in other populations—as reflected in the debate about the use of league tables which include the results of studies using a variety of methods and which were undertaken to answer a variety of context-specific questions *(12)*. For most countries, but particularly for low- and middle-income countries where the majority of the world's poor live, there has been little progress towards the goal of providing affordable and timely information on the costs and effects of a wide array of interventions to inform policy. Fourth, the difficulties of generalizing context-specific CEA studies have been institutionalized by the proliferation of multiple national or subnational guidelines for CEA practice, all using slightly different methods. International guidelines have not to date been developed.

In this section, we outline some of the uses of CEA, the limitations of current methods, directions for revising these methods including the

development of generalized cost-effectiveness analysis (hereafter called GCEA), and some of the remaining technical challenges facing this revision.

1.2 Two sectoral uses of cost-effectiveness analysis

The appropriate methods, transferability of results and policy applicability of CEA depend critically on the intended use. CEA can have many applications beyond informing health sector resource allocation decisions across interventions[2] but the focus of this section is on two potential applications. These will be outlined briefly, after which the strengths and weaknesses of current methods of undertaking CEA will be discussed in relation to the two uses.

First, *CEA of a wide range of interventions can be undertaken to inform a specific decision-maker. This person faces a known set of resource constraints (hereafter called a budget), a set of options for use in the budget, and a series of other (ethical or political) constraints.* The set of constraints in this highly context-specific use of CEA for sectoral decision-making will vary tremendously from setting to setting. A decision-maker may be able to reallocate an entire budget or only allocate a budget increase; the decision-maker might be a donor, a minister of health, a district medical officer, or a hospital director.

Choices available, at least in the short- to medium-term, might be limited by factors such as the currently available physical infrastructure, human resources or political considerations. For example, in systems with substantial public provision there is a relatively fixed stock of hospital beds that cannot be increased or decreased easily.

Decisions could also be constrained by the current mix of interventions that are delivered; perhaps for political reasons specific interventions may not be reduced or eliminated without providing some alternative for that class of health problem. The set of constraints facing a decision-maker defines the decision space or the set of possible options from which choices can be made.

Second, *CEA of a wide range of interventions can be undertaken to provide general information on the relative costs and health benefits of different technologies or strategies which contribute through multiple channels to a more informed debate on resource allocation priorities.* Such general information should be seen as only one input into the policy debate on priorities. Because it is not meant to provide a formulaic solution to resource allocation problems it need not be highly contextualized. This general approach will contribute to judgements on whether interventions are highly cost-effective, highly cost-ineffective or somewhere in between. Such general perceptions of relative cost-

effectiveness can have far-reaching and constructive influence on policy formulation, defining the set of options that are debated without defining the allocation of resources in a precise or mechanical fashion. An alternative way to conceptualize this more general use of sectoral CEA is that the results define the mix of interventions that would be health maximizing in the absence of any constraints on possible decisions except a finite resource constraint. That health maximizing mix of interventions, which does not pertain to any specific decision-maker, can be a useful starting point for evaluating the directions for enhancing allocative efficiency in a variety of settings.

The first use of sectoral CEA, to inform a given decision-maker in a specific context, is more likely than the second to be used in a formulaic way to determine resource allocation. In this case, the challenges of incorporating explicitly other social concerns are more pressing but efforts to incorporate legitimate context-specific social concerns into the calculation of cost-effectiveness through devices such as equity weights inevitably make the results more difficult to communicate to some decision-makers and to the public. Such efforts also decrease the transferability of results. At some point in the continuum of complexity, the goal of informing a given decision-maker in a specific context may become impossible because of the cost and time required to generate the information.

In some sense, there is a trade-off between making CEA information precise to a given context and the time and resources required for that contextualization. Our preference for the more general use of CEA is an indication of how we see the outcome of that trade-off. We believe that the more general use of CEA, to inform sectoral debates on resource allocation, is where CEA can make the greatest contribution to health policy formulation. Such analysis indicates the general directions for resource reallocation required to enhance allocative efficiency. The results can be weighed alongside other social goals and considered together with the other constraints on decision-makers which are inevitable in specific contexts. The more generalized approach will enhance transferability and will make it possible to provide useful, timely and affordable information on the health generating characteristics of interventions.

1.3 Intervention mix constrained cost-effectiveness analysis

Various attempts have been made to codify a standard practice for CEA *(13–30)*. These guidelines differ for certain technical assumptions such as standard discount rates, the treatment of unrelated medical costs or the valuation of health outcomes. The broad approach, however, is similar. Intervention costs and health benefits are evaluated with respect to

current practice so that the numerator in the cost-effectiveness ratio is the change in cost due to the application of an intervention and the denominator is the change in health benefit. The interpretation of results is only straightforward in cases where the intervention produces more health benefits at lower cost in comparison with current practice, in which case the intervention should always be chosen. In any other case, results need to be compared with the cost-effectiveness of other interventions, by the implicit or explicit use of league tables.

The results of this type of analysis suggest replacing a less efficient intervention aimed at a particular condition by a more efficient alternative aimed at that condition, or if current practice involves doing nothing, it might suggest adding a new intervention. It is not used to evaluate if the existing interventions against that condition are themselves worth doing, however. It takes as the starting point the fact that some intervention against that condition will be undertaken. For this reason, we will refer to this standard practice as intervention mix constrained CEA or IMC-CEA, that is, one where there is a constraint against eliminating interventions that are currently in place unless they are replaced by another intervention targeting the same disease or condition.

Interestingly, IMC-CEA as currently practised does not consider other possible constraints on decision-making. It is worth noting that the policy environment in which decision-makers come closest to facing a constraint to continue current practice (or expand benefits in areas where there are existing interventions) but face no physical infrastructure, human capital or other constraints is in the United States. The combination of service provision through the private sector funded by third party payers with the presence of ethical guidelines on standards of care encourages the adoption of most health-enhancing interventions.

To further demonstrate the advantages and disadvantages of standard cost-effectiveness methods, consider Figure 1.1 that depicts the costs and benefits of six mutually exclusive interventions.[3] Following standard practice, intervention costs are on the y-axis and health benefits on the x-axis. In this and subsequent diagrams, each intervention should be thought of as a national programme or policy which can be purchased at only the point on the figure shown.

If a population has purchased intervention a1 then IMC-CEA would evaluate the cost-effectiveness of interventions a2-a6 with respect to the origin set equal to a1—indicated by the light grey axes centred on a1. The cost-effectiveness of each alternative, compared to a1, is equal to the slope of the line from a1 to the new point, illustrated for intervention a2—this slope is labelled as $\alpha1\alpha2$. The incremental cost-effectiveness for moving from a2 to a4 is the slope $\alpha2\alpha4$. For reasons that will be discussed in detail below, the origin in Figure 1.1 has been set as the costs and health benefits in the absence of any of the interventions a1-a6, called the null set. The line joining intervention a2 to the origin is the incremental cost-effectiveness with respect to doing nothing, labelled simply $\alpha2$.[4]

Figure 1.1   Costs and benefits of six mutually exclusive interventions
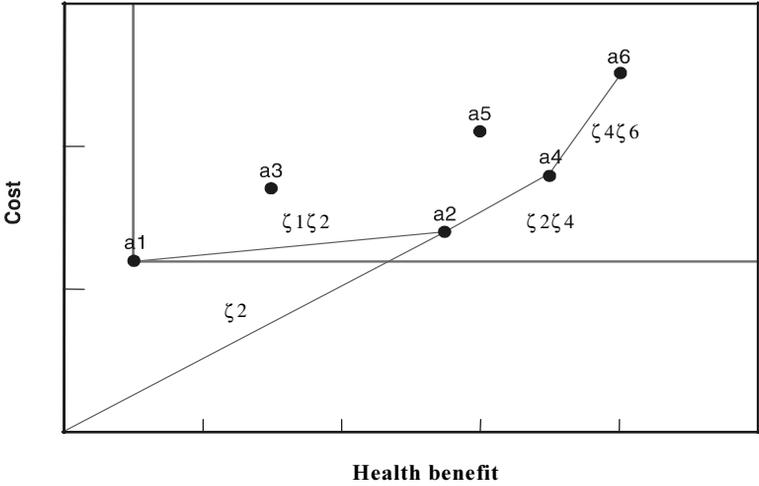
Health benefit

Figure 1.2 will be used to illustrate one of the main limitations of IMC-CEA. Eleven interventions are shown divided into three sets of mutually exclusive interventions, a1-a4, b1-b3, and c1-c4. Costs and health benefits for each intervention are shown with respect to the null set of this set of 11 interventions; health benefits could be denominated in DALYs averted, QALYs gained or some other physical measure of population health. A line from the origin to any individual intervention shows the

Figure 1.2   Costs and benefits of three sets of mutually exclusive interventions
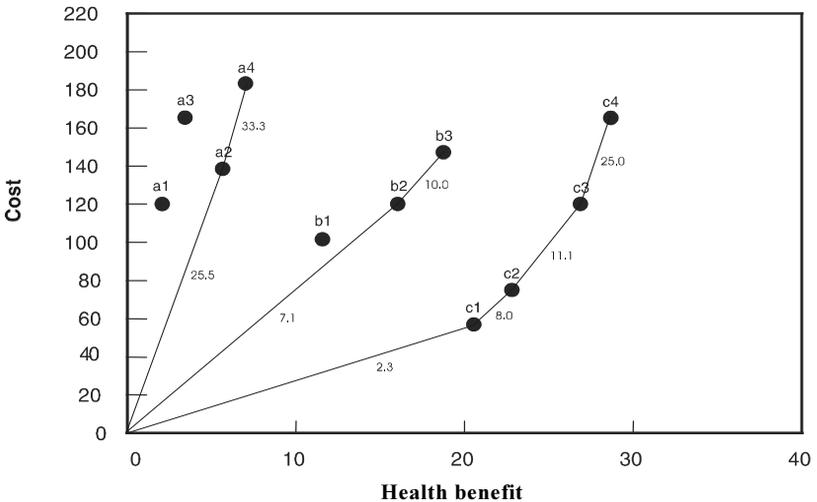
Health benefit

costs and health benefits compared to doing nothing—or the absence of any of these interventions. Table 1.1 provides the costs and benefits for each intervention and the incremental cost-effectiveness of each with respect to doing nothing, hereafter called the null set.

Table 1.1   Incremental CERs for 11 interventions

Intervention	Costs	Health benefits	Incremental CER compared to the null
a1	120	1.0	120
a2	140	5.5	25.45
a3	170	3.0	56.67
a4	190	7.0	27.14
b1	100	12.0	8.33
b2	120	17.0	7.06
b3	150	20.0	7.50
c1	50	22.0	2.27
c2	70	24.5	2.86
c3	120	29.0	4.14
c4	170	31.0	5.48

Consider a population where a budget of 170 is currently spent to purchase a1 and c1 producing 23 units of health. Next, consider an increase in the budget from 170 to 190. The remaining set of mutually exclusive interventions with respect to a1 would be evaluated. It shows that a3 is dominated and yields the incremental cost-effectiveness ratios in Table 1.2, which also shows similar calculations for the independent sets of interventions. A decision-maker would choose to purchase a2 instead of a1 because moving from a1 to a2 has the lowest incremental cost-effectiveness ratio. The final combination of a2 and c1 yields 27.5 units of health.

Table 1.2   Sequential incremental CERs starting from a1, c1

Category "a"		Category "b"		Category "c"	
$\Delta C/\Delta E$		$\Delta C/\Delta E$		$\Delta C/\Delta E$	
a2	4.4	b1	8.3	c2	8.0
a3	dominated	b2	7.1	c3	11.1
a4	33.3	b3	10.0	c4	25.0

   Consider another population where a budget of 170 is currently spent on a3 yielding only 3 health units. In this population, incremental cost-effectiveness analysis of the remaining interventions with respect to the starting point of a3 would yield the ratios in Table 1.3. If the budget now increases from 170 to 190, the decision-maker would first choose to save money and increase health output by moving to a2. With the savings of 30 and the increased budget of 20, the next most attractive intervention would be to purchase c1 with the resulting allocation of resources being a2 and c1 yielding 27.5 units of health.

Table 1.3   Sequential incremental CERs starting from a3

Category "a"	Category "b"	Category "c"
$\Delta C / \Delta E$	$\Delta C / \Delta E$	$\Delta C / \Delta E$
a2 12.0	b1 8.3	c1 2.3
a4 33.3	b2 7.1	c2 8.0
	b3 10.0	c3 11.1
		c4 25.0

   In both examples, IMC-CEA analysis identified health-enhancing resource allocations but the basic fact that the C and B category interventions are much more cost-effective than the A category interventions does not emerge from the analysis.
   This is because the cost-effectiveness of the starting point is not evaluated in current practice. As detailed below, it is relatively straightforward to identify the health maximizing combination of interventions for a budget of 170 as c1 and b2 which yields 39 health units and the health maximizing combination of interventions for a budget of 190 is c2 and b2 yielding 41.5 health units. In reality there is likely to be substantial allocative inefficiency in current allocations of health resources in many settings, and this example demonstrates that the application of IMC-CEA may fail to identify major opportunities for enhancing the overall cost-effectiveness of the health system.
   The intervention mix constraint on CEA means that major allocative inefficiencies may not be evaluated and thus not identified. If the current intervention mix is an unavoidable constraint on decision-makers in a given context then this is appropriate for context-specific CEA analyses. In most situations, however, other constraints on decision-makers may be more pervasive. As described above, in many health systems with a large share of public provision there is a fixed stock of community and referral hospitals that cannot be modified in the short- to medium-term for powerful political reasons.

Likewise, in many countries the supply of different types of health provider (nurses, general practitioners, specialists or community health workers) may limit the set of interventions that can be delivered. These decision constraints may be more common than the strict commitment to the mix of interventions that is assumed in current practice—it may be easier to shift spending from the treatment of ischaemic heart disease to childhood immunization programmes than to shut district hospitals or import ophthalmologists.

If the focus of sectoral CEA is to inform context-specific decision-making, then methods need to be developed to incorporate these and other constraints on the set of possible decisions. In that case, the choice of interventions cannot be guided through the use of league tables but more complex optimal resource allocation planning models are required *(3;31–33)*.

1.4 GENERALIZED COST-EFFECTIVENESS ANALYSIS

For some decision-makers, the development of complex resource allocation models that explicitly incorporate a range of decision constraints and multiple objectives may be useful. But such efforts are information intensive, time consuming, costly and very often difficult to communicate to the full set of actors in any health policy dialogue *(34;35)*. We believe that CEA can be most useful with more modest goals by focusing on the more general use of cost-effectiveness information to inform health policy debates without being completely contextualized. Moreover, sectoral CEA should identify current allocative inefficiencies as well as opportunities presented by new interventions. For this reason, we propose a modification of the standard IMC-CEA lifting the constraint on the current mix of interventions to evaluate the cost-effectiveness of all options, including currently funded interventions.

In brief the basic modification can be summarized in two propositions.

1) The costs and benefits of a set of related interventions should be evaluated with respect to the counterfactual of the null set of the related interventions. This was illustrated in Figure 1.2 for the 11 interventions. This provides the complete set of information for evaluating both independent and mutually exclusive options to identify the health maximizing combination of interventions for any given budget.
2) Results of CEA should initially be presented in a single league table as the first step of policy analysis. Subsequently the decision would be made about the appropriate cut point for classifying interventions as very cost-effective, very cost-ineffective and somewhere in between, as described earlier. Within the set of cost-effective interventions, the relative size of the cost-effectiveness ratio would be of little importance to the policy debate.

In constructing the league table, for each set of mutually exclusive interventions, the intervention with the lowest cost-effectiveness ratio (the lowest slope in Figure 1.2) with respect to the null set should appear first in the league table. The second intervention from the set (if there are at least two) is the one with the lowest incremental CER compared to the first intervention—the lowest slope with respect to the intervention with the lowest CER. The third intervention is the one with the lowest slope with respect to the second intervention, etc. Weakly dominated interventions should not appear in the league table.[5]

The results for all sets of mutually exclusive interventions are shown in the same league table according to the same principles. The application of this simple approach to the 11-intervention example in Figure 1.2 is shown in Table 1.4. Interventions a1, a3 and b1 are weakly dominated and do not appear. For heuristic purposes, the health maximizing combination for any budget level can be selected from the table. These decision rules are similar to those that have been derived for IMC-CEA but the analysis starts from the origin *(36)*.

Table 1.4   Generalized cost-effectiveness league table

Intervention	Cost-effectiveness ratio
c1	2.3
b2	7.1
c1 → c2	8.0
b2 → b3	10.0
c2 → c3	11.1
c3 → c4	25.0
a2	25.5
a2 → a4	33.3

By analysing the costs and benefits of sets of related interventions with respect to the null set of those interventions, the results are likely to be more transferable from one population to another. Clearly, the costs of different resource inputs to the production of a given intervention vary across populations as do some of the determinants of effectiveness. But one major factor limiting the transferability of IMC-CEA results from one population to another, namely different current mixes of interventions, can be removed by using the GCEA approach. To put it another way, the null set for a group of related interventions is more comparable across populations (or at least sets of populations) than the current mix of interventions. Nevertheless, there are clear limits to the

comparability across populations of the counterfactual null set. It will depend on the development of the health system and on the epidemiological pattern. Clearly, global comparisons of the cost-effectiveness of interventions with respect to the null set, even if input costs and effectiveness parameters are adjusted, are unlikely to be useful. This is the reason for WHO's decision to estimate and report cost-effectiveness for 14 epidemiological subregions.

The strategy for the development of this idea will be to define a limited set of average health system and epidemiological contexts within which null set comparisons are likely to be informative. Many groupings of countries or communities could be developed on the basis of income per capita, region, public/private splits in health care finance or provision, burden of disease, etc. Choosing the appropriate basis for grouping will be one major challenge for the development of this approach.

The benefits of analysing the costs and health benefits of interventions with respect to the null set for a group of related interventions appears to be great but the technical challenge of estimating the conditions in the null set counterfactual needs to be addressed. In theory, in IMC-CEA, costs and benefits of each intervention are evaluated with respect to the current mix of interventions. However, many studies are based on retrospective analysis where the intervention cost and benefits are evaluated with respect to a past mix of interventions that are not necessarily the current mix. Likewise, estimates of benefits of interventions that involve a time lag between purchase and benefit, such as hepatitis B immunization, are based on relatively implausible assumptions that the current mix of interventions will apply in the future. A symptom of this problem is demonstrated by the standard practice in IMC-CEA of estimating the benefits of life saving interventions using period life tables when in fact the cohort life expectancy at each age would be a more accurate (but more difficult to estimate) estimate of the years of life gained. Cohort life expectancy in some settings might be 10 to 15 years higher at birth than period life expectancy so that this is not a minor bias *(37)*.

Although the GCEA approach seems preferable to the identification of the efficient mix of interventions, IMC-CEA is still required as an additional analysis once the efficient mix of interventions has been identified. IMC-CEA is then used to provide information on how this efficient mix of interventions can best be achieved, starting from the current mix. In the above example, it was shown that the current mix of interventions a1 and c1 was not optimal in terms of health gains for a budget of 170, but that health resources should be reallocated to b2 and c1 in order to maximize health (Figure 1.2). In practice, this reallocation of resources can entail a complex process of taking resources away from a1 (perhaps a high technology surgical intervention) to allocate them to b2 (perhaps preventive care). Where GCEA is instrumental in defining regionally efficient mixes of interventions as a long-term goal for

decision-makers, the actual optimal pathway towards this goal is context-specific and could benefit from context-specific IMC-CEA. The current set of guidelines relates only to the use of GCEA to define regional efficient mixes of interventions.

1.5 INFORMING DECISION-MAKERS

Wider use of cost-effectiveness studies to analyse the allocative efficiency of health systems and recommend resource allocations has led to a number of challenges. It appears that the field can develop in two distinct directions, towards increasingly contextualized analyses or towards more generalized assessments. Cost-effectiveness studies and the sectoral application of CEA to a wide range of interventions can become increasingly context specific—at the individual study level by directly incorporating other social concerns such as distributional weights or a priority to treat the sick and at the sectoral level by developing complex resource allocation models that capture the full range of resource, ethical and political constraints facing decision-makers.

We fear that this direction will lead ultimately to less use of cost-effectiveness information in the health policy dialogue. Highly contextualized analyses must by definition be undertaken in each context; the cost and time involved as well as the inevitable complexity of the resource allocation models will limit their practical use. The other direction for sectoral cost-effectiveness, the direction that WHO is promoting (see Annex A), is to focus on the general assessment of the costs and health benefits of different interventions in the absence of various highly variable local decision constraints. A generalized league table of the cost-effectiveness of interventions for a group of populations with comparable health systems and epidemiological profiles can make the most powerful component of CEA readily available to inform health policy debates. Relative judgements on cost-effectiveness—e.g. treating tuberculosis with the DOTS strategy is highly cost-effective and providing liver transplants in cases of alcoholic cirrhosis is highly cost-ineffective—can have wide ranging influence and, as one input to an informed policy debate, can enhance allocative efficiency of many health systems. Information on GCEA can be used alongside consideration of the effect of different resource allocations on other important social goals such as equity. Because we believe this is the most constructive use of cost-effectiveness information, WHO is proposing to modify standard cost-effectiveness methods. The modifications proposed, that is to remove the current intervention mix decision constraint, will expose current allocative inefficiencies to analysis and at the same time enhance the transferability of results from one population to another.

For some narrower applications of CEA, such as the appraisal of new drugs in a specific country where it is not possible to change the mix of interventions currently provided, the currently practised IMC-CEA

remains appropriate. Nevertheless, even in these circumstances it would be useful for analysts to also estimate the costs and health benefits of interventions with respect to the null set. This would not only help to build a picture of the most efficient mix of interventions in the local context if policy-makers were able to reallocate resources, but would also substantially improve the world's body of knowledge on the cost-effectiveness of different interventions. In this way, each new study would add to our collective knowledge of the relative costs and effectiveness of different interventions.

2 Undertaking a study using GCEA

Before undertaking CEA, analysts must make a number of key decisions that have implications for the estimation of costs and health effects, and that collectively define and describe the work to be undertaken. In this section, we discuss the theoretical framework of analysis, the definition of interventions, the concept of the "null" or counterfactual, and the choice of intervention implementation period.

2.1 The theoretical framework of analysis

A variety of methods have been used in applied cost-effectiveness studies to estimate the costs and effects of different interventions. It is, therefore, difficult for policy-makers to know whether differences in reported costs and effects are truly due to differences in the efficiency of interventions, or whether they are simply a result of differences in the methods used by the analysts. For this reason, numerous attempts have been made to standardize practice for CEA *(13–30)*. Despite this, a number of controversies about the correct methodological approach to CEA remain and a variety of methods can still be observed in practice. Many of the remaining controversies relate to which items to include as costs and how to value them—for example, whether and how to include the impact of interventions on economic production, the time of informal care givers, and the costs incurred in extra years of life gained by an intervention *(18;38–41)*. There is unlikely to be full agreement between economists on all these issues in the near future, yet it is important for policy purposes to ensure that the results of cost-effectiveness studies are as comparable as possible. In this respect, a discussion of the theoretical foundations of CEA can help to identify which recommendations for standardization have a strong theoretical basis and which are based on practical or

pragmatic considerations *(18;39)*. This is the reason for briefly reviewing the theoretical foundations here.

This is a guide on CEA. In doing this, we are clearly not choosing the path of cost-benefit analysis in which social welfare or well-being is the aggregation of individual utility or well-being, and individual utility is a function purely of the consumption of goods and services *(42;43)*. In that approach, the value of additional consumption to an individual—including the consumption of health and medical services—can be measured using the individual's willingness to pay (WTP) for it and the question of whether an intervention should be undertaken becomes one of determining if aggregate WTP exceeds costs. In its standard application, cost-benefit analysis requires a number of strong assumptions. Among other things, consumers have perfect information on the consequences of their consumption choices. Only then does the expression of their WTP reflect the value to them of the proposed consumption. This is rejected here largely on the grounds that the assumptions required for WTP to be meaningful do not apply in health. For example, there is virtual unanimity in the literature that people do not have the information or training necessary to value the benefit to them of using a particular health service or intervention *(44–47)*. This is the reason they seek the input of health professionals to act as their agent in making treatment decisions.[6] In most applications of CBA, potential Pareto improvements are identified where the winners would in principle be able to compensate the losers. When such redistributions between the winners and the losers do not occur, choosing social decisions with a positive cost-benefit ratio would not necessarily increase social welfare and could in fact decrease it. CBA is often associated with the term money metric utility, where in practice individual's utility is measured in dollars.

Accordingly, WTP cannot be used in practice to value the benefits of health interventions. CEA is used, therefore, in preference. CEA is based on the belief that health contributes to social welfare separately to the consumption of non-health goods and services. If it is assumed that the budget for health has already been decided in some way, the simplest form of analysis would be to take the perspective of a single benevolent decision-maker or provider, who seeks to maximize population health subject to the resources that are available. The decision-maker focuses only on the resources under their direct control and has a fixed time horizon. The only relevant costs are those met from the fixed health budget *(38)*.

This approach, sometimes called the decision-maker's approach *(38)*, is too narrow for our purposes. By focusing only on the budget strictly under the control of the decision-maker, it is not consistent with WHO's concern that governments should be stewards of the entire health system, seeking to ensure that all health resources, no matter who controls them, should make the maximum possible contribution to a set

of key social goals. It is also difficult to measure costs in this context. Economists think of the costs of using resources for a particular activity as being the benefit foregone because the resources were not used in the next best alternative—the opportunity cost. Strictly speaking, opportunity cost in this case would be the foregone health involved in using the given resources in one type of health intervention rather than another. Only under very restrictive assumption of markets working perfectly in health and non-health would money values such as market wages reflect this opportunity cost.

In addition, as soon as it is allowed that health improvements can affect income, thereby influencing the budget available for health expenditure, questions of how much of the additional income should be allocated to health must be addressed. This requires considering the opportunity cost of expenditure on health in broader terms, taking into account that the resources could also have been used to produce non-health consumption.

Most CEA guides, therefore, argue that a "social perspective" should be taken when estimating costs *(13–30)*. The social perspective is taken to mean that all costs should be included regardless of who pays them, and resources used or created by health interventions should be valued at the benefit foregone because society could not use the resources in their next best use. This next-best use might be in health or non-health.

Implicitly or explicitly, therefore, the assumption of CEA is that non-health consumption and health both contribute to welfare. Health is important for its own sake, but in measuring the opportunity costs of using resources for a health intervention it is important to take into account the fact that they could have been used for non-health consumption.

All CEA requires the effect of an intervention to be measured against the counterfactual state of an alternative intervention being undertaken. In GCEA, this alternative is the null, or the situation in which the intervention did not exist. Given that interventions effect welfare through their impact on non-health consumption and on health, CEA requires both effects to be measured for the intervention and the counterfactual. Both types of effect should be taken into account, regardless of the time period in which these effects occur. This should include any changes in non-health consumption resulting from changes in health.

Distributional issues are also important if the social perspective is to be strictly applied. It is generally accepted that the utility or well-being an individual gains from an extra unit of consumption declines as consumption increases—reflected in the concavity of utility functions. This means that who pays the costs of an intervention and who receives any increases in non-health consumption also influence social welfare. A poor person has, by definition, lower levels of consumption than a rich

person, so would lose more well being by being asked to contribute $1 to the health system than would the rich person.

Related to this is the consideration that if improved health results in improved capacity to work, which results in additional production, and thereby to additional non-health consumption, the value of this additional consumption varies across people. It means more to poor people than to the rich. On the other hand, productivity also differs across people, so improving the health of one person might lead to greater increases in non-health consumption opportunities than improving the health of someone else. To adequately capture the total welfare effect of an intervention through its impact on health and non-health consumption requires knowledge of each person's marginal product, who benefits from the additional consumption, and the current level of consumption of non-health goods and services of those who benefit.

These distributional questions are related purely to the fact that the increase in welfare associated with increments of consumption depends on the current level of consumption (i.e. the concavity of utility functions) and to differences in the marginal productivity of labour. They are even more important if societies have "inequality aversion" *(48)*. If people are offered a choice between two states of the world in which total consumption was the same, but in one it was distributed more equally across the population than in the other, most societies are likely to prefer the first option. This would provide an additional reason to weigh resources used and created according to who contributes and receives them.

Although the separation of the welfare effects of health interventions into those related to health and those related to non-health consumption provides a theoretically consistent framework, there remain a series of pragmatic barriers in applying this to CEA in practice. The effect of interventions on welfare through health and through non-health consumption, taking into account who contributes resources and gains additional units of consumption, is not typically done in CEA for a variety of pragmatic reasons *(18;49)*. We make similar compromises, explained in the subsequent sections on costs and health effects. However, the guiding principle should be that it is important to measure the impact of an intervention on welfare through its impact on health as well as through its impact on non-health consumption. In essence, the benefits of an intervention should be the welfare gain resulting from any health improvement, while the costs represent the welfare foregone because the resources could not be used in the next best use—health or non-health consumption.

2.2 Defining health interventions

A key issue in GCEA as well as in any other CEA is defining an intervention. The term "intervention" is used in its broadest sense here.

It includes any use of resources aimed at improving health outcomes be they preventive, promotive, curative, rehabilitative or palliative. It includes clinical care and public health programmes and strategies.

Many interventions interact in terms of either costs or effects at the population level. The health impact of undertaking two interventions together is not necessarily additive, nor are the costs of the joint production. To understand whether they are efficient uses of resources independently or in combination requires assessing their costs and health effects independently and in combination. For example, the health effects of the introduction of insecticide treated nets in malaria control is likely to be dependent on whether there is ongoing residual spraying of houses with insecticide. The costs of introducing early case detection and management for malaria will depend on whether insecticide treated nets are widely used. It is not very useful for policy-makers to be given information on the cost-effectiveness of nets alone, or treatment alone, when interventions are rarely undertaken in isolation. Accordingly, interventions that interact should be evaluated as a group (see Annex B). In the malaria case the individual interventions should be evaluated alone, but then in combination with the others, producing a more policy-relevant set of estimates of cost-effectiveness.

It is important to note that traditional incremental analysis would emerge from this. For example, the cost-effectiveness of adding nets to a setting in which there is already residual spraying of houses would emerge from the comparison of the residual spraying alone option with the combination of residual spraying and nets.

The case of mutually exclusive options is similar, i.e. interventions which by definition cannot be implemented simultaneously in the same population. An example is population-based annual and biannual breast cancer screening. These interventions must be evaluated as part of the same set which will ensure that only one of the interventions appears in an optimal mix.

Figure 2.1 illustrates interactions using hypothetical data for a cluster of interventions for tuberculosis: passive case detection and treatment with DOTS (a), Bacille Calmette-Guérin (BCG) vaccine at 50% coverage (b1), BCG at 75% coverage (b2) and BCG at 100% coverage (b3). The data are reported in Table 2.1. In addition, three other mutually exclusive options are presented, passive case detection and treatment with DOTS combined with the three different levels of BCG coverage (respectively ab1, ab2, ab3). Costs interact. For example, if BCG is delivered, the number of cases of tuberculosis that will occur, be detected and accept treatment will be smaller so that the variable cost component of the treatment programme will decline but the fixed cost component will not. Likewise, the health benefits of BCG in the presence of a treatment programme will be less because many of the deaths from tuberculosis expected in the absence of treatment will be avoided.

Figure 2.1 Costs and health benefits of interventions with cost-effectiveness interactions
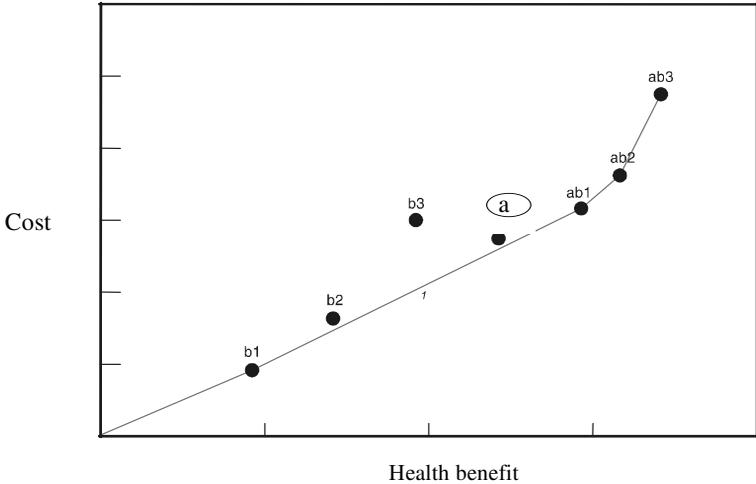
The interaction of the benefits of the two programmes can be estimated, here using a multiplicative model. In Figure 2.1 the cluster of interventions—including each individual intervention and the possible combinations—is depicted and can be used to develop a league table. The intervention with the smallest slope (BCG at 50% coverage – b1) is the most efficient and should be done first if funds are available. The slope from b1 to any other point should then be assessed and the lowest slope chosen, and the process repeated. This results in the following sequence of choices for this set of mutually exclusive interventions: BCG at 50% coverage (b1), BCG at 50% coverage combined with passive detection and treatment (ab1), BCG 75% with detection and treatment (ab2) and BCG 100% with detection and treatment (ab3). The lines connecting these points form the "expansion path", revealing the mix of interventions that would be chosen on cost-effectiveness grounds for any given level of resource availability.[7]

BCG at 75%, BCG at 100% and passive detection and treatment alone do not appear in the list because they are dominated by other alternatives. They lie to the north east of the expansion path in Figure 2.1 showing that they are both more expensive and less effective than one of the options that lies on the expansion path.

The decision rules, therefore, require all interventions and combinations to be evaluated initially compared to doing nothing. Subsequently, incremental analyses are built on top of the most cost-effective option. It is clear from the TB example that CEA is a valuable tool for assessing the appropriate approach to control a specific health problem in addition to being valuable at the sectoral level.

Table 2.1   Costs, health benefits and cost-effectiveness ratios of a set of interrelated interventions

Intervention	Costs ($)	Health Benefits	CE Ratios
a	550 000	500	1 100
b1	180 000	200	900
b2	325 000	300	1 083
b3	600 000	400	1 500
ab1	631 000	600	1 052
ab2	726 500	650	1 118
ab3	952 000	700	1 360

In the literature on cost-effectiveness there has been considerable concern about non-linear cost-effectiveness functions *(33;50;51)*. For example, the cost per disability-adjusted life year (DALY) averted through the expansion of measles coverage from 50% to 90% is likely to be much lower than the cost per DALY averted through the expansion of coverage from 90% to 99%. Because interventions at different levels of coverage are mutually exclusive at the population level, then the same approach outlined above can be used to capture a non-linear cost-effectiveness function in a series of discrete points.

In Figure 2.1, the set of interventions ab1-ab3 could be different strategies or different levels of coverage for the same strategy. By picking a parsimonious set of coverage rates, a set of independent and mutually exclusive interventions can be defined capturing the key consequences of non-linear cost-effectiveness functions in a single league table as described above.[8]

Defining the intervention

It is critical to describe the "intervention" accurately using all information which is essential to interpret the estimated costs and benefits. This includes the strict definition of the treatment pathway for clinical interventions, which procedures are incorporated and which are not. This approach combats the problem sometimes encountered in current CEA where interventions are evaluated "as implemented", comprising a wide variety of possible treatment pathways and outcomes. This makes the interpretation of results difficult, since this variety of treatment pathways is often context-specific and does not provide information on the relative cost-effectiveness of each of the alternative treatment options. Alternative treatment pathways in a disease area should be represented by the analysis of separate interventions.

> The definition of an intervention should include information on the setting where the intervention is delivered or undertaken (site of delivery, e.g. facility or community-based and level of care, e.g. primary, secondary or tertiary); the target population covered by the intervention; the time frame of the cost data included; the regimen of therapy (in curative interventions); the frequency of delivery of the intervention (e.g. for screening); the extent of coverage of the target population; and any other important information.
>
> For example, a definition of a programme to deliver directly observed short course therapy (DOTS) for a newly diagnosed tuberculosis patient might require the following additional information: at 95% geographic coverage, diagnosis of symptomatic cases presenting to government health facilities by detection of acid fast bacilli at least twice in initial sputum smears (or other specified diagnostic criteria) and treatment of smear-positive cases with directly observed chemotherapy (three times weekly) using fixed drug combinations. The regimen consists of a two-month intensive phase of Rifampicin, Isoniazid, Pyrazinamide and Ethambutol followed by a four-month continuation of Rifampicin and Isoniazid, all at recommended doses.

2.3 DEFINING THE COUNTERFACTUAL

2.3.1 INTERPRETING THE NULL

GCEA requires the analyst to consider what would happen, starting from today, if all resources in the health sector could be reallocated. The counterfactual against which all interventions should be evaluated is what would happen if none of the current set of interventions were implemented. The cost-effectiveness of all possible interventions—individually and in combination—is assessed in relation to this counterfactual, or null. The next question is how to define "did not exist". It is not possible to estimate the current levels of population health in every setting assuming that none of the current or past interventions had ever been undertaken. It is possible, however, to estimate what would happen if all current interventions ceased forthwith, that is, what would happen if they were eliminated today. This is the counterfactual proposed in these guidelines.

This does not necessarily reflect an equilibrium situation with a stable prevalence of disease. Rather, because of the sudden absence of interventions ("what would happen if all current interventions were eliminated today?"), hazard rates change and cause the epidemiological situation to be out of balance; eventually, prevalence of disease will adapt to the null hazard rates and the equilibrium will be restored. The time lag

involved, however, will differ per intervention-cluster and may pass the time horizon of analysis. The null, therefore, does not represent a stable epidemiological situation, but a transition of the epidemiological profile of disease over time.

The epidemiological profile associated with the null depends on the past history of interventions and country-specific factors such as climate. For example, where childhood immunizations have achieved high rates of coverage historically, the present epidemiological profile will be characterized by a relatively low prevalence of vaccine-preventable disease. Since the present epidemiological profile is the starting point for the null, the null will be also characterized by a relatively low initial prevalence of disease.

In addition, the ability of countries to improve health varies according to such factors as the current availability of infrastructure within and outside the health sector. This includes the existing number and training level of health personnel which will influence the effectiveness of interventions. It also includes the availability of roads, which determines the costs of expanding coverage, and the education level of the population which could affect the current epidemiological profile of the population and the effectiveness of interventions.

The definition of the null is likely to vary across populations as does the ability of countries to improve health. In the ideal case, separate analyses should be done for every country, and possibly for every sub-population within a country. It will not be possible to undertake the sheer volume of work required to assess the costs and effects of all possible interventions, for all subpopulations in the world, in the foreseeable future so populations that are relatively homogeneous are analysed together. This is the reason why the WHO-CHOICE project conducts its analysis for 14 epidemiological subregions which are similar in terms of starting points.

2.3.2 CAN PARTIAL NULLS BE USED?

In defining the null, it is not practical or necessary to assume that no interventions at all exist. Since it is clear that the costs and health effects of many interventions are unaffected by the existence of others, a key issue in GCEA is to define clusters of interventions that are interrelated either because they interact on either costs or effects or because they are mutually exclusive (see Section 2.2). The null, therefore, is defined as what would occur if a specific group of interrelated interventions were eliminated today. Given this definition of the null as that state in which groups of interrelated interventions no longer exist, different so-called "partial nulls" can be defined within the GCEA framework. These partial nulls are defined by the elimination of a specific group of interrelated interventions, while interventions in other groups of interrelated interventions continue to exist but are assumed to have no impact on incremental costs and health effects of the interventions under study.

For example, while the overall population health effect of introducing insecticide treated nets will depend on whether residual spraying of houses with insecticide takes place at the same time, it is unlikely to be effected by whether preventive or curative services for cardiovascular disease exist. Accordingly the analyst can evaluate the set of interventions for reducing the burden associated with malaria independently of those focusing on cardiovascular disease control. When defining the null for the set of malaria interventions, it is not necessary to assume that no interventions for cardiovascular disease exist, and vice versa. In defining the partial nulls, it is simply necessary to ensure that the costs and effects of an intervention in one intervention cluster are largely independent of the existence of interventions outside that group.

Defining partial nulls

In some areas, costs and health effects of a large number of interventions are interrelated, if remotely, and pragmatic boundaries need to be drawn. For example, costs and health effects of a cluster of interventions in diarrhoea control are interrelated with a large number of other interventions, such as vitamin A supplementation, oral rehydration therapy (ORT) and breastfeeding. Whether the benefits of evaluating oral rehydration therapy and breastfeeding in a single framework outweigh the extra analytical effort is a pragmatic decision which depends on the believed strength of the relationship between the different types of interventions.

2.4 IMPLEMENTATION PERIOD AND TIME HORIZON OF ANALYSIS

The time period over which to evaluate an intervention is a vexing question *(17;18;52;53)*. Many interventions are evaluated for a typical year, with start up and capital costs annualized. This would bias choices against programmes which take a number of years of activity at full scale to start producing benefits, such as health education programmes. Yet the time horizon of health decision-makers is probably rather short. As a balance, we suggest that GCEA should evaluate all interventions over a period of 10 years at full implementation.[9] That means that the costs are those of full implementation, and include annualized start-up costs prior to these 10 years, and total costs during the 10-year period. Depending on the intervention, annual health effects may be identical each year during the 10-year period (likely for curative interventions) or may gradually increase each year (likely for preventive interventions).

It is important to note that the 10-year period refers only to the period that the intervention is implemented. The time-horizon for the analysis is

obviously longer: analysis must include all the health effects of the intervention that occur because of the 10 years of activity, whether these benefits accrue during the 10 years or subsequently; this will frequently involve modelled data. A classic example is immunization of infants against hepatitis B with the primary aim of preventing liver cancer. In the analysis, yearly cohorts of infants would be immunized over a 10-year period, but the expected impact on liver cancer would not begin to emerge for approximately 35 years later.

SUMMARY OF RECOMMENDATIONS

1. Groups of interventions where there are major interactions in either costs or health effects should be evaluated together.
2. Analysts should evaluate all interventions initially against the "null", i.e. the situation that would exist if none of the set of interacting interventions were implemented.
3. Interventions should be described in detail, which includes information on the setting, target population, time frame, regimen, and frequency of obtaining the intervention.
4. All interventions should be evaluated under the assumption that they are implemented over a period of 10 years. However, costs and health effects related to the intervention should be followed for the duration of the lifetime of the beneficiaries. This could be varied by country-analysts adapting the results or undertaking studies in their own settings.
5. Resource use and health effects should be identified and valued from the societal perspective.

3 ESTIMATING COSTS

In Section 2.1 it was argued that the strict application of the societal perspective in CEA would require all effects of an intervention on welfare, through changes in non-health consumption and health, to be measured and included in the analysis. Distributional considerations would need to be incorporated, requiring identification of who contributes resources to the intervention, who is able to produce more because of improved health, and who gains the benefits of this increased production.

The benefit of a health intervention is the gain in welfare associated with the health improvement. The cost is the loss of welfare associated with the non-health consumption forgone because the resources are used to provide the health intervention. This section focuses on the questions of what changes in non-health consumption should be included in the numerator, how they should be valued to represent changes in welfare, and how the resulting estimates should be included in a cost-effectiveness analysis. The framework of Section 2.1 is used to identify what should be done in theory and then the pragmatic considerations that lead to modifications of the first-best approach are discussed.

3.1 IDENTIFICATION OF COSTS

3.1.1 DEFINING THE NULL WITH RESPECT TO COSTS

In applying GCEA, groups of related interventions are analysed with respect to the null set or the counterfactual of those interventions not existing. In theory, all costs relating to the interventions that are being analysed with respect to the null set would be zero. However, part of the overhead costs of ensuring that interventions take place relate to the availability of trained staff, and some level of central administration such as central stores, auditing, budgeting, etc. While it might be argued that

the investment in basic training of health personnel is reflected in their salaries, this is not true in many settings where government controls public sector wages and the private sector is not well established for many types of health personnel. Moreover, it is not practical to try to allocate the costs of a department of audit, for example, across all health interventions. For that reason, we propose that it should be assumed that these costs exist and will continue at the same level regardless of the different mix of interventions that are delivered. GCEA would focus on resources that could realistically be reallocated over the time horizon of the analysis.

Two major types of ongoing costs are identified. The first involves some of the costs of central administration, such as the overall planning and management of the health system that are unrelated to the development and implementation of particular interventions aimed at improving health. Some activities of a ministry of health, for example, would exist and have a certain staffing profile independent of any particular set of interventions that may be done in the country for the available resources.

The second type of ongoing costs relates to the current level of education of health professionals. If the skills required to deliver an intervention are not available (or not yet available to the full extent necessary) in the country under study, training costs to develop those skills should be included as part of the intervention costs. However, if those skills are already acquired, and no further training is required, the cost of the previously acquired training can be assumed to exist.

Accordingly, some types of administrative costs and those related to the formal education of health professionals have not been included in the costs of interventions for GCEA. These define the "starting point" for the analysis, and would vary across settings, which is one reason for undertaking the WHO-CHOICE analysis at a sub-regional, rather than a global level.

3.1.2 Costs of providing health interventions

The costs of providing health interventions—such as an outpatient visit, an inpatient stay or a population-based programme—are the resources used in making the intervention available. The resources include labour, capital such as building space and equipment, consumables such as medical supplies and medications and overhead costs such as electricity, water and maintenance. By using these resources to improve health, they cannot be used to produce other goods and services, thereby incurring a welfare loss.

Resources to fund health interventions can be financed in various ways, including taxation, insurance, and direct out-of-pocket payments by households. Out-of-pocket expenses with near perfect markets can be valued in dollar terms because the expenditure undertaken by consumers

will reflect the value they place on the purchased service, taking into account their budgetary constraints. Markets, however, are not even approximately perfect for most areas related to health and do not equilibrate payments with the perceived value of the service obtained. This is particularly true for tax payments and insurance contributions. In addition, as people consume more of a certain product, the value of each unit of additional consumption falls (known as diminishing marginal utility, illustrated by a concave utility function). This means that the welfare loss of a dollar contributed by a poor person is greater than that of a dollar contributed by the rich. Because of this, the resources used to provide health interventions should be evaluated in terms of the welfare loss associated with how the funds are raised, and not in simple monetary terms.

It could be argued that once the pool of funds for health has been raised, expenditure on each intervention is not subsequently linked to payment mechanisms. If this is true, the rank ordering of costs and cost-effectiveness would not be affected by different weights for each dollar contributed by the poor and rich within a particular country. This is not strictly true for situations where significant co-payments exist and where patients incur out-of-pocket payments, where the welfare loss should theoretically be estimated.

This type of analysis is not typically undertaken because of the difficulties involved in estimating the welfare loss of each dollar to each contributor. We also follow the traditional approach of measuring the costs of providing health interventions in money terms while recognizing the limitations.

3.1.3 Costs of accessing health interventions

The costs of accessing health interventions include the resources used by patients and their families to obtain an intervention. This does not include payments for the intervention itself, which are part of the costs of providing an intervention discussed above. Access cost has two components. The first consists of the resources used in seeking or obtaining the intervention—such as the cost of taking a taxi to a hospital, or special food for diets connected with therapy. The second involves *time costs* related to seeking or obtaining the intervention. This time has an opportunity cost in that it cannot be used to produce consumption in other areas so that, in theory, it should be valued and included in the numerator.

The way to treat the resources (as opposed to time costs) used in seeking or obtaining care is no different to the costs of providing an intervention discussed above. In theory it is necessary to identify the welfare losses associated with each payment made by each household, but this has proven intractable for practical CEA. Accordingly, we use the standard approach of measuring these costs in money metric terms. Time costs are related to production gains and losses, discussed below.

Many types of health interventions—such as rehabilitation, prevention or life-saving programmes—can affect the ability of people to work, and through this, the total resources available to society. In the social welfare framework, productivity costs or gains affect the consumption of goods and services and, therefore, social welfare. They should be included in the analysis.

The valuation of such changes is not straightforward. First, the impact of one unit of health on production will differ according to such factors as the age and occupation of the individual. Perhaps this can be measured in terms of the wages gained or lost if labour markets work relatively efficiently, although this is rarely the case. Secondly, the welfare effect of each dollar gained depends on each individual's marginal utility of an additional unit of consumption and this will be higher for the poor than the rich. Thirdly, each dollar gained may affect other people's welfare, within households and/or in society as a whole. For example, the related social welfare changes will be larger for productivity changes of rich people than for the poor in the presence of a progressive tax system, since rich people's tax contributions are higher and will add more to a society's welfare. Finally, where there is inequality aversion,[10] society would value a dollar of production gain by the rich to be less value than the equivalent gain accruing to the poor.

If the distribution of productivity gains or losses across diseases and interventions is not random, the inclusion or exclusion of these costs in CEA could affect the ranking of interventions. Even if it is possible to quantify the productivity changes resulting from gains in health by modelling the relationship between life-years and economic growth, as proposed by the Commission of Macroeconomics and Health *(54)*, it would still require weighting each dollar gained by who gains it. This has not yet been done in CEA.

Certainly market wage rates do not reflect the welfare gains from a health intervention that allows people to work longer or more productively. To do this would require firstly estimating the change in GDP over time with and without the intervention, something that requires heroic assumptions about trends in economic growth rates. Secondly, the value of non-market production with and without the intervention would need to be calculated over time—non-market production can be a large component of the economy in many poorer countries *(55)*. Thirdly, the question of who gains the benefits of increased production would need to be addressed. Any attempt to measure this welfare change in money terms would simply introduce noise into the calculations, so we recommend that changes in production are not included.[11]

In doing this, we recognize that the correct approach is to include the welfare effects of changes in production, but that no conceptually

appropriate way of measuring these welfare effects in money terms is yet available. Work to this end is continuing, but in the meantime, we suggest that where productivity gains are believed to be important, analysts should attempt to quantify them in physical units as rigorously as possible and report them separately.

This recommendation is identical to common practice, including the recommendations of the Washington Panel *(18)*, although for different reasons. Gold et al. argue that people take into account the impact of an intervention on their future production when providing utility weights for QALYs, so that the effects are implicitly included in the denominator of the CEA ratio. Accordingly it would be double counting to include them in the numerator as well. We do not believe that this is the case. DALYs, the summary measure of population health used in the WHO-CHOICE project, measure health and do not incorporate the welfare associated with any income-enhancing properties of an intervention. But even if QALYs are used, it is very debatable that the utility weights used in QALYs capture the effect of an intervention on people's future income. Meltzer *(56)*, for example, has shown that utility weights change if people are provided information on the effect of an intervention on their income. This is not usually done when eliciting utility weights.

The same argument applies to the time of informal caregivers[12] and the time patients and their families spend in seeking and obtaining care. To be consistent with the decision to take a social perspective in measuring costs, each unit of time would have to be weighted according to the different impacts of each person's production on overall output, and also according to who receives the benefit of the additional consumption. Gold et al. *(18)* argue that this time should be valued in terms of lost wages and included in the numerator of the CEA, but this misses the point that these effects should be measured in terms of the change in welfare.

Even a shadow wage rate is not strictly correct because it takes into account only the impact of time on market production rather than the welfare effect of this time. The alternative practice of using average GDP is not correct either because GDP divided by the workforce overestimates the marginal product of labour by a considerable margin.[13] There is no way of knowing how good an approximation of the welfare losses these alternatives are, and we argue that including time costs valued using any of the above methods simply adds noise to the calculations. It is not possible even to estimate the direction of the likely bias. Even knowing that this time would not be valued at zero, to be consistent with the treatment of production gains and losses, they should be excluded.

In general, time costs are unlikely to be substantial, but for certain conditions or interventions, their exclusion may introduce bias into the comparative estimates of cost-effectiveness, e.g. where people require repeated contacts with the system. A case in point would be treatment for a chronic condition, such as kidney dialysis, where the omission of time

costs would undervalue the attractiveness of home dialysis compared to a facility-based intervention. Accordingly, where they are likely to be substantial, the analyst should report them separately.

3.1.5 Health costs in extended years of life

Controversy surrounds the question of whether and how to incorporate health costs in extended years of life into CEA. On the one hand, Weinstein and Stason *(57)* argue that because interventions that extend life result in increased medical expenditures in the extended years, these additional costs should be added to the costs of the intervention in the numerator, even if they are unrelated to the intervention under consideration. On the other hand Gold et al. *(18)* draw on the model of Garber and Phelps *(58)* to argue that future health costs unrelated to the intervention under consideration can be excluded because the ranking of interventions will be the same whether they are excluded or included. The important point to emphasize is that the analyst should always be consistent in the choice whether or not to include them.

The theoretical framework outlined earlier suggests that the net change in non-health consumption over time as a result of the intervention should be measured, whether the change results from the production of additional resources due to healthier workers or from the additional use of resources because people live longer. A problem emerges, however, because of the need to value these changes in money terms. As in the case of changes in future production resulting because people live longer discussed earlier, society may value an extra dollar accruing to the poor more than a dollar accruing to richer people. There is no way to determine the relationship between net changes in non-health consumption valued in money terms and the resulting changes in welfare, so we recommend that they be excluded on the same grounds discussed above.

3.1.6 Joint or overhead costs

Joint costs are those resources that are shared with other interventions or programmes. To the extent that these resources could be reallocated to other activities, a share of their cost must be attributed to the intervention. Joint costs usually, but not always, constitutes overhead costs such as buildings, maintenance, electricity, water, etc. Other types of joint costs include personnel or equipment such as those involved in diagnostic tests, which are typically shared between several interventions. In practice, joint costs are estimated by applying some allocation rule related to the usage of the resource item *(17;59)*. For example, personnel costs can be allocated to the intervention on the basis of the proportion of time they devote to it. Vehicle costs could be allocated according to the proportion of the total distance travelled for each intervention, while building costs are typically allocated using the proportion of the space used by each intervention. Assumptions for allocating joint costs are often arbitrary but it is helpful to think of the particular component of

the resource that determines its distribution between interventions. For more discussion on the allocation of joint costs, see Creese and Parker *(59)*.

There are two levels of overhead costs that are related to the provision of an intervention. The first is where the intervention is actually delivered (e.g. hospitals) and the second is at higher levels of the organizational system, e.g. district and national administrative levels, referred to as programme costs in these guidelines. To be consistent, these costs should be traced and incorporated in the analysis.

3.1.7 COST OFFSETS OR RELATED HEALTH COSTS

For GCEA, interactions between interventions in terms of costs and outcomes are taken into account by defining combinations of mutually exclusive interventions as described above. For example, if administration of the BCG vaccine reduces the subsequent number of cases of TB requiring treatment, at least three mutually exclusive interventions would be defined at the population level: x% of children covered by BCG with no treatment of subsequent cases; treatment of all TB cases only; and x% coverage of BCG with treatment of subsequent cases.[14] If all are evaluated against the null, cost offsets directly linked to BCG vaccination—i.e. savings in treatment costs due to fewer subsequent cases—would automatically be included in the option of BCG vaccination combined with treatment.

Classification of costs

One of the first steps of any practical costing exercise is to identify the production process of the programme or intervention. There are several ways of classifying costs. For example, they can be classified by *input category* (e.g. salaries, medical supplies, capital), *intervention activity* (e.g. administration, planning, supervision) or *organizational level* (e.g. national, district, hospital). The most important point to consider when choosing a classification scheme is to make sure that all the relevant costs are included and that the classification categories do not overlap *(59)*.

In most interventions, different types of activities are involved at the programme level as distinct from the point of delivery, during both start-up and post start-up periods. **Programme costs** are defined here as those associated with the development and administration of an intervention, outside the point of delivery. They may be incurred at any level—e.g. national, district or provincial. Examples are planning, training, media and information activities, development of information, education and communication (IEC) materials, monitoring, some types of

supervision, and social mobilization. Annex C provides a detailed description of the types of costs that are relevant to each of these activities. Programme costs are often ignored in published studies *(60;61)* although they can account for a major component of the total costs of interventions.

It is important therefore that they are included in any CEA and WHO-CHOICE reports programme and patient costs separately. **Patient costs** are defined as any costs incurred at the point of delivery. They are usually associated with the delivery of curative care, but can also include certain types of health education and preventive activities. The former would include health education provided to women attending maternal and child clinics, for example, while the latter would include childhood immunization.

3.2 Cost valuation

This section discusses the main concepts and recommendations for the economic valuation of resources used in an intervention. Particular reference is made to issues related to developing countries. For a detailed presentation of the arguments and recommendations on resource valuation, see Hutton and Baltussen *(62)*.

3.2.1 Economic prices

In providing recommendations for valuation in cost-effectiveness analysis, in particular for resource-poor countries, it is important to keep in mind the need to develop approaches that can be applied widely in many settings, that do not have stringent data needs, and that can be applied by non-specialists who do not necessarily have an in-depth understanding of economic principles. However, it is important to decide first which methods are methodologically correct before asking if they can be applied in practice.

It is generally agreed that the economic definition of costs should be used in cost valuation, not the accounting (or financial) definition. This is based on the concept of "opportunity cost", i.e. here defined as the value forgone by not using the same resource in the best alternative activity. It is important to distinguish between prices on the one hand (usually determined by a market, but which also can be determined from other sources), and economic value on the other. It is well known that observed prices or charges do not necessarily reflect the economic value *(63;64)* as shown briefly in the following examples.

- In company or government accounts, buildings and equipment are depreciated over time, so that after a few years they have an

accounting value of zero. However, even after this period these items still have a cost from the economic point of view. For example, there is an opportunity cost of using the buildings for tuberculosis control rather than for a factory or an office building, or using resources to make an X-ray machine rather than a machine that could be used to produce computers. Therefore, a value of capital items needs to be established even after they have a book value of zero. In many developing countries, the ministry of health receives many inputs free of charge or at reduced price, such as donated drugs, radio or television time for health education and communication, or volunteer labour. Some of these resources still have an opportunity cost in terms of foregone non-health welfare because they have alternative uses—the television time, for example, could be used to advertise consumer goods. A value needs to be established that can be given to donated or reduced-cost items even when the ministry of health does not pay for them.

Where the donated goods are specific to the health intervention—for example, pharmaceuticals or volunteer labour dedicated to a particular disease or person who is ill—the question becomes whether the intervention could always be provided using donated goods. If this is the case, the opportunity cost in terms of foregone non-health welfare in the country under consideration is zero. If not, the inputs should be valued at the cost that would be incurred if they needed to be obtained in the market place.

• In many resource markets, there may be distortions that cause the current market prices to diverge from opportunity costs. On the one hand, the observed price can be higher than the opportunity cost, due to monopoly power or taxes/import tariffs, while on the other hand the going price can be lower than the opportunity cost, due to subsidies or "dumping" of products in the market at below cost. Again, a value needs to be established that more closely reflects the opportunity cost than the observed prices or charges.

Ingredients approach

In CEA it is important to be able to distinguish clearly quantities from the prices used to value them rather than to report only total costs or total expenditures, for two reasons. Firstly, total expenditures might well be estimated from a financial perspective and might not include all resources nor value them appropriately. It is important for analysts and policy-makers to be able to judge the appropriateness of the way costs have been estimated. Secondly, an ingredients approach allows analysts from one country to more easily assess if costs collected in another country can be used or

modified to their settings. They can, for example, decide the extent to which they use, or would use, different types of inputs, in different quantities, or at different prices. This cannot be done if total expenditures are used as the basis of the analysis, or if only total costs or total expenditures are presented and reported. For this reason we recommend using the ingredients approach to estimating costs. A software known as CostIt *(65)* is available to assist in this process. It is a costing template that uses the ingredients approach to report and analyse cost information in a standardized way (available from www.who.int/evidence/cea).

3.2.2 TRANSFER PAYMENTS

Some interventions may result in financial flows within society from the government to individual patients. Examples are unemployment or sickness benefits. Such transfer payments are a financial cost to the paying government (or to taxpayers in general), a financial gain to the patient, but do not use or create resources.[15] These money streams signify a change in command over resources, not a change in the aggregate value of resources available to society. Transfer payments are, therefore, generally excluded from a CEA.

However, as argued above, social welfare is influenced by who makes and receives these payments. In theory, these welfare changes should be taken into account but this type of analysis has not typically been done so we do not recommend their inclusion. On the other hand, any related administrative costs do use real resources and should be included.

3.2.3 THE UNIT OF ACCOUNT

In nearly all economies, domestic market price levels are higher than world market price levels, which may be caused by exchange controls, import quotas, and other trade restrictions. This creates a need to bring all resource inputs to a common basis so that they can be aggregated into an estimate of the costs of a health intervention. To do this it is necessary to define a unit of account, that is, to choose a *numeraire* or price level—domestic or world market price level—and to choose a currency—national or foreign currency—in which to express all resource inputs. The unit of account affects the valuation of traded and non-traded goods, and needs to be chosen first.

The numeraire

Because the analysis is typically undertaken from the perspective of an individual country, it is often argued that the world price level is the most appropriate starting point for analysis *(64)*. The Organisation for Economic Co-operation and Development guidelines for project

appraisal provide the rationale for using world prices as "they represent the actual terms on which a country can trade" *(64)*. The cost to a country of importing goods is the foreign exchange given up to purchase them, so internationally traded goods are valued at their traded or "border" prices, termed here "international prices". Non-traded goods are, however, subject to local market distortions and the use of observed market prices might not reflect true opportunity costs. For this reason it would not be appropriate to add traded goods valued in international prices to non-traded goods valued at local market prices *(64)*.

The solution generally used in cost-benefit analysis in other sectors is to revalue non-traded goods in terms of international prices, taking into account distortions that exist in the domestic goods markets. To convert domestic prices to international prices requires application of a "conversion factor", which is the proportion by which domestic prices exceed international prices because of market distortions. To do this for many goods and services is a considerable amount of work, and instead the standard conversion factor can be used, which is defined as the weighted average of all the conversion factors in the economy (for all goods). The standard conversion factor (SCF) can be estimated as the ratio of the value of traded goods and services at the international price level to the value of traded goods and services at the domestic price level, or can be approximated by the weighted average import tariff *(37;66)*. For example, if the average tariff were 25%, the SCF would be 0.8.

Adjustments for market distortions are rarely enforced in CEA—particularly the use of conversion factors to ensure that traded and non-traded goods are expressed in terms of the same numeraire. In addition, shadow pricing of non-traded goods is unusual, except perhaps for the exclusion of some transfer payments, possibly because many studies have originated in the USA where market distortions are arguably lower than in other settings. Several years ago, in an evaluation of immunization programmes in developing countries, Creese *(67)* showed that the use of appropriate shadow pricing can make a difference to the conclusions about whether or not an intervention is cost-effective, but little further work has been done since then. WHO-CHOICE is currently exploring this question to determine if it is possible to identify the circumstances in which it is critical to use these methods. For the moment, CHOICE uses the traditional approach. The way traded and non-traded goods have been treated is described below.

3.2.4 TRADED GOODS

Traded goods (e.g. equipment, supplies and pharmaceuticals) are the commodities that are, or could be, available on the international market, and could be available to all countries at an international market price. Most countries are too small to effect the international price—either for goods they import or export. So the opportunity cost for imported goods can be considered the foreign exchange that leaves the country in order

to pay for the inputs—e.g. they should be valued at the international price. Similarly, the value of an input to an intervention that is produced locally but could be exported is the value that could have been obtained for it on the international market. Traded goods are, therefore, valued at the international market price, adjusted to include cost, insurance and freight (CIF) for imported goods and free on board (FOB) for exported goods.

The CIF price should exclude import duties and subsidies (transfer payments), and include the selling price of the producing country, freight, insurance and unloading charges. If the goods are imported, the costs of local transport and distribution (termed "domestic margin") should be added to the landed price to approximate the local opportunity cost because local transport and distribution does use resources which cannot then be used elsewhere *(63)*.

The FOB price of exports should include the production cost as well as the costs to get the product to the harbour of the exporting country, and includes local marketing and transport costs and local port charges *(63)*.

The prices of some internationally traded goods might still include market imperfections to an extent (e.g. patented drugs are protected by definition). If an intervention in a particular country uses a patent drug, the question is whether a generic substitute exists and has similar effectiveness. If a generic substitute exists and has the same effectiveness, then its price should be used. The logic is that decision-makers need to know the cost-effectiveness of an intervention if the appropriate inputs were used. In some settings it might also be useful to show how the cost-effectiveness would alter with the use of the brand name substitute. If no generic product exists, or is unlikely to in the lifetime of the project, or the health programme does not have access to it, then the price of the patented product should be used. If a generic is predicted to become available later in the life of a project, then the expected generic price should be used after this time.

3.2.5 NON-TRADED GOODS

In general, personnel, utilities, buildings and domestic transport are treated as non-traded goods. Some considerations about the valuation of these items are highlighted below.

Labour costs

Labour market prices might not reflect true opportunity costs. To determine the economic value of labour employed in health interventions, these prices must be adjusted for distortions in the labour market, and so-called shadow wage rates (SWR) then could be estimated. Labour has traditionally been broken into two basic categories: scarce labour and labour which is not scarce locally. A third and fourth category, voluntary labour and patients' and care-givers' time, are also discussed. The distinction between "scarce" and "non-scarce" labour will vary by setting. In some countries, it is not uncommon for doctors and nurses to

be unemployed—e.g. it could not be argued that their skills are scarce. In other countries there are consistent shortages of medical personnel and government controls salaries. Similarly, in countries where agriculture is an important activity, low-skilled labour can be in short supply during certain periods of the year, such as harvest time, and readily available at other times. Therefore, the analyst should make their own judgements, and justify their choices.

Scarce labour. Scarce labour is typically labour that involves skilled workers for which there is little or zero unemployment. For this type of labour, it is recommended to take prevailing market wages and fringes plus the monetary value of housing and other allowances to give an approximation of the opportunity cost. This may well underestimate the true opportunity cost of skilled health workers in countries where the private sector does not function and governments control salaries.

The opportunity cost of labour is the gross salary plus fringe benefits. This represents the total resources that society pays to employ someone. Fringe benefits include the employer's contributions to social security, other pension plans, health and life insurance, and perks such as use of a car, free use of accommodation or financial contributions to private accommodation.

An important question is what to do about the valuation of expatriate labour employed in a country on salaries that are much higher than those paid to people with similar skills locally. The general answer to this question for GCEA is that it depends on whether the intervention needs this type of labour or whether the expatriate labour could be replaced with local labour with the same qualifications, skills and efficiency. If for some reason the intervention absolutely needs the expatriate labour, it should be considered as a traded good and valued accordingly. However, if the intervention would be normally undertaken with local labour, and the goal is to evaluate whether an intervention undertaken efficiently is worth doing, then local labour costs should be used.

Non-scarce labour. In many countries unskilled labour is not scarce— there are many more people who apply for positions in the modern sector than posts available. The cost to the economy of using unskilled labour in a health intervention is the value of the net output lost elsewhere.

Where labour is drawn from rural areas and would have been employed in agricultural production, the opportunity cost is often taken to equal the value of lost production. An indirect way of estimating this is to use the rural wage rate, adjusting for seasonal fluctuations in demand. At some times of the year this might be close to zero. Where labour is drawn from the informal sector in urban areas, the economic price of labour in the urban areas can be approximated by estimates of annual incomes in the urban informal sector. The urban formal sector wage rate is likely to be an overestimate, especially where minimum wage laws apply.

Voluntary labour. Voluntary labour is, by definition, free to an intervention. It should be treated similarly to expatriate labour. If it can

be assumed that the intervention will always be able to call on this volunteer labour, it would be valued at zero cost. If not, the cost of employing others to undertake this task should be used—effectively this means that it would normally be valued at the wage rate of health personnel who would normally be employed to do the same tasks.

Buildings

One of the differences between the accounting and the economic methods of cost valuation is the treatment of capital. For example, capital items such as buildings that have been written off in the accounts and no longer incur a depreciation cost would still have a cost from the economic point of view. There are two possibilities for the valuation of a building or space used by the intervention. The first is to use the annualized value of the building *(17)*. This is done using the replacement costs of the building, i.e. the cost of constructing a similar building today, and the annualization factor that incorporates the useful life of the building (depreciation) and the opportunity costs (interest rate) of the funds tied up in this asset. Details about how to calculate annualization factors are discussed in detail in Section 3.3.2 below. The second option is to use the rental value of a similar space in the same location, which could provide the same function, e.g. a private clinic or hospital. The rental value incorporates both the depreciation and the opportunity costs of the asset. However, this method is only appropriate if competitive rental markets exist which is certainly not the case in the rural areas of many of the poorest countries of the world. As a result, WHO recommends and uses the former method.

WHO-CHOICE estimates of unit costs

One of the most important inputs to a costing process is estimating unit costs of patient services, e.g. cost per bed day or cost per outpatient visit. Methods to estimate these unit costs are covered in most guidelines, e.g. in Drummond et al. *(17)*, but there are many countries for which reliable estimates are not available. Although the long-term solution is to encourage appropriate costing studies in all settings, for short-term use, WHO-CHOICE has developed models to predict country-specific unit costs, which is used to estimate the cost-effectiveness of health interventions by region *(68;69)*. See www.who.int/evidence/cea for regional estimates of unit costs per bed day, outpatient visit, category of personnel, litre of fuel, etc. for the 14 WHO subregions. The results of these models are not only useful for WHO-CHOICE but to analysts and policy makers who can estimate unit costs for their own setting.

3.2.6 Transferability of costs across time

Sometimes it will not be possible to obtain unit prices from the year of the study and it will be necessary to extrapolate from earlier years. Because general price levels change over time, it is necessary to adjust costs from other time periods to the base year used in the analysis (e.g. 2000 US dollars)—that is, they need to be valued in *constant* or *real* terms (net of inflation). Money values that are unadjusted for inflation are referred to as *nominal* or current prices.

There are several possible measures of inflation:

1. *The consumer price index* (CPI) reflects the change in the cost to the average consumer of acquiring a fixed basket of goods and services.[16] However, it is questionable if its determinants (i.e. choice of goods and services to include, and the weights used) are reflective of health costs as a whole. Moreover, the CPI is only appropriate if the price of the resource in question is changing at the rate of the general price inflation *(70)*.
2. *The Gross Domestic Product* (GDP) implicit price deflator is defined as the price index that measures the change in the price level of GDP relative to real output. It measures the average annual rate of price change in the economy as a whole. It also takes into account changes in government consumption, capital formation, international trade and final household expenditure, and therefore covers virtually the whole economy. It is the broadest-based measure of inflation *(71)*, and our recommended deflator for making health sector cost adjustments over time.[17]
3. *The rate of wage inflation* reflects the average annual increase in wages throughout the economy, or in specific sectors of the economy (e.g. public service). It is too narrow to used as a general index of inflation.
4. *The rate of inflation of specific product groups* reflects the rate of inflation for individual or groups of products, such as agricultural products, raw commodities and food. Some countries have an index of inflation for health goods and services, but not enough to recommend its use broadly.

For consistency and clarity reasons, a single indicator of the inflation rate should ideally be recommended in costing guidelines. The most appropriate inflationary measure for adjusting costs for CEA is the one which reflects most closely the general price level of the resources used to produce health interventions *(71)*. This would probably be the health component of the GDP deflator, but this is available in only a few countries. Therefore, we recommend using the GDP deflator. If no GDP deflator is available for a country, the CPI can be used as the second best alternative.

3.2.7 TRANSFERABILITY OF COSTS ACROSS SETTINGS

Most of the above discussion about costs is appropriate for collecting data about costs for a CEA undertaken in one setting. There may be several reasons to transfer cost estimates across settings:

1. Studies done in one country would show the quantities and prices of inputs in their setting. Analysts in other settings could adjust quantities and prices as needed. No special methods for transferring prices are required.
2. Sometimes analysts may not have the price/cost information for their own setting. In that case, the distinction between traded and non-traded goods should be made, and traded goods should be valued in their international price as described above and purchasing power parities (PPPs) could be used to convert prices of non-traded goods.
3. Some unit costs might not be available in a local setting and it would be very complicated to break them into their traded and non-traded components. An example is the unit cost per inpatient bed day. In these cases, PPPs could be used to translate unit costs from another setting and would give at least an approximation of local costs. This is not an ideal solution because it is often very difficult to tell from the published literature if the reported unit costs of inpatient days include all relevant costs and have been valued appropriately.

Presentation of results in international dollars

Since global cost-effectiveness estimates are not readily adaptable to individual countries, and since very few countries are able to estimate the costs and effects of all possible interventions in their own settings, WHO-CHOICE is making available the costs and effects of a wide range of interventions for 14 epidemiological sub-regions of the world.

To capture differences in purchasing power, results are presented in international dollars. Costs in the local country currency units are converted to international dollars using purchasing power parity (PPP) exchange rates. A PPP exchange rate is the number of units of a country's currency required to buy the same quantities of goods and services as one unit of currency in a reference country, in this case the United States. An international dollar is, therefore, a hypothetical currency that is used as a means of translating and comparing costs from one country to the other. See Annex D for more detail about how to interpret results presented in International dollars. The PPP exchange rates used in this analysis were developed by WHO and are available on the WHO-CHOICE web site www.who.int/evidence/cea.

3.3 COST ANALYSIS

3.3.1 CHOICE OF DISCOUNT RATE

As discussed in Section 2.4, WHO-CHOICE recommends estimating the costs of projects over a 10-year period. The costs incurred in each of these years cannot simply be summed without any adjustment. Individuals and society prefer to pay costs in the future rather than now, so from today's perspective, a cost of $100 payable after 10 years is not seen to be as high as a cost of $100 payable today. The present value of $100 payable in 10 years is, therefore, less than $100. Discounting is the process of converting future costs to their present value, to reflect the fact that, in general, individuals and society have a positive rate of time preference for consumption now over consumption in the future. For comparability across studies, it is important that analysis is performed using a common discount rate. For that purpose, WHO-CHOICE uses a discount rate of 3% for the base case, as suggested in a number of guidelines. A discount rate of 6% is also explored using sensitivity analysis. If country analysts wish to use country-specific rate of return of long-term government bonds as the social discount rate for costs, they may do this using sensitivity analysis. The question of discounting is discussed in detail in Section 6 below.

3.3.2 ANNUALIZATION OF CAPITAL INVESTMENTS

Capital goods are defined as inputs that last for more than a year, e.g. vehicles and buildings. As they are bought in one year and used for several years more, there is a need to spread the costs over the study period. The economic cost of using capital consists of two components: the opportunity cost of making the investment (resources invested in its purchase that cannot be used elsewhere) and the rate at which the capital is "used up" (commonly called depreciation). Where rental markets exist and can be considered to work relatively well, the rental price of the capital item can be assumed to incorporate both components. However, in many situations rental markets do not function or do not function well, in which case it is best to annualize the cost of the item using the following formula *(17)*.

Let P be the value of the good when bought where resale is netted out and resale nominal value be denoted by S. If K is the nominal purchase value, then

$$P \mid K\, 4\, \frac{S}{(1\, 2\, r)^n}$$

where r is the period interest rate and n is the period in which the capital is replaced. Assigning E as the equivalent cost per period, we get

$$E \mid \frac{P}{A(n,r)} \quad \text{where} \quad A(n,r) \mid \frac{1\,4\,(1\,2\,r)^{4n}}{r}.$$

The method described above is defined for new equipment. For old equipment, instead of *P*, use the replacement cost of the equipment or the original cost indexed to current dollars and a full life. To ensure comparability between studies undertaken in the same setting analysts should seek to ensure consistency in their assumptions about the useful life (longevity) of capital. Locally appropriate life spans should be standardized in different settings. In its regional tables, WHO uses averages appropriate to each region and will publish these life spans as they are developed further over time.

3.3.3 START-UP COSTS

The start-up period is defined as the period between deciding to implement an intervention and starting to deliver it to the first beneficiary. The start-up period can vary from several months to several years. Resources used in the different start-up activities include personnel, supplies, overhead and capital items. Start-up costs, whether recurrent or capital, should be annualized over the lifetime of the start-up activities. This is done in two steps: first the annualized value of capital items and the total costs of recurrent items are summed over the whole start-up period; second this value is annualized over the lifetime of the programme. The choice of the period over which start-up costs should be annualized is arbitrary and should be done on a case by case basis. In interventions where health technologies are not expected to change over a short time period, we recommend the use of a 10-year period as the basis of annualization of start-up costs. This is again arbitrary but will help improve the comparability of the results.

3.3.4 CAPACITY UTILIZATION

It is not uncommon, especially in developing countries, for capital (e.g. hospitals, health centres, laboratory equipment) and labour to be used at less than full capacity. GCEA requires estimates of the total costs of providing an intervention against the counterfactual of the intervention not existing. Overhead costs that are required to provide the intervention are included, some of which can be shared by other programmes. The total costs estimated against this counterfactual can vary substantially according to the level of capacity utilization. For example, the costs of treating TB patients in hospitals might appear to be high if a study were undertaken in a hospital where occupancy rates were low and the capital costs of the building are allocated across few patient stays. If these results are compared with the costs of providing care for children under five years of age derived from health centres that are always crowded,

differences in costs would reflect differences in capacity utilization rather than the costs of each intervention run relatively efficiently.

Given our interest in making recommendations on what types of interventions would be appropriate if policy-makers could reallocate resources, it is important that regional cost-effectiveness league tables control for capacity utilization and report the cost-effectiveness of interventions that are done efficiently. It is not useful for policy-makers to know the cost-effectiveness of interventions undertaken in a technically inefficient manner. This means making sure that the target efficiency is attainable in the region. WHO is currently using 80% capacity utilization as the norm for its sectoral analysis, consistent with recommendations made in previous CEA guidelines *(17;18)*.

Analysts wishing to adjust the WHO-CHOICE estimates to their countries can modify this assumption as appropriate. An important implication is that analysts undertaking individual cost-effectiveness studies should report the capacity utilization that forms the basis of the estimates. Otherwise it will not be clear to policy-makers or analysts in other settings if the results simply reflect excess capacity or the fact that an intervention is not cost-effective compared to others, even if it were done in an efficient manner.

Cost of scaling up interventions

An important question that is facing many governments is the cost of scaling up interventions to achieve target coverage levels. As coverage expands into remote areas, the marginal costs of providing an intervention to each additional person usually increase. The cost of scaling up interventions, including economies and diseconomies of scale, should be taken into account. For this reason, WHO-CHOICE presents cost-effectiveness estimates of different interventions e.g. at coverage levels of 50%, 80% and 95%. This involved the development of price multipliers to provide a conversion factor for prices at different levels of coverage *(61)*, and unit costs of outpatient visits to health facilities at different coverage levels. More detail of the methods used and results of this analysis are available from the WHO-CHOICE web site www.who.int/evidence/cea.

Summary of Recommendations

1. Ideally, analysts should follow the ingredients approach and collect and report information on the quantities and prices of the resources used in addition to total expenditures.
2. The cost of providing health interventions should be included in the analysis as should the resources used up in seeking or obtaining an intervention (e.g. transport costs). It is recommended that

productivity gains and losses due to an intervention, including time costs of seeking or obtaining care, should be excluded from the CEA. Where they are believed to be particularly important, they should be measured (rigorously) in physical units (e.g. time gains or losses) and reported separately.

3. Transfer payments should not be included in CEA. However, any related administrative costs should be included.

4. Costs of central administration and the education of health professionals can be regarded as existing or ongoing costs and should not be included in the analysis. This does not include training costs for a specific intervention, which should be included.

5. Shadow pricing should be used to determine the economic costs of goods that have no market price or if market prices are believed to have major distortions.

6. Prices of traded and non-traded goods should, in theory, be expressed in terms of a common numeraire, and we recommend using the world (international) price level to allow for comparability of results.

7. The annual costs of capital investments can be approximated by their rental price where a rental market exists and works relatively well. But because this is often not the case, the preferred approach is to annualized them taking into account purchase value, resale value, interest rate and working life.

8. Costs should be discounted at an annual rate of 3% in the base analysis. The sensitivity of the results to using a 6% rate should also be explored (see Section 6).

9. Analysts should report the capacity utilization that drives their cost-effectiveness estimates. WHO-CHOICE consistently uses 80% capacity utilization to obtain estimates of the cost-effectiveness of interventions if they are undertaken relatively efficiently.

10. Prices should be adjusted to a common year using the GDP deflator where possible. If this is not available, the Consumer Price Index can be used.

4   Estimating health effects

This section examines issues related to the estimation of the denominator of the cost-effectiveness ratio, which is the difference in population health with an intervention compared to the counterfactual to which it is being compared (the net health effect). This section begins by explaining how time lived in non-fatal health states can be combined with survival data in order to aggregate the impact of an intervention on fatal and non-fatal health outcomes into a single outcome measure. Then the methodology to estimate the counterfactual for analysis (the null) is examined, one of the most important challenges in conducting GCEA. This is followed by a discussion of the estimation of intervention effectiveness, defined in relation to the counterfactual.

The primary objective of this section is to review the process of estimating effectiveness specifically for GCEA, not effectiveness in CEA in general. Since the two share many of the same techniques—for example, the use of models to combine data from various sources—these are not discussed in the current set of guidelines. In those instances, the reader is referred to the existing literature *(17;18;72;73)*.

4.1 Defining the outcome of interest

The first use of GCEA is to compare a set of interrelated interventions, usually interrelated because they have the same goal—e.g. to reduce the risk of cardiovascular disease, or to improve the health of children under five years of age. The second use is to provide information useful to the decision about how to allocate scarce resources from the perspective of the sector as a whole, which involves comparing the costs and outcomes of all the different types of health interventions that would be possible. This requires cost-effectiveness to be estimated using an outcome indicator which measures changes in health taking into account fatal and non-fatal outcomes. Disability-adjusted life years (DALYs), healthy year

equivalents (HYEs), or quality-adjusted life years (QALYs) are all time-based measures of health that include the impact of interventions on years of life lost due to premature mortality and years of life lived with a non-fatal health outcome, weighted by the severity of that outcome. WHO-CHOICE employs DALYs in its CEA, and we recommend that other analysts also use DALYs in any GCEA for purposes of comparability. The metric has become increasingly used in the cost-effectiveness literature *(74–80)*.

4.1.1 Disability-adjusted life years (DALYs)

DALYs were first used in The Global Burden of Disease and Injury (GBD) study, a joint study between the World Bank, the World Health Organization (WHO) and Harvard School of Public Health *(81;82)*. This study began in 1988 with the objective to quantify the burden of disease and injury of human populations and define the main health challenges at the global level using a measure that could also be used for cost-effectiveness analysis. Using DALYs, the GBD was measured for 1990 and projections were developed to 2020. Estimates of disease burden were combined with estimates of cost-effectiveness using DALYs in *The World Development Report of 1993 (74)* in order to define priorities for investments in health. This summary measure of population health has since been refined and is used routinely by WHO for reporting on the health of populations and as an outcome measure for CEA *(37;83)*.

The DALY extends the concept of potential years of life lost due to premature death (PYLL) to include equivalent years of "healthy" life lost by virtue of being in states other than good health. DALYs lost due to a disease or health condition are calculated as the sum of the years of life lost due to premature mortality (YLL) in the population and the equivalent "healthy" years lost due to non-fatal health conditions (YLD):

$$DALY = YLL + YLD$$

The loss of healthy life due to non-fatal health conditions (YLD) requires estimation of the incidence of the health condition (disease or injury) in a specified time period. For each new case, the number of years of healthy life lost is obtained by multiplying the average duration of the condition (to remission or death) by a severity weight that measures the valuation of the loss of healthy life. Time lost due to premature mortality (YLL) is a function of the death rate and the duration of life lost due to a death at each age. The DALY is described in detail in Murray & Lopez *(37)* while additional information on summary measures of population health are described in Mathers et al. *(84)*.

4.1.2 NON-FATAL HEALTH OUTCOMES

In order to use time as a common currency for non-fatal health states and for years of life lost due to mortality, it is necessary to define, measure and numerically value time lived in non-fatal health states. The "valuation" of time lived in non-fatal health states formalizes and quantifies social perceptions of how time lived in a particular state compares with full health. Depending on how these valuations are derived and in which summary measure they are used, they have variously been called disability weights, health state valuations, health state preferences or health state utilities. When measuring the burden of disease, or the decrement in health due to a disease or condition, each year of life lost is given a weight of 1. Years lived, but in states less than full health, are given a weight between 0 and 1, with 0 representing full health.

While death is not difficult to define, non-fatal health states are. Non-fatal outcomes of disease differ in their impact on the individual, and the impact on the individual is mediated by contextual factors including personal characteristics and the physical and social environment. Non-fatal outcomes involve multiple domains such as mobility, anxiety and pain: health state valuations provide the means to weight and then aggregate individual functioning on these domains of health. Methods for eliciting health state valuations are described in Section 4.1.3.

The health state valuations used to estimate the burden of disease in terms of DALYs lost do not represent the lived experience of any disability or health state, or imply any societal value of the person in a disability or health state. Rather they quantify societal preferences for health states in relation to the societal "ideal" of good health.

Thus a weight for paraplegia of 0.57 does not mean that a person in this health state is almost "half dead", that they experience their life as halfway between life and death, or that society values them as a person less than anyone else. It means that, on average, society judges a year lived with paraplegia (weight 0.57) as less preferable than a year lived in full health, and a year with paraplegia is less preferable to a year with blindness (weight 0.43). It also means that, on average, society would prefer a person to have a year in good health followed by death than a year with paraplegia followed by death. In addition, society would prefer a person to live three years with paraplegia followed by death than have one year of good health followed by death (3 years x (1-0.57) = 1.3 "healthy" life years is greater than 1 year of good health).

Other things being equal, society would prefer to prevent or cure a case of paraplegia (weight 0.57) rather than a case of low back pain (weight 0.06), if each could be restored to full function for the same cost and there were insufficient resources to do both.

For CEA, the denominator is the gain in health due to an intervention rather than the loss measured in burden of disease (BOD) calculations.

Even though the same metric is used, the DALY is a negative concept in BOD calculations—DALYs lost—and a positive concept in CEA—DALYs averted. For that reason, the valuations used in CEA calculations are the complements (e.g. 1—health decrement) of the weights used in BOD analysis. For CEA, full health is valued at 1, with 0 representing the worse possible state of health, in this case death.

4.1.3 HEALTH STATE VALUATION

For countries undertaking national GCEA studies, it is recommended that a baseline set of estimates should apply the valuations used in the GBD study, which are being updated as described above for a new set of estimates for the GBD 2000.[18] For countries wishing to assess local health state valuations as well, we recommend using the data collection strategy following the standardized protocols developed at WHO *(85)*.

The two-tiered data collection strategy consists of:

• collection of health state valuations in the general population using a multiple-state ordinal ranking and visual analog scale (VAS) questionnaire; and
• collection of health state valuations among a smaller sample of respondents with high levels of educational attainment using a multi-state, multi-method survey.

The primary objective in the first component of the strategy is to estimate aggregation functions that can be used to generate overall valuations based on the levels on multiple core domains of health. Data collection in the general population is also designed to address questions of variation in health state valuations across different types of respondents and settings. The results are used to analyze the possible determinants of variation both across and within countries. The second part of the strategy is required in order to translate responses obtained using the simple methods applied in the population-based surveys (rankings and VAS) to an interval-scaled measure of strength of preference that may be used in constructing summary measures of population health *(86)*.

4.1.4 DALYs AND THE THEORETICAL FOUNDATIONS OF CEA

In Section 2.1 it was argued that the numerator in CEA should be the loss of welfare because of the reduction in non-health consumption that results from allocating scarce resources to a health intervention. The denominator should be the gain in welfare associated with the increase in health. DALYs are a summary measure of population health that capture the impact of an intervention on both fatal and non-fatal health outcomes. The bridge between mortality and non-fatal outcomes is the set of health state valuations that represent the overall health levels associated with time spent in different states.

It is important to note explicitly what these health state valuations do and do not capture. They do not measure health utility as defined in terms of the Von Neumann-Morgenstern axioms of expected utility. This is partly because risk attitude is a separate consideration which, we argue, is not relevant for quantifying health levels *(87)*, and partly because health-related utility encompasses a broader range of dimensions of well-being that extend beyond the focus on health that is sought in the health state valuations. Health, as defined in DALYs, interacts with these other domains but can be conceptualized as being distinct from them *(88)*.

Health levels can be understood without reference to the contribution of health to overall well-being, and without relying on the language of individual preferences. Even if DALYs, however, are intended to reflect a health construct that is more narrowly defined than well-being, DALY maximization—the implicit objective of GCEA—can be consistent with the welfare maximization implied by the theoretical framework described earlier under certain, admittedly strict, but reasonable conditions.

The first is separability of individual utilities across time. Using a time-based measure of population health implies that years are additively separable—the total health impact of an intervention is the sum of its impact at different moments in time, and the valuation of a health state at one moment is independent of the health state in another. There has been considerable debate in the literature about this assumption, which is common to QALYs, and one of the reasons why the healthy year equivalent (HYE) was proposed as an alternative outcome measure for CEA *(89;90)*. At one level, the appeal of this requirement is highly pragmatic, allowing impacts over different periods of time to be combined in a relatively simple way.

The second condition is separability across people *(89)*. The welfare impact of a health event on a given individual does not depend on the health levels of other members of a population. This assumption allows DALYs (and other summary measures of population health including QALYs) to be constructed from overall measures of individual health that may be assessed without reference to the states of health of other people in society. While inequality is clearly an important consideration for health policy, it is useful to develop measures of the average level of population health separately from a measure of the distribution of health in order to allow explicit accounting for both. Thus, while there have been proposals for summary measures that reflect both levels and distribution of health (e.g. *(91)*), DALYs are intended to capture only the former. This implies that considerations of changes in the distribution of health across individuals brought about through a health intervention should be introduced into the policy debate through mechanisms other than the way health is aggregated in the denominator of a cost-effectiveness analysis.

The third condition under which DALY maximization is consistent with welfare maximization is that the health and non-health components

of welfare *(92;93)* are separable and welfare is a linear function of years of healthy life. While declining marginal utility of increased consumption is an assumption with clear face validity, a similar assumption that extra increments in health have less utility value as individuals have lived longer healthier lives does not have face validity. Debates on the allocation of scarce resources, such as organs, clearly indicate that most individuals do not accept the premise of declining marginal utility of health *(94;95)*. In fact, survival can be thought of as the capacity to generate utility from consumption. Extra years of healthy life could have a greater welfare value, not less, if people expect to have higher levels of consumption in those years. As an operational assumption, it seems reasonable to assume that welfare is a linear function of years of healthy life.

Although many examples can be found that would imply that these assumptions do not hold strictly true in all circumstances, they are the basis of all cost-effectiveness analyses. Not only are they widely accepted, but also alternatives to these assumptions have not proven feasible in practical application. It is useful to recognize that these assumptions have considerably more plausibility than the assumptions embedded in cost-benefit analysis without equity weights, in which a year of healthy life for a poor person generates much less welfare than a year of healthy life for a wealthy person.

4.1.5 Discounting

DALYs capture the impact of an intervention on the future stream of healthy years of life lost due to disease or injury. Future healthy years gained by an intervention are discounted at a 3% rate to their present values for the base case analysis in WHO-CHOICE, so that a year of healthy life gained in 10 years' time is worth 24% less than one gained now. The rationale for discounting future health benefits is discussed in more detail in Section 5.

4.1.6 Age-weighting

The Global Burden of Disease study weighted a year of healthy life lived at young and older ages lower than a year lived at other ages, something that has subsequently proven to be controversial. Some critics have argued that age weights were unacceptable on equity grounds (i.e. every year of life should be of equal value regardless of the age of the person who gains it) *(96)*. In response, Murray *(97)* argued that age weights are not in themselves inequitable, because everyone potentially lives through every age. Moreover, the preference given to years gained at particular ages is consistent with societal priorities expressed in a number of studies that have indicated there is a broad social preference to value a year lived by a young adult more highly than a year lived by a young child or at older ages *(98;99)*.

For the purpose of GCEA at the country level, individual analysts have the choice whether to use age weighting. WHO reports its cost-

effectiveness results with age-weighting in its base case and without age-weighting as part of the sensitivity analysis. A more detailed discussion of the principles and techniques of age-weighting can be found in Murray and Lopez *(37)*.

4.1.7 ESTIMATING YEARS OF LIFE LOST DUE TO MORTALITY

For measuring the global burden due to premature mortality associated with different types of diseases, the standard expected years of life lost (SEYLL) method has been adopted by WHO. This uses the expectation of life at each age *x* based on some ideal standard to estimate the loss of years of life associated with a death. See Annex E for more detail on the various possible approaches to the measurement of premature mortality in DALYs.

The needs for CEA are, however, different to those for measuring burden of disease and the method of estimating YLLs gained by an intervention needs to be more complex. To show this, assume that an intervention, e.g. insecticide-treated nets (ITN) in malaria, can reduce infant mortality by 50%. What would be the benefits from the implementation of ITN in terms of years of life gained? Imagine that this intervention is applied to the population represented in Table 4.1. It is commonly assumed that preventing a neonatal death (at age 0) gains years of life equal to the life expectancy at that age (74.68 years).

This is a good approximation as long as the changes caused by the intervention do not change age-specific and overall life expectancies substantially. A 50% reduction in infant mortality does, however, substantially change life expectancy at birth. In that case, the total number of life-years gained calculated according to the common assumption would be the number of lives saved (627) times the life expectancy (74.68 years) which equals 46 809 . This is not correct. Table 4.1 shows the life expectancies without and with the intervention. The reduction in childhood mortality increases life expectancy at birth by 0.47214 years. This means that the total number of life-years saved equals 0.47214 x 100 000 = 47 214, larger than the number estimated above.

Another way to think about the effect of an intervention on life expectancy and years of life saved is in terms of shifts in the survivorship curve. Survivorship curves are also a useful representation of mortality rates in a closed or stationary population. The y-axis shows the proportion of a birth cohort (exposed to a set of age-specific death rates) that survives to any age (shown on the x-axis).

Table 4.1   Life table without and with the intervention

Age (years)	Without intervention				With intervention			
	Pop. (l)	Death (d)	Person years lived (T)	Life exp. (e_x)	Pop. (l)	Death (d)	Person years lived (T)	Life exp. (e_x)
0	100 000	1 254	7 468 328	74.68	100 000	627	7 515 542	75.16
1 to 4	98 746	140	7 369 205	74.63	99 373	141	7 415 980	74.63*
5 to 9	98 607	46	6 974 499	70.73	99 232	46	7 018 769	70.73
10 to 14	98 561	82	6 481 581	65.76	99 186	82	6 522 722	65.76
15 to 19	98 479	223	5 988 983	60.82	99 104	224	6 026 997	60.82
20 to 24	98 256	296	5 497 147	55.95	98 879	298	5 532 039	55.95
25 to 29	97 959	422	5 006 610	51.11	98 581	424	5 038 388	51.11
30 to 34	97 537	448	4 517 868	46.32	98 157	451	4 546 544	46.32
35 to 39	97 089	522	4 031 301	41.52	97 706	525	4 056 889	41.52
40 to 44	96 568	774	3 547 159	36.73	97 181	779	3 569 674	36.73
45 to 49	95 794	1 248	3 066 256	32.01	96 402	1 255	3 085 719	32.01
50 to 54	94 546	1 809	2 590 407	27.40	95 146	1 820	2 606 850	27.40
55 to 59	92 737	3 423	2 122 198	22.88	93 326	3 445	2 135 669	22.88
60 to 64	89 314	5 154	1 667 069	18.67	89 881	5 187	1 677 651	18.67
65 to 69	84 160	8 655	1 233 383	14.66	84 694	8 710	1 241 212	14.66
70 to 74	75 505	12 806	834 220	11.05	75 985	12 887	839 515	11.05
75 to 79	62 700	16 213	488 708	7.79	63 098	16 316	491 810	7.79
80 to 84	46 486	18 957	215 743	4.64	46 781	19 078	217 113	4.64
85+	27 529	27 529	30 705	1.12	27 704	27 704	30 900	1.12

* Other than for age 0, e_x differs for intervention and non-intervention scenarios only at the 5th decimal place.

Figure 4.1 shows the survivorship curve for the life table in Table 4.1. As a result of the intervention reducing infant mortality by 50%, the curve shifts upwards. Years of life lived for the population, equal to the area under the survivorship curve, is increased through this intervention by 47 214 life-years. This can also be seen in the life table as the difference in the total number of life-years lived (T) with and without the intervention (7 515 542–7 468 328). The preceding paragraphs show that for interventions with a relatively large impact on age-specific mortality implemented over a number of years, period life expectancy (before the intervention) will not be a good approximation of the population health gain. In such cases, a population projection model is required in order to estimate fully the intervention benefits in terms of years of life lived *(37;100)*.

Figure 4.1   Survivorship curve
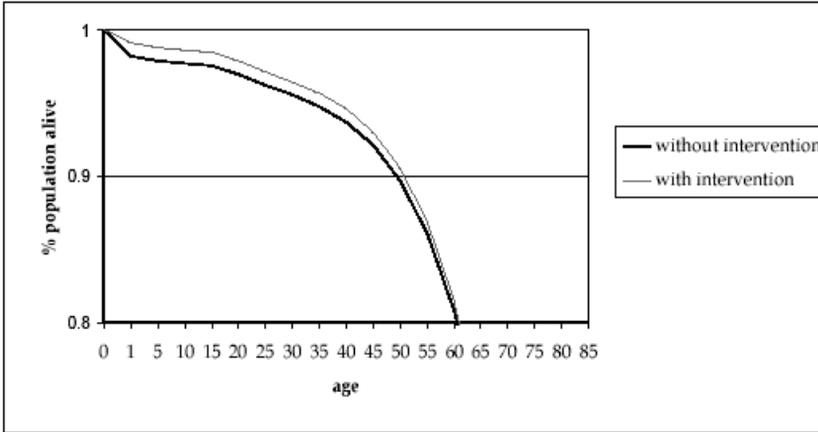
Population models calculate the benefits of an intervention as the difference in healthy years lived in the population with and without the intervention. They can capture the effects of the actual age-distribution of the population, the age-specific prevalence of disease, and competing risks, as well as the effects of discounting and age weighting if desired. WHO has developed a population model known as PopMod (see Section 4.2.3).

The need for a population model to estimate years of life gained by an intervention

If the intervention under study will lead to relatively small changes in mortality rates, it may not be necessary to use a population model to obtain a relative accurate estimate of the health gain. Preston and Gribble *(100)* have analysed the number of life years added to a population through different types of interventions and have shown that subtracting the age of death from local period life expectancy is a good first-order approximation of the years of life gained for some single-year interventions. Where age-specific mortality rates are changing over time, it would be more accurate to use cohort than period life expectancy. However, a population model should be used where interventions substantially effect life expectancy.

4.2 ESTIMATING POPULATION EFFECTIVENESS

The denominator in a CEA, the difference in population health between the intervention and the counterfactual (null set), must be expressed as a single number. In GCEA, the number of healthy years lived (HYL) for the

intervention scenario is compared to that under the null scenario using PopMod.[19] The difference is the number of DALYs gained by the intervention (see Annex F). A number of steps are required in these calculations. The first involves the definition of the cluster of interrelated interventions which captures the most important interactions between interventions in terms of costs and/or effects. In the second step, the partial null for that set of interventions is assessed by estimating what would happen to population health if all interventions in this cluster were eliminated. The third step is to estimate the impact of introducing all interventions, singly and combined on HYL at the population level. These steps are discussed in turn. Wherever the term "null" is used in this section, it refers to "partial nulls" as defined in Section 2, except where explicitly stated.

4.2.1 STEP ONE: DEFINING THE CLUSTER OF INTERRELATED INTERVENTIONS

As noted in Section 2, GCEA requires interventions that interrelate in terms of costs and effects to be analysed together. This is required in order to assess the joint effect of interventions undertaken at the same time. It was also shown that the analyst must consider what would happen, starting from today, if all interventions were eliminated. Only through the analysis of this counterfactual or null can the impact on population health of current interventions be assessed.

   In defining the null, it is not practical or necessary to assume that no interventions at all exist because the costs and health effects of many are unaffected by the existence of others. The null, therefore, was defined as what would occur if a specific group of interrelated interventions were eliminated today. Given this definition of the null as that state in which groups of interrelated interventions no longer exist, different so-called "partial nulls" can be defined within the GCEA framework. Section 2 provides more detail on the definition of sets of interrelated interventions and the partial null.

4.2.2 STEP TWO: DEFINING THE EPIDEMIOLOGICAL PROFILE OF THE NULL

 The starting point for defining the epidemiological profile of the null is usually the current epidemiological situation. The Global Burden of Disease study has estimated epidemiological profiles for a comprehensive set of diseases for many epidemiological subregions in the world. Many of the details are reported in the annex tables to the annual World Health Reports of WHO.

   How the null is to be estimated in practice depends to a large extent on the nature of the intervention cluster under study. If the current intervention mix comprises purely of preventive interventions, their elimination will affect only the incidence of disease. Where the current intervention mix includes only curative interventions, the null can be characterized by a change in the remission and/or case-fatality rates. The hypothetical elimination of rehabilitative and palliative interventions is

reflected in a change in the severity and, maybe, duration of the non-fatal health outcomes associated with the particular disease. In many intervention clusters, the hypothetical absence of interventions will be reflected by a change in all of these hazard rates simultaneously.

Hazard Rates

One of the most common concepts in epidemiology is the concept of a rate. Of particular interest are rates denominated in units of time, specifically in units of population time at risk. "Time at risk" for an event like the onset of disease, or incidence, is simply the aggregate time spent without the disease experienced by a given population group. As soon as a member of the population gets the disease, he or she stops contributing "time at risk". The incidence rate would be, therefore, the number of disease onsets divided by the total time at risk in the population. Incidence rates are conventionally denominated in units of years, i.e. the number of disease onsets in a year divided by the total time spent without the disease in the population during the year. However, incidence rates can vary within a year due to seasonal or other patterns. A good analogy is the velocity of a car. Although velocity can be constantly changing, in practice—as for incidence—we are often interested in the average velocity over a certain interval, i.e. between point A and B. At other times we might be more interested in the velocity at a point in time. The term instantaneous velocity can be used to distinguish this latter concept from average velocity. The same applies to rates: the rate at a point in time is termed the instantaneous rate. In mathematical models, it is often the instantaneous rate that is of primary interest, frequently called a "hazard rate". Demographers tend to refer to instantaneous rates as "hazard rates" or as "forces", e.g. the "force of mortality" is the mortality hazard. In epidemiology, the term "rate" is sometimes used to mean an instantaneous rate, or hazard, and sometimes an average rate. In PopMod, instantaneous rates are used, and they are referred to as hazard rates in the text.

It is important to note that interventions in WHO-CHOICE are evaluated assuming that they exist at full capacity for 10 years. Accordingly, the null must assume that current interventions cease to exist for 10 years. It should be remembered that even though an intervention might exist (or not exist for the null) for 10 years, the impact on population health can be felt over a much longer period of time and all these effects must be measured and included in the effectiveness estimate for the with- and without-intervention scenarios. PopMod allows the null, and the impact

of interventions, to be traced over the lifetime of the individuals currently alive, which has pragmatically been defined at 100 years (see Section 2). The null hazard rates should be estimated for this period as well.

In some instances, the null hazard rates may be easy to define. For example, in the case of cataract, the only possible interventions are surgical. Since there is no natural remission from cataract, the hypothetical absence of these interventions implies a zero remission rate. However, for most intervention clusters, deriving the null hazard rates is more complicated. A number of approaches can be used[20] reflecting the absence of interventions, including the use of observed patient data, or using the method of back-adjusting. These are described in turn.

Using observed patient data

Natural history models. Natural history models describe the progression of a disease in the absence of treatment. In theory this approach refers directly to the null, and has much appeal. Natural history models can be based on data from various sources. Firstly, data may refer to deprived populations whose access to health services is limited. In the absence of interventions, observed hazard rates could be interpreted as natural hazard rates. Secondly, the natural history of disease can be mimicked by considering study groups whose disease indicators reflect the failure of care in a particular area. For example, the natural (or null) incidence rate of diabetes-related complications in diabetes patients can be assumed to be similar to that of diabetic patients with very inadequate blood sugar control.

Trial data. Another option is to use data collected in trials such as randomized controlled trials (RCTs). RCTs randomly assign subjects to different types of interventions (sometimes using placebo or the "do nothing" option) to study the comparative effectiveness (or efficacy) of interventions. The randomization enhances the comparability of the different study groups and provides a valid basis for inferring that the intervention actually caused any observed difference in outcome between the groups. These studies can be valuable for estimating null hazard rates if one of the interventions is the "do-nothing" scenario. The analyst needs to ensure that "do-nothing" actually refers to the absence of any intervention, and does not imply "usual care". If the "do-nothing" option involved the administration of a placebo, analysts must in addition be aware of the fact that the placebo might effect the health outcome. Although a recent meta-analysis has shown that, in general, there is no evidence of a placebo effect, a positive moderate effect in placebo arms of studies with subjective outcomes like pain was detected *(101)*. It might, therefore, be necessary for analysts to make adjustments to extract the placebo effect from trial data where there are subjective outcomes.

Observational study data. In some cohort studies, a defined population is followed over time to observe the rate of occurrence of a particular outcome and to assess if it is related to the presence or absence

of treatment. Hazard rates (such as incidence and mortality) observed in people not receiving treatment can be used as null hazard rates as long as those people do not receive any intervention that is related to the cluster of interrelated interventions.

Back-adjusting

Conceptually, the null hazard rates can also be assessed by eliminating the impact of interventions, i.e. by considering their coverage and effectiveness. This approach is labelled "back-adjusting" since the analyst derives the null hazard rates from knowledge of current epidemiology and the characteristics of current interventions. For a single intervention,

$$\lambda_N = \lambda_C/(1-c.e)$$

where λ_N = null hazard rate (e.g. incidence of disease)
  λ_C = current hazard rate
  c = current coverage of intervention
  e = current effectiveness of the intervention.

Where interventions interact, the multiplicative form of the interaction is often appropriate:

$$\lambda_N = \lambda_C/(1-c_1.e_1)*(1-c_2.e_2)*\ldots\ldots*(1-c_n.e_n)$$

where the subscripts on c and e represent the number of the intervention.

The approach requires information on the (i) mix of interventions currently provided (ii) their coverage rates and (iii) their effectiveness.

General quality issues

When using the above approaches, analysts need to be careful when extrapolating identified null hazard rates to contexts other than those under study. For example, there may be important genetic or environmental differences in the natural history of a disease that may differ across populations. With infectious diseases, the context of where the disease occurs can change transmission dynamics. Another problem is that the observed patient data used to determine the natural course of the disease without treatment might come from studies undertaken many years ago. Where the relevant health outcome is also dependent on non-health variables that change over time, such as educational status, these data should be extrapolated to the present with care. Similar arguments apply to the use of "back-adjusting": analysts should ensure the generalizability of their estimates of effectiveness beyond the actual study context.

A critical issue in defining the null is the internal consistency of the various null hazard rates. Internal consistency is more likely when all relevant hazard rates stem from the same epidemiological data set (e.g. from a single well-designed longitudinal study), but may be problematic if the rates are obtained from different studies or if the single study has

measurement or design problems. For example, the rate of natural remission from one study may be higher than would be consistent with the prevalence of disease in another, and it is important that analysts check for consistency before using hazard rates taken from different sources. Software such as DisMod[21] has been developed to ensure that rates used in the analysis are internally consistent.

In most analyses, age- and sex-specific hazard rates need to be used. However, available studies sometimes only provide a single hazard rate that is not disaggregated by age or sex because of the small sample size involved. Again, DisMod can be of use, since it incorporates a number of algorithms for translating a single population hazard into age- and sex-specific hazard rates.

4.2.3 STEP 3: CONSTRUCTING A POPULATION MODEL

What is a population model?

A population model describes the health experience of a population conditional on a number of health states (such as healthy, dead, ill) and events (or transitions between health states), such as incidence, remission and mortality, reflected in hazard rates. The model describes the healthy years of life lived (HYL) by the population over the lifetime of all the individuals that are initially alive, and future births, as a function of the hazard rates and the distribution of the starting population across the various health states. It can be used to trace the HYL under the null and intervention scenarios.

Figure 4.2 shows the population model developed by WHO, named PopMod. This model currently has five boxes representing different health states: full health; death; disease X; co-morbid condition C; disease X plus co-morbid condition C. The arrows reflect the transitions between states; *r* indicates a generic transition, *f* indicates case fatality and *m* indicates mortality. T represents the total population and B represents births. The model allows for different hazard rates to be used for the different age- and sex-specific cohorts. It is possible to add or subtract health states if necessary.[22] The simplest form includes only three health states: those who are healthy (S); those who have the disease or condition of interest (X); and those who die (D) as a result of either disease X or background mortality. PopMod is not suitable for modelling transmission dynamics that can be important for diseases such as tuberculosis or HIV/AIDS. More information is provided in Lauer et al. *(102)* and in the users' guide included with the PopMod software.

Figure 4.2 Model structure with five boxes
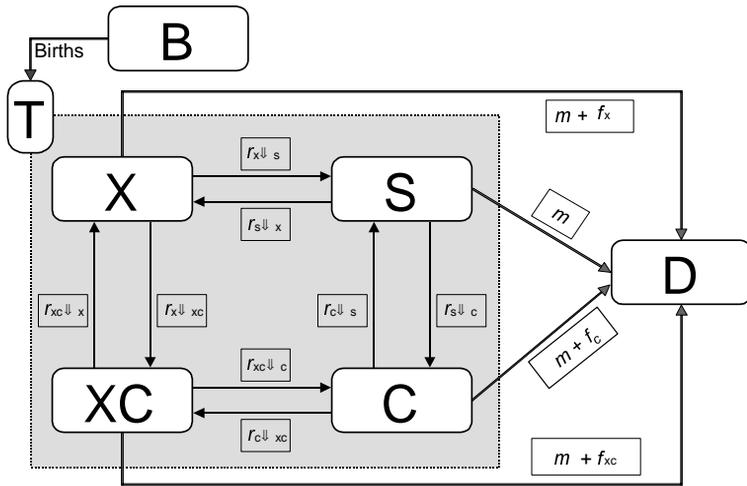
Determining the starting distribution in year 0

In any population model, the population of interest is first distributed into the defined disease states. This distribution will normally be based on the latest prevalence rates for the modelled diseases as reported, for example, in the burden of disease studies of WHO or other epidemiological studies. Such studies seldom report the prevalence of co-morbid conditions (XC) and the analyst might need to make assumptions about the distribution of co-morbidity[23].

To estimate HYL for the null and intervention scenarios, a value needs to be attached to time lived in the disease states X, S, XC and C—the valuation is zero for D. The population in box S cannot be assumed to be in perfect health because they will suffer from a variety of non-fatal outcomes not associated with X or C. Accordingly, the average (or background) health state valuation of the population without the diseases or risk factors of interest (e.g. background morbidity) is appropriate. For individuals in boxes X, C, or XC, the health state valuation of the condition of interest also includes the background disability of all the other conditions (not associated with X or C) that they may have. This allows the health benefit of moving from X back to S to be calculated correctly as the benefit associated with the removal of X alone. Section 4.1.2 provides more information on health state valuations.

Summary of necessary steps to measure population health impacts

The necessary steps to measure population effectiveness for a disease/health state or a cluster of diseases/health states are:

1. define the cluster of interrelated interventions.
2. define the epidemiological profile of the null

    a) define current epidemiological profile
    b) define null hazard rates.

3. construct a population model to calculate HYL

    a) define population of interest and allocate to the different states of health
    b) define intervention hazard rates
    c) define health state valuations for each of the boxes
    d) estimate HYL for the lifetime of individuals initially alive under the null and intervention scenarios.

4.3 ESTIMATING THE EFFECTIVENESS OF INTERVENTIONS

Many methodological issues in the estimation of effectiveness of interventions, such as the use of systematic review studies or the application of models, have been described in detail elsewhere *(18)*. Of special interest to GCEA is that the effectiveness of interventions should initially be compared to the null, i.e. the absence of those interventions. In that respect, the data sources for the estimation of effectiveness parameters are very much related to those suggested for use in defining the null.

Trial data must be treated carefully for three reasons. Often trials measure efficacy, or the effect of the intervention in ideal circumstances rather than the circumstances that would apply in practice. Secondly, the population groups included in trials often exclude the people who would benefit from the intervention in practice, such as the chronically ill, children and pregnant women. Thirdly, trials typically estimate the effectiveness (or efficacy) of interventions in comparison to current practice rather than in comparison to doing nothing.

However, useful information can be still obtained through adjusted indirect comparisons. If there are studies comparing interventions A to B directly, and there are also studies comparing A to doing nothing (or a placebo), it is possible to do a careful indirect comparison, with appropriate adjustments, to obtain some information on the effectiveness

of B compared to doing nothing. The validity of adjusted indirect comparisons will depend on the internal validity and the general similarity of the studies involved *(103)*.

4.3.1 OBTAINING DATA ON EFFECTIVENESS

Data Sources. The main sources of data ideally need to be obtained from a systematic review which, if done well, will ensure that there is a low probability of obtaining a biased estimate of the overall efficacy of an intervention. Through data pooling, a meta-analysis done as part of a systematic review can also increase the power to detect a difference in the efficacy of different interventions. Data pooling can also be used to explore questions of whether the effectiveness of interventions differ across different population groups (effect modification) by analysis of subgroups *(104)*.

Summary measures of efficacy. The summary measure of efficacy can be risk differences or relative risk for outcomes which are expressed as proportions (case fatality or incidence proportions) or effect size for outcomes which are expressed as continuous variables, e.g. measures of disability or functioning.

A relative risk is a measure of how much more likely an outcome is among individuals in a group given an intervention compared to a group not given an intervention. A relative risk of one means that the intervention is not efficacious with respect to the outcome being measured *(105)*.

An effect size is a standardized difference. It is the difference in means of the two groups being compared expressed in terms of standard deviation shifts or formulaically:
(mean of group 1 – mean of group 2)/standard deviation.

An effect size of 0 means there is no difference between the two groups whereas an effect size of 1 means that 88% of the control group would rank below the average person in the experimental group (assuming normal distributions) *(106)*.

SUMMARY OF RECOMMENDATIONS

1. It is recommended that analysts express population effectiveness in terms of DALYs, although measures such as QALYs and HYL could also be used.
2. For interventions that alter life expectancy, years of life saved by an intervention should be estimated from a population model.
3. The counterfactual scenario for estimating population effectiveness is the null set, defined as the lifetime health experience of a defined population in a situation where all related interventions directed against a disease or condition are stopped. The null set can be estimated using natural history models, using trial data or by back-adjusting using coverage rates and effectiveness of currently implemented interventions.

4. Data on the efficacy of interventions ideally comes from systematic reviews of studies. Efficacy can be expressed as relative risks for rates and effects sizes for means. Efficacy should be adjusted to reflect population effectiveness, taking into account factors like coverage, quality of care, adherence and other local factors, all of which can modify efficacy.

5. The health state valuations derived in the GBD study can be used until regional estimates are available from WHO. Results should be presented with and without age-weighting as part of the sensitivity analysis (see Section 6).

5   DISCOUNTING

Discounting is the process of converting future values—e.g. costs or health effects—to their present values to reflect the belief that, in general, society prefers to receive benefits sooner rather than later, and pay costs later rather than sooner. There is general agreement in the literature about the need to discount costs, and some agreement about the likely value or range of the appropriate discount rate. There is less agreement about the need to discount health effects and, if so, whether the appropriate discount rate should be identical to that used for costs *(18;107–109)*.

5.1 DISCOUNTING COSTS

The logic for discounting costs is that the value of a unit of consumption to individuals and society decreases over time, for three possible reasons. First, individuals take into account the fact that they might not be alive to benefit from future consumption, and society takes into account the possibility of catastrophe—the possibility that any or all interventions might at some point in the future become valueless due to the technology becoming obsolete, climate change or social chaos, for example *(110)*. Second, people and society might simply prefer consumption now to consumption in the future—called the pure rate of time preference or, sometimes, myopia. Third, if it is expected that incomes will increase, the marginal welfare gain from an additional unit of consumption will be lower in the future, when people are richer, meaning that any given increase in consumption is more valuable now than in the future. Accordingly it is standard practice to discount future costs to their present values to allow for differences in the value of one extra unit of consumption over time.

The mechanics are straightforward. The discrete time formula[24] for estimating the present value of any stream of costs is:

$$Cost_{presentvalue} \mid \sum_{t=0}^{T} \frac{Cost}{(1+r)^t}$$ Equation (5.1)

where r is the discount rate and t is the time period when the cost occurs. The appropriate rate of discount r is more controversial *(18)*. It has been argued that the societal discount rate should be derived from the market, in which individuals equate their willingness to trade off future consumption for present consumption with the interest rate. The interest rate reflects their ability to trade off future for present consumption by borrowing or lending. This is not convincing for a variety of reasons. For example, market imperfections such as taxes mean that the rate of return available on investment (the pre-tax interest rate) is higher than the rate of return obtained by any individual investor, so exceeds the rate of time preference expressed by each investor. Even if the average rate of time preference of individuals could be estimated from their market behaviour, some individuals with high rates of time preference may not have been able to borrow to the extent that they would have liked *(110)*. Their actual borrowing behaviour would be less than optimal from their perspective so their behaviour in the market place does not reflect their rates of time preference. In addition, Sen *(111)* has argued that individual discount rates do not take into account the interests of future generations whereas people making decisions on behalf of society should.

Following Sen *(111)*, we argue that it is not appropriate to aggregate individual discount rates in choosing the appropriate social discount rate for costs, even if people are asked about their discount rates for social, rather than private, decisions *(112)*. The appropriate social discount rate takes into account the possibility of catastrophe outlined above, pure time preference and the rate at which the value of a unit of consumption might decline over time. A number of guidelines have recommended 3% *(17;18)*, while a recent review by the UK Department of Health has argued that it is around 3.5% *(110)*. We suggest using 3% as the base case for costs and testing the sensitivity of the results to a rate of 6%.

5.2 DISCOUNTING HEALTH EFFECTS

It is standard practice in most cost-effectiveness studies to discount future health benefits at the same rate as costs, a rate between 3% and 5% per year *(18;113)*. Despite this, the practice is widely debated *(108;114;115)*. Some individuals certainly discount their future health, although not necessarily in a way that is consistent with the constant geometrical discounting assumed in Equation 5.1 *(116)*. There is, however, considerable variation across individuals, with some expressing a negative discount rate while some discount their own future health at very high rates *(117)*. It has also been shown that individuals might trade

off future health gains for social decision-making at a lower rate than they use for their own private decision-making, and recent work suggests that the average rate at which people discount health effects might be higher than that for costs *(112)*. However, the concept of aggregating only the preferences of people currently alive when their decisions affect the health of future generations does not seem appropriate or morally acceptable *(118)*, nor would we expect current generations to give adequate weight to the welfare of future generations along the lines argued by Sen *(111)*.

For these reasons, it is important to define a social rate of discount in the same way that the social rate of discount for costs was defined. A number of authors have shown that there are good theoretical reason why the social rate of discount for health effects should be lower than the rate for costs *(108;114;118)*. Gravelle and Smith *(108)*, for example, argue that the discount rate for health effects should be the discount rate for costs minus the rate at which the value of future health effects in terms of future income increases over time. This rate of change, they believe, will be positive in most settings, for example because the marginal utility of consumption falls with increases in income while the marginal utility of health does not. Acharya and Murray go further by arguing that there is no reason why society would choose to discount future health benefits at all *(118)*.

Earlier, Murray and Acharya *(119)* had concluded that the strongest argument for some positive rate of discount for health is the disease eradication/research paradox: assuming that investment in research or disease eradication has a non-zero chance of succeeding, then without discounting, all current expenditure should be shifted to such investment because the future stream of benefits is infinite. They noted, however, that the choice of a discount rate for health benefits, even if technically desirable, may result in morally unacceptable allocations between generations.

Two practical arguments have commonly been made to support the use of the same discount rate for costs and effects. The time paradox of Keeler and Cretin *(120)* claimed that if a lower rate of discount is used for health effects, postponing a project for one year would lead to an increase in the present value of the CER. The same health effects would accrue, and the same costs. Both would be discounted, but costs would be discounted at a higher rate than the health effects—the present value of the CER would be higher than for the same project undertaken a year earlier. Because of this, they argued that, discounting using lower rates for health effects than for costs would lead decision-makers to delay all health spending indefinitely. The time paradox has now been shown to be irrelevant to the choice of discount rate for a variety of reasons, including the fact that it assumes an infinite time horizon and it assumes that the opportunities for translating expenditure into health do not vary over time *(114;121)*. This is clearly untrue—for example, in many

circumstances it would be more expensive to gain each unit of health next year if opportunities to improve health this year are missed—e.g. epidemic control and preventive interventions.

The second justification is the consistency argument of Weinstein and Stason *(57)*. They argued that the opportunity of individuals or society to transform income into health is constant over time, which requires the CER of interventions to remain constant over time. This will only happen if costs and effects are discounted at the same rate. Again, the assumption that the ability to transfer income into health is constant is not likely to be true, and van Hout has shown that even if it is true, it does not require that the same discount rate should be used *(114)*.

The Washington panel seemed to accept that the discount rate for health should be lower than that for costs if the value of health in terms of income increases over time *(18)*. However, they argued that because CEA had not yet incorporated dynamic wealth effects into the analysis, it was premature to depart from the standard use of the same discount rate for costs and effects. This poses a problem for any analyst. They are correct that most current practice is to use the same rate—only the UK Department of Health has departed from this by recommending that costs be discounted at 6% and health effects at 1.5%.[25]

Accordingly, the standard approach has been used in the base case for WHO-CHOICE, with both costs and health effects discounted at 3%. In the sensitivity analysis (see Section 6) we recommend testing the sensitivity of the results to the use of a 0% discount rate for health effects and a rate of 6% for costs.

Constant rate discounting

Recently it has been argued increasingly that the social discount rate decreases over time. A number of possible reasons have been offered. The first is that individuals express preferences for lives saved that are more consistent with hyperbolic than constant exponential discounting *(122)*. Second, the appropriate discount rate for consumption from economic theory is a function of the pure rate of time preference plus the product of the value of a marginal increase in consumption multiplied by the rate of change of consumption. If the rate of change of consumption is expected to fall over time, so will the discount rate *(123)*. Third, Newell and Pizer *(124)* have argued that when the future path of interest rates is uncertain, but correlated over time, the distant future would be discounted at lower rates than suggested by the current rate. Finally, Reinschmidt *(125)* showed that if the social discount rate is derived from the aggregation of individual discount rates, and if there is a

distribution of discount rates across the population, the social discount rate will be smaller than the average of the individual rates and will decline over time. This holds even if individuals use a constant rate of discounting.

As yet, declining rates of discount have not been used in CEA. They present a problem for decision-makers because the relative attractiveness of a set of interventions could change over time purely because of the time the analysis is undertaken rather than because of any change in the costs or effects of the intervention *(18)*. This raises a number of problems for CEA which are yet to be resolved. Accordingly, we follow the standard practice of using a constant rate.

SUMMARY OF RECOMMENDATIONS

1. Costs and effects should be discounted at 3% in the base-case analysis.
2. In the sensitivity analysis (see Section 6) we recommend testing the sensitivity of the results to a 0% discount rate for health effects and a 6% discount rate for costs.

6   UNCERTAINTY IN COST-EFFECTIVENESS ANALYSIS

6.1 INTRODUCTION

All estimates of costs and effects are subject to uncertainty and the sources can be categorized in a number of ways *(18;126;127)*. Here we focus on three types—parameter uncertainty, model uncertainty and generalizability uncertainty. The question of how policy-makers can make decisions in the face of these uncertainties is also considered.

Parameter uncertainty arises for two reasons. The first is due to sample variation around estimates of variables used to calculate a CER, such as unit costs, adherence rates, and the efficacy of an intervention. The second is because there is no agreement about value judgements required for the CEA analysis—the choice of the appropriate discount rate is an example *(18;128)*.

Model uncertainty relates to uncertainty around the appropriate functional form of a model used to estimate a particular parameter and the explanatory variables that should be included. For GCEA this is most relevant when considering the joint effect of interventions on health. Trial data are often available for the effectiveness or efficacy of interventions undertaken singly, but rarely for the joint impact of two or more interventions undertaken together. In this case, the joint impact needs to be modelled—most commonly assuming a multiplicative relationship between the effectiveness of the individual interventions—but it is not certain that this is "truth" *(18)*.

The third type of uncertainty relates to the need to extrapolate the results of studies. For example, clinical trials of a pharmaceutical product might have been undertaken in a low-risk patient group but policy-makers need to know the cost-effectiveness of the product as it would be used in the general population. Or costs might have been collected sometime in the past, and it is necessary to extrapolate them to the present time period for the CEA.

Methods of addressing these types of uncertainty are considered in turn. The section ends by describing how policy-makers can interpret results where there are multiple, sometimes overlapping, cost-effectiveness uncertainty intervals. This type of question has only recently been addressed in the literature *(129;130)*.

6.2 SENSITIVITY ANALYSIS

Sensitivity analysis shows the impact on the CER of varying different parameters. With "one-way" analysis, each uncertain component of the evaluation is varied individually, while the others retain their base-case specifications, in order to establish the separate effect of each component on the results. A "multi-way" sensitivity analysis involves varying two or more inputs at the same time, and studying the effect on outcomes *(131)*. The analyst could choose to recalculate the CER for a range of plausible values, and if there is some value of the parameter at which the intervention would no longer be considered cost-effective, this threshold value could be identified. The important policy question then is how likely is it that the threshold value of the parameter will occur—i.e. how likely is it that the intervention would not be cost-effective. An alternative is to use "analysis of extremes", in which only the extremes of the range of plausible values are included in a sensitivity analysis, to see if the policy implications would change. If the intervention would be considered to be cost-effective even at the extremes, the analysis is said to be robust to changes in key assumptions.

Sensitivity analysis could be undertaken for any parameter used to construct the CER. However, we suggest that it is better to use probabilistic uncertainty analysis to explore the impact of variability in parameters which can be measured and for which there is an underlying probability distribution *(127)*. Sensitivity analysis is more relevant for variables that cannot be measured and for which there is no probability distribution. For CEA, this applies to the two key social choice variables—the discount rate and age weights.[26] There is no probability distribution for either parameter. The analyst either believes that age-weighting is important or that it is not. The analyst believes that the discount rate for health effects should be zero or that it should be some positive number. The base case for WHO-CHOICE includes age-weighting and a 3% discount rate for both costs and health effects. For sensitivity analysis, the recommendation is to use analysis of extremes for these parameters. This involves recalculating all CERs in the absence of age-weighting and to explore the sensitivity of the results to a zero discount rate for health effects.

6.3 PROBABILISTIC UNCERTAINTY ANALYSIS

Uncertainty around the distribution of parameters such as unit costs, population effectiveness, or initial incidence of disease has traditionally

been dealt with by sensitivity analysis, but recently there has been an interest in applying statistical methods to quantify the effect of these sources of uncertainty. The impact of uncertainty surrounding one parameter or around multiple parameters can be explored. For a decision-maker, the most important piece of information is whether the results are robust to all possible sources of uncertainty at the same time, so the statistical approaches generally consider multiple sources of uncertainty simultaneously. However, it is often useful for the analyst to report uncertainty around key parameters individually as a way of helping policy-makers understand the sources of the overall uncertainty.

The main application of this approach has been to use probabilistic uncertainty analysis employing the method of bootstrapping. For "one-way" uncertainty analysis, repeated draws are drawn from the distribution around each key variable to determine the probability distribution of the cost-effectiveness ratio. The number of draws should be sufficiently large to allow the estimated CER to stabilize, usually a minimum of 1000. From this, 90% uncertainty intervals around the CER can be generated using the simple percentile method, which involves omitting the lower and upper 5% of estimates *(132)*.[27]

For "multi-way" uncertainty analysis one draw is taken from the uncertainty range around each parameter simultaneously. The CER is then estimated. This procedure is repeated a minimum of 1000 times, and the 90% confidence interval calculated for the CER taking into account the variation around all parameters simultaneously.

Bootstrapping can be applied to sampled or non-sampled data. With sampled data, non-parametric bootstrapping is preferable—repeated draws can be taken from the sampled data with no need to specify a particular distribution *(132)*. Where sampled data are not available, the analyst needs to specify the upper and lower limits for each parameter to be used in the draws, and the type of distribution that is likely to characterize the parameter. Analysis can be undertaken using a variety of standard statistical programs—the analysis for WHO-CHOICE is based on @RISK 4.0 (Palisade Decision Tools).

To illustrate the procedure, consider the hypothetical example introduced in Section 2 related to four interventions for tuberculosis: passive case detection and treatment with directly observed therapy, short course (DOTS) (a); BCG vaccination at 50% coverage (b1); BCG vaccination at 75% coverage (b2); BCG vaccination at 100% coverage (b3). This set of interrelated interventions should be evaluated together. This means estimating their cost-effectiveness when they are undertaken individually or in combination. Each of the interactions is defined as a separate intervention: DOTS combined with the different levels of BCG coverage, i.e. ab1, ab2, and ab3 in turn.

Costs and benefits interact: the variable cost component of DOTS decreases when vaccination is given and fewer cases of tuberculosis occur.

The health benefits of BCG vaccination will be less in the presence of a treatment programme because many of the deaths from tuberculosis expected in the absence of treatment will be avoided. Total costs and effects of the interventions at the population level were presented in Section 2.2 and are reproduced in Table 6.1.

Table 6.1   Hypothetical costs and health benefits of TB interventions

Intervention	Costs	Benefits	Mean CER	90% Uncertainty Interval (low)	90% Uncertainty Interval (high)
a	550 000	500	1 100	877	1 360
b1	180 000	200	900	451	1 730
b2	325 000	300	1 083	734	1 581
b3	600 000	400	1 500	1 194	1 954
ab1	631 000	600	1 052	878	1 251
ab2	726 500	650	1 118	942	1 338
ab3	952 000	700	1 360	1 171	1 572

In Section 2.2 it was shown that the expansion path based on the point estimates of the CER involved introducing interventions in the following order, depending on resource availability and assuming that costs and effects were measured with certainty: BCG at 50% coverage (b1), BCG at 50% coverage combined with passive detection and treatment (ab1), BCG at 75% coverage combined with detection and treatment (ab2) and BCG at 100% coverage with detection and treatment (ab3). Table 6.1 shows the results of assuming that costs and effects are no longer measured with certainty. Costs are assumed to be log-normally distributed with a standard deviation of 50; effects are assumed to be normally distributed with a standard deviation of 50. The covariance between costs and effects is assumed to be zero.[28] All three assumptions can be varied as necessary.

A correlation between the costs of the different interventions, and between their effects, was assumed. For example, if the price of the BCG vaccine is a source of uncertainty and could take a value between $0.50 and $2.50 per dose, it does not make sense to allow the iterations to choose $0.50 per dose for coverage at 75% and $2.50 per dose for coverage at 50. If the cost happens to be higher than expected, it would be higher regardless of which intervention the policy-maker chose. The same price, therefore, should be used in each iteration for all interventions using the BCG vaccine.[29] The same is true for effects where the same level of individual effectiveness for the BCG vaccine must be used in any given iteration for all interventions using the vaccine.

To account for these interactions in our current example, we arbitrarily assume a correlation coefficient of 0.9 for the uncertainty in costs between different coverage levels of the same intervention (e.g. between b1 and b2), and a correlation coefficient of 0.6 between individual interventions and combinations of interventions (e.g. between b1 and ab1). The same correlations were also used for population effectiveness. More complicated assumptions could be built-in depending on the nature of the likely correlations.[30]

Table 6.1 also shows the 90% uncertainty intervals around the point estimate of the CERs for the seven interventions. Although b1 has the lowest mean CER, its uncertainty interval overlaps with that for a and ab1. The usual interpretation is that it is not possible to be 100% sure that b1 would be the first choice, although at low levels of resource availability (e.g. at a total cost of $180 000), b1 is the only choice because the other options are not affordable.

To generate uncertainty for the full expansion path, the incremental CER of all interventions compared with the best option could then be calculated with their uncertainty intervals. The steps are:

1. Take one draw of costs (C) and effects (E) from the distributions of costs and effects from b1: C_{b1} and E_{b1}, and one draw of cost and effects from the distribution of costs and effects of b2: C_{b2} and E_{b2} allowing for the correlations described earlier.
2. Estimate the incremental CER for that draw as C_{b2}–C_{b1} divided by E_{b2}–E_{b1}.
3. Repeat this process a large number of times (minimum 1000) to obtain a vector of estimates which is the empirical sampling distribution of the incremental CER statistic.
4. Repeat steps 1–3 for all interventions that could be added after b1—in this case only ab1 in addition to b2.
5. Repeat these steps for each addition to the expansion path.

Table 6.2 shows the results assuming that b1 is initially chosen. Although the point estimate of the incremental cost-effectiveness ratio (ICER) of moving from b1 to b2 is higher than that of moving to ab1, the uncertainty intervals overlap. Again, the usual interpretation is that it is not possible to be 100% sure that it is best to move from b1 to ab1 rather than to b2.

Table 6.2   Incremental Cost Effectiveness Ratios (ICER) of TB interventions

Intervention	Mean ICER	90% Uncertainty Interval (low)	90% Uncertainty Interval (high)
b1→b2	1 450	914	2 273
b1→ab1	1 128	886	1 439

The problem with Tables 6.1 and 6.2 is that they are difficult to interpret when there are correlations in both costs and effects between the interventions, as previously discussed. For example, the ICER of moving from b1–b2 of Table 6.2 is correlated with the incremental CER of moving from b1–ab1, because if a high cost of the vaccine is drawn from the distribution of vaccine costs for b1 it would also be drawn for ab1 and for b2. Therefore, the decision whether to move from b1 to ab1 or to b2 cannot be determined by inspecting the uncertainty intervals around the ICER alone (or the CER in Table 6.1), as they do not reveal how the costs and effects of b1–b2 and b1–ab1 interact.[31] The appropriate way to do this is to inspect the paired simulation data to determine how often the ICER of b1–ab1 is lower than the ICER of b1–b2. This is one of the foundations of the stochastic league table approach described below.[32]

To this point the discussion has dealt with parameter uncertainty. Model uncertainty is more difficult to incorporate into CEA in a formal manner and we do not explore this further *(18)*. Generalizability uncertainty can be incorporated in the same way as parameter uncertainty using probabilistic uncertainty analysis. The analyst must simply decide on the likely upper an lower limits of key parameters in the group or time period to which the results will be extrapolated—for example, the upper and lower limits of efficacy for a pharmaceutical product in the general population rather than in the low risk population in which a clinical trial was undertaken.

Reporting uncertainty results

To improve the usefulness of sensitivity and uncertainty analysis to decision-makers and analysts, an explanation should be provided for each parameter that has been varied. This should include the upper and lower limits used for probabilistic uncertainty analysis, the source(s) and the nature of the assumed distribution. Furthermore, providing a summary of the impact on the CER of uncertainty in each key variable separately is useful to understanding the source of overall uncertainty. This can be investigated formally using regression or correlation analysis of the simulation data, a feature contained in @Risk 4.0 (Palisade Decision Tools). The value of this is that it informs the analyst which uncertainty variable impacts most on the CER, and can guide researchers as to what future prospective research is the most beneficial in terms of reducing uncertainty in the CER. For example, in an analysis of a screening programme for hepatitis B surface antigen, the authors reported partial derivatives which showed the impact on cost-effectiveness of a small change in each parameter separately *(127;133)*.

6.4 Policy-making under uncertainty

Traditionally, the CERs reported in Tables 6.1 and 6.2, minus their uncertainty intervals, would be placed in a single league table to inform decision-makers about the relative value of a set of interventions. Rank ordering in the league table approach would be made on the basis of point estimates of the CER or ICER. Increasingly, however, uncertainty intervals are also being reported. Where uncertainty intervals overlap, as in the TB example, it is not clear how decision-makers should interpret the results. This problem is common. A recent example was reported by Goodman et al., who argued that because the uncertainty intervals for their estimates of the cost-effectiveness of interventions against malaria overlapped, they could not decide which ones should be given preference in the event of a shortage of resources *(134)*.

We believe that additional information is contained in the data used to produce the uncertainty intervals, information that could be used to guide policy-makers more than simply saying that it is not possible to be sure which intervention is preferable. WHO recommends the use of stochastic league tables and has developed a tool for this purpose known as MCLeague. The approach shows the likelihood that a single intervention, or a particular mix of interventions, maximizes health gain (i.e. the optimal mix) at any given budget level given the uncertainty surrounding all competing interventions.

The construction of stochastic league tables requires four steps described in more detail elsewhere *(135)*. In the first, CERs are calculated for the respective programs by drawing single samples from distributions of both costs and effects, using Monte Carlo simulations. Distribution of costs and effects can be based on sampled or non-sampled data, and should take into account co-variance between costs and effect distributions. Second, the optimal mix of interventions is defined based on this information, applying the resource allocation decision rules described in Section 1. Third, this exercise is repeated a large number of times (>1000) to obtain a distribution of the number of times each intervention would be chosen for a certain budget. This provides information on the likelihood that each intervention would be included in the optimal mix. Fourth, this procedure is repeated for various budget levels. This provides a "stochastic budget expansion path" which shows the probability that any interventions would be chosen at the different budget levels.

Table 6.3 summarizes the results as probabilities (in percentages) that a particular intervention or combination of interventions in the hypothetical tuberculosis example would be chosen at different levels of resource availability. The same assumptions outlined above were used.

Firstly, consider the case where a new programme is established and low levels of resource are available e.g. $200 000. At this budget level, intervention b1 is chosen but in only 64% of all cases. No other option

appears in the table, implying that in 36% of the iterations, the costs of b1 would exceed the budget limit, and all the other interventions are always too costly. It is important for policy-makers to know that although b1 is the most cost-effective option where resources are scarce, there is some chance (e.g. 36%) that the costs of intervention b1 could exceed the available resources.

Table 6.3 Stochastic league table, with probabilities of inclusion of interventions (%) for different resource availability

	Budget (US$,000)											
	100	200	300	400	500	600	700	800	900	1000	1100	1200
a	0	0	0	0	16	**58**	9	0	0	0	0	0
b1	5	**64**	**69**	5	0	0	0	0	0	0	0	0
b2	0	0	31	**95**	**83**	7	0	0	0	0	0	0
b3	0	0	0	0	2	9	0	0	0	0	0	0
ab1	0	0	0	0	0	26	**62**	8	2	0	0	0
ab2	0	0	0	0	0	1	29	**91**	**84**	16	1	1
ab3	0	0	0	0	0	0	0	0	15	**84**	**99**	**99**

*Bold numbers are the chosen options based on the league table without uncertainty.

Secondly, consider the case where a new programme is established and the resources available amount to $600 000. In this case, the decision is different—in the majority of cases (56% of the time), the decision-maker begins by purchasing intervention a (DOTS). The reason why DOTS is chosen in the majority of cases is due to a complex interplay between the cost-effectiveness ratio, absolute health gain and affordability (absolute costs). For example, combined DOTS/BCG (ab1) may be more cost-effective than DOTS (a), but budget constraints may not allow the decision-maker to purchase this intervention. On the other hand, while b1 may be more cost-effective than a, option a produces a greater absolute health gain and in most cases is affordable at a budget level of $600 000. This demonstrates how stochastic league tables take into account the interactions between efficiency, health gain and affordability.

The expansion path showing which interventions are added after the first intervention is purchased depends on both the magnitude of the additional budget that becomes available as well as which intervention or combination is already funded. In the first example where b1 is purchased before other interventions, if the budget was to increase to $700 000 it is most likely (62%) that the expansion path requires adding

option a to b1, i.e. moving to ab1. From this point, the expansion path is clear—expand coverage of BCG to ab2 at a budget level of $800 000, and then to ab3 at a budget level of $900 000. This is the same expansion path derived (Section 2.2) for the case where uncertainty was not considered.

Consider the alternative scenario where the budget was to increase to a level of $600 000 only. Table 6.3 shows that the optimal mix is likely to be DOTS (a at 58%). In this scenario, however, the decision-maker has already purchased BCG (b1), and funds would need to be cut from BCG in order to fund DOTS—in order to move from b1-a[33]. This is not likely to be feasible given the penalties often associated with shifting resources from an existing intervention to a new intervention[34]. A more likely scenario is that the decision-maker would wait or lobby for more funding until the higher budget of $700 000 is reached where it is more certain that ab1 will be chosen. Alternatively, the decision-maker may choose to fund ab1 anyway, given that there is a 26% chance that ab1 will be in the optimal mix at a budget level of $600 000, or assume that further funds will be available in the near future.

These examples demonstrate that it would not have been possible to identify the appropriate path by considering only the CERs and the uncertainty intervals of Tables 6.1 and 6.2. Stochastic league tables, therefore, present decision-makers with the probability that an intervention will be included in the optimal mix at a given level of resource availability and are more informative than the traditional approach that implies that no decision can be made when uncertainty intervals overlap.

6.5 CONCLUSION

Various methods can be used to show the impact of uncertainty on estimates of cost-effectiveness. Sensitivity analysis is useful to indicate uncertainty related to value judgements, while the bootstrapping approach can be applied to capture uncertainty related to the distribution of parameter estimates.

Simply reporting the uncertainty range around each estimated CER ignores the fact that there can be correlations between the costs, or benefits, of different interventions, and between the costs and benefits of the same intervention. It also avoids the question of how policy-makers should interpret the results where uncertainty intervals overlap and how decisions may be affected by budget constraints. The stochastic league table is a new way of presenting this information in a way that is intuitively obvious to decision-makers. It provides additional information beyond that offered by the traditional treatment of uncertainty, presenting the probability that each intervention should be included in the optimal mix for given levels of resource availability.

SUMMARY OF RECOMMENDATIONS

1. Uncertainty related to variables that carry value judgements should be subjected to one-way, and sometimes multi-way, sensitivity analysis. The base case analysis for WHO-CHOICE involves using a 3% discount rate for both costs and health effects, with age-weighting. We recommend examining the sensitivity of the results to the use of a 0% discount rate for health effects, 6% for costs, and with no age-weighting.
2. Uncertainty related to parameter estimates should be quantified through probabilistic uncertainty analysis using bootstrapping.
3. Stochastic league tables should be used to provide additional information to policy-makers about how to interpret results in the face of uncertainty.

7   POLICY USES OF GENERALIZED CEA

The results of GCEA can be used to guide policy in a number of ways. This section provides an overview of these options with reference to WHO-CHOICE, the database on costs and effects of interventions that WHO has established using GCEA. However, the policy uses are also relevant to other types of CEA.

7.1   GLOBAL DISSEMINATION OF NEW INTERVENTIONS

Some new technologies and approaches to improving health are rapidly disseminated and adopted. Others take a longer time to be accepted. The former group includes technologies that improve the health of people who can afford to purchase them or who have the ability to demand the service in other ways. The latter group includes interventions aimed largely at poor people. GCEA can identify interventions that are potentially very good buys but which are not currently used, either because they are new or because they have not been widely adopted.

Once an intervention has been identified as cost-effective, it can be promoted at an international and national level, shortening the lag time between development and adoption of the technology. For example, the analysis on micronutrient supplementation and fortification undertaken by WHO-CHOICE showed that these interventions were cost-effective in all regions (see www.who.int/evidence/cea). But while vitamin A supplementation has been promoted actively by international agencies and some governments, zinc supplementation has lagged behind and many countries are still not providing this intervention. Reduction of the salt content in processed foods has been shown to be an efficient way of reducing the risk of cardiovascular events, but governments have only recently begun to think about how to encourage this type of action. Information from CEA can help to ensure that new or under-

used interventions are used more rapidly. The corollary is that it can also help to discourage the use of inefficient technologies that are widely used.

7.2 NATIONAL PRIORITY SETTING

7.2.1 PRIORITY SETTING AT THE PROGRAMME LEVEL

GCEA is particularly suited to identifying a set of the most cost-effective interventions that can be used by decision-makers to improve the performance of their health systems. Unlike earlier work on sectoral CEA, WHO-CHOICE has evaluated sets of interventions at different coverage levels and in different combinations. This accounts for non-linearities in cost-functions and for any interactions in the impact of interventions being undertaken simultaneously. Groups of interventions that are interrelated are evaluated together in a cluster, as discussed in Section 2.

The first use is to set priorities within any set of interrelated interventions. This was illustrated in Section 2.2 where a hypothetical example from TB was used to illustrate how an expansion path can be calculated for the set of interrelated interventions and the most appropriate mix chosen for any given level of resources. An example relating to the prevention of cardiovascular disease through reducing blood pressure and cholesterol levels is found in Part Two of this volume, reproducing work originally reported in *The Lancet (136)*. There it is shown that secondary prevention based on the identification and treatment of elevated blood pressure alone, or of elevated levels of cholesterol alone, are not on the expansion path in any subregion. They are less cost-effective than interventions which first identify the overall risk individuals have of suffering a cardiovascular event in the next 10 years, then identified individuals at risk with the combination of a cholesterol lowering agent, a blood pressure-reducing agent and aspirin.

CEA within a group of interrelated interventions is a powerful tool even where there is uncertainty. Often it is very clear that one intervention is both less costly and more effective than another option, for all combinations of assumptions. The analysis can also show which interventions are very costly ways to improve health within any set.

The use of this type of information in identifying potential improvements in the efficiency of the health system is demonstrated in Figure 7.1, which illustrates the maximum gain in health that could be derived if the most efficient set of interventions were chosen to improve child health for any given level of resource use. The "frontier" that is depicted has been developed from an analysis of interventions

versus childhood pneumonia and diarrhoea, and zinc and vitamin A deficiency in one epidemiological subregion of Africa characterized by high adult and high child mortality (called AFRO-D—for a description of WHO-CHOICE and the definition of epidemiological subregions, see Annexes A and G respectively). The gain in healthy life expectancy (HALE) is given on the y-axis (obtained by transforming the DALYs averted by the interventions to HALE gain) and resource use or costs on the x-axis.[35]

Figure 7.1   Maximum possible health gains from selected child health interventions, AFRO-D
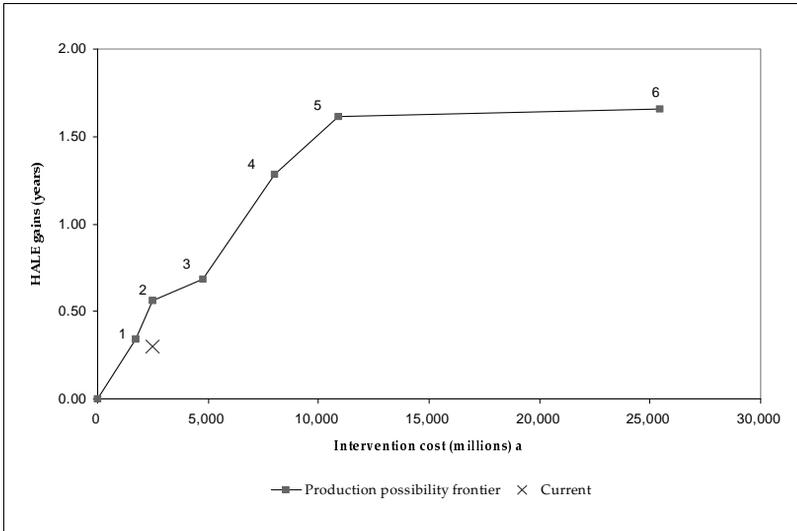
^a International dollars.
1 Vitamin A fortification (VS), zinc fortification (ZF) and case management for pneumonia (CM) at 80%.
2 VF, ZF and CM 95%.
3 VS, ZS and CM 95%.
4 VS, ZS and oral rehydration therapy (ORT) at 80%.
5 VS, ZS, ORT and CM 95%.
6 VS, ZS, provision of targeted supplementary feeding, ORT and CM 95%.

The costs of the current set of interventions used in AFRO-D and the associated gains in HALE were then estimated, represented by point X. This implies that it would be possible for countries in AFRO-D to reallocate the resources currently devoted to interventions focusing on children under the age of five years in a way that would achieve more health than is currently the case. This can be done, for example, by increasing the provision of micronutrients, particularly zinc but also vitamin A, either through fortification or supplementation.

7.2.2 PRIORITY SETTING AT THE SECTORAL LEVEL

The potential improvement of efficiency is even greater if a sectoral approach is taken and all interventions are considered together. Figure

7.2 represents the efficiency of a health system in the production of overall health. The x-axis denotes the inputs used to achieve health improvements. The y-axis represents the level of health. Line L represents the minimum health level that would be observed in the absence of a health system or the absence of any health expenditure—it is not zero because people would still be alive even if the health system did not exist. Line M represents the maximum level of health that could be achieved for any given level of resources, or the production frontier for the system as a whole, as opposed to the case of Figure 7.1 which represented only the frontier for interventions focusing on child health. It reflects the fact that increasing expenditure is associated with increasing health.

Figure 7.2   Health system efficiency and cost-effectiveness
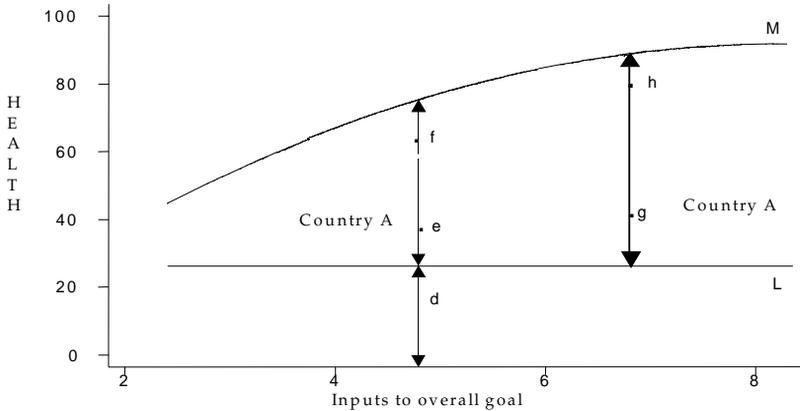
Country A is observed to provide a set of interventions resulting in the costs and health level of point e. Efficiency is usually defined as the level of actual goal attainment divided by the maximum that would have been possible for the resources available. In this case, because some level of health would exist even if no resources were spent (e.g. line L), it is defined as the health gain achieved above the minimum possible (L), divided by the maximum health gain that would have been possible for those inputs (also above the minimum). Efficiency at e is the distance from line L to e in a vertical direction, divided by the distance from L to M at that point. Efficiency at g is Lg divided by LM. Assuming all other variables contributing to health are held constant, countries below line M are producing less health than is possible for their given level of resources.

Inefficiencies in the production of health may derive from two sources: problems with technical efficiency—how an intervention is delivered— and problems with allocative inefficiency—which set of interventions is

provided. If a spontaneous vaginal delivery at a health facility utilizes seven days of in-patient stay, but the same health outcomes could have been obtained with an in-patient stay of 48 hours or less, or even a home birth, there is technical inefficiency. A given health gain is obtained at a higher than necessary cost. If several magnetic resonance imaging devices have been purchased and placed in health facilities within walking distance of each other, resulting in under-utilization of those devices, there is again technical inefficiency. The same benefits of MRI technology could have been obtained through selective placement of the machines in referral facilities, at lower cost.

Allocative efficiency traditionally is used to describe the optimal mix of inputs (such as capital, labour and supplies) to a production process, given their respective prices. As interventions are inputs to the production of health, allocative efficiency requires choosing the most cost-effective mix of interventions for any set of resource constraints. Country A operating initially at point e could improve health by spending more—moving to g, for example. The alternative would be to change the mix of interventions it is providing. It could have reallocated existing resources from cost-ineffective to cost-effective interventions, gaining more health for the same resources, e.g. moving to point f. If it has additional resources to invest, by choosing a more cost-effective mix of interventions and spending more, it could move from point e to h.

This illustrates how GCEA can help decision-makers to assess and potentially improve the performance of their health systems in terms of one goal, the level of health. It indicates which sets of interventions provide the highest "value for money" and helps policy-makers choose the interventions and programmes which maximize health for the available resources. In principle, GCEA could also be used to define the overall frontier M for the entire health sector. This would require information on the entire set of intervention options but, at least in theory, it is a way of assessing the efficiency of the overall system—similar to the situation depicted in Figure 7.1 but for the entire sector.

7.3   Reimbursement and financing decisions

Related to the uses described above, GCEA can also be used to guide or re-examine financing decisions. It can help inform decisions on whether to fully reimburse, subsidize, or refuse to cover the costs of providing a service. It could be used to decide the extent or frequency of coverage—for example, for screening programmes. This provides valuable information for a health insurance scheme covering all types of health interventions, or a component of the health system such as a hospital. On the grounds of efficiency, it can be argued that there should be no attempt to provide cost-ineffective interventions. This type of use of CEA is becoming increasingly common—for example, the Pharmaceutical Benefits Advisory Committee in Australia takes cost-effectiveness into

account when making decisions on which new drugs will be publicly reimbursed. The National Institute for Clinical Excellence in the United Kingdom has also used cost-effectiveness information in providing guidance for the use of new drugs in the National Health Service.

7.4 Research and development priorities

WHO-CHOICE has analysed approximately 200 interventions at a subregional level. An immediate research need is to expand the number of interventions in the database and to contextualize the results to as many countries as possible. Analysts in countries can contribute to both activities by contacting the WHO-CHOICE project team.

GCEA can be used to estimate the contribution of interventions, or combinations of interventions, to decreasing the burden of disease. If it is shown, for example, that all combinations of cost-effective interventions together have a relatively small impact on the total burden of a particular disease or risk factor, research into new ways of reducing this burden is required. Interventions targeting child under-nutrition illustrate this. They are relatively costly and not very effective. Research to improve the effectiveness of current technologies, to reduce their costs, or to develop new technologies is warranted.

A variation of this theme is that technologies may exist but there may be system-wide constraints which prevent them being used. For example, access to skilled midwives is a cost-effective way to reduce maternal mortality, but there might be a shortage of skilled midwives that prevent this intervention being scaled-up to high levels of coverage. It is critical for decision-makers to know if a high disease burden is due to the lack of cost-effective intervention options, or if it is due more to health system constraints.

The final possibility is that a cost-effective intervention exists but is not widely used. Research is needed to determine why this is the case—it may, for example, be related to cost, or perhaps providers or members of the community are not convinced that it is effective—and to examine how it can be used more widely.

7.5 Goals and functions of health systems

Cost-effectiveness analysis focuses on the improvements in health that result from different choices about how health resources should be used. It is important to remember that improving health is only one goal of health systems. According to the WHO framework of health systems performance assessment, there are five indicators of the three intrinsic social goals to which the health system contributes: namely, improving the level and distribution of health, improving the level and distribution of responsiveness and ensuring that the financial burden of paying for the health system is distributed fairly (137) (see Table 7.1).

Table 7.1   WHO health system performance framework

Goal	Level	Distribution	
1. Health	X	X	EFFICIENCY
2. Responsiveness	X	X	
3. Fair financing		X	
	GOODNESS	FAIRNESS	

This means that the results of CEA should *not* be used formulaically. Cost-effectiveness provides information on how current resources and any new resources could be allocated to obtain the greatest possible improvement in population health. This enters the policy debate and decision-makers then must weigh the costs of changing the intervention mix and the impact of different mixes against other goals of the health system. In fact, CEA at the sectoral level is probably most powerful when it is used to classify interventions into broad groups. In the first round of WHO-CHOICE three categories were used—those that are: very cost-effective, cost-effective, and not cost-effective. Policy-makers would be encouraged to choose from the first set, and to avoid the third, other things being equal, but they would also need to assess the impact of any proposed mix of interventions on poverty and other types of inequality, for example.

To illustrate, in Figure 7.1 the intervention combination nearest the upper right corner includes targeted provision of supplementary food to infants. This combination cannot be considered cost-effective for the available resources. However, even if it exceeds the threshold of what is considered to be cost-effective in that setting, countries in the region might opt to provide it on equity grounds because under-nutrition has a disproportionately high burden in the poor. At the same time, having identified that there is no cost-effective intervention against under-nutrition, policy-makers could also recommend setting aside research funds to determine how to decrease the costs or improve the effects of the interventions, or even fund a research programme with the intent to discover a different technology altogether.

7.6 ETHICAL ISSUES

A number of ethical issues may arise when using CEA for health care resource prioritization. For example, how can concerns for equity or justice be incorporated in decision-making in addition to the concerns for efficiency and benefit maximization? And, should all QALYs or DALYs count equally regardless of the age of the recipient of the health benefit?

Another questions is what priority should be given to the sickest or worst off? Since many of the issues are relevant to GCEA as well as to other forms of CEA, the interested reader is referred to an overview in Part Two on ethical issues in CEA and for a more thorough discussion, to *Fairness and Goodness: Ethical Issues In Health Resource Allocation (138)*.

8   REPORTING CEA RESULTS

Reports on CEA results must provide sufficient information to enable independent analysts to critically evaluate the estimates of the costs and effectiveness of the interventions studied. In addition, they should be able to interpret the findings of the CEA and assess the possibility of generalizing them to their own decision-making context.[36] Since it may not always be possible to document this information in a journal article, additional information should be provided in background reports or on the World Wide Web. To enhance transparency and ensure accountability, all reports and all data inputs, including assumptions, used in deriving the estimates, should be placed in the public domain.

A CEA report usually contains, or indicates sources for, a detailed description of the inputs and methods used to estimate costs, effectiveness and cost-effectiveness ratios of the interventions studied. The ten-point checklist introduced by Drummond et al. *(17)* or a similar format may be used as a guide to analysts seeking to improve the quality of their study reports.

The following section outlines the key information to be reported with respect to the elements of CEA. A short description of WHO's approach to reporting GCEA results can be found in Annex A.

8.1 COST INFORMATION

Reports should contain or discuss:

- information on unit prices and quantities for the main factor inputs used to estimate programme costs (e.g. personnel, vehicles, office space etc.);
- how patient costs were estimated—for example the cost per visit or bed-day, the costs of laboratory tests—and what assumptions were used, including questions of intervention coverage levels, capacity utilization, depreciation rates used to obtain capital costs, etc.;
- whether the costs used in the study have face validity, in terms of other costs reported in the literature, for example, and whether they were

obtained from a sample of costs that are likely to be representative rather than based on a single observation;
- results of sensitivity and uncertainty analysis; and
- space permitting (e.g. for web-based presentation of results), a detailed listing of quantities and prices of factor inputs used in the analysis.

8.2 Effectiveness information

Reports should contain or discuss:

- whether a systematic search for evidence on baseline epidemiology and effectiveness was undertaken, the criteria used for selecting sources, the assumptions made, etc.;
- quantitative documentation of the sources and assumptions used for: (1) the main input variables in the analysis such as prevalence, incidence or remission rates and relative risk ratios, all of which should be reported for both the null and the intervention scenarios; (2) how the effectiveness of each intervention was modelled (e.g. through a decrease in incidence, in duration, in remission, or in mortality rates); and (3) other factors related to modelling health effects such as intervention coverage rates, patient adherence to medicines and follow-up visits and quality of services provided;
- the healthy life years lived by the population under both the null and interventions scenarios, and the difference between the two scenarios—representing the health gain of the intervention; and
- results of sensitivity and uncertainty analysis.

8.3 Cost-effectiveness ratios

Reports should contain or discuss:

- both a numerical and a graphical documentation of cost-effectiveness ratios;
- cost-effectiveness ratios compared to the null for all interventions studied, and incremental cost-effectiveness ratios for those interventions on the expansion path;
- expansion paths clearly identified either in tabular or graphical form, for each set of inter-dependent interventions;
- results of uncertainty analysis including use of stochastic league tables where appropriate.

SUMMARY OF RECOMMENDATIONS

1. Reports on CEA results must provide sufficient information in the public domain to enable independent analysts and policy-makers to critically evaluate the validity of the estimates of the costs and effectiveness of the interventions studied.

9   SUMMARY OF RECOMMENDATIONS

Overall study design

1. Groups of interventions where there are major interactions in either costs or health effects should be evaluated together.
2. Analysts should evaluate all interventions initially against the "null", i.e. the situation that would exist if none of the set of interacting interventions were implemented.
3. Interventions should be described in detail, which includes information on the setting, target population, time frame, regimen, and frequency of obtaining the intervention.
4. All interventions should be evaluated under the assumption that they are implemented over a period of 10 years. However, costs and health effects related to the intervention should be followed for the duration of the lifetime of the beneficiaries. This could be varied by country-analysts adapting the results or undertaking studies in their own settings.
5. Resource use and health effects should be identified and valued from the societal perspective.

Estimating costs

1. Ideally, analysts should follow the ingredients approach and collect and report information on the quantities and prices of the resources used in addition to total expenditures.
2. The cost of providing health interventions should be included in the analysis as should the resources used up in seeking or obtaining an intervention (e.g. transport costs). It is recommended that productivity gains and losses due to an intervention, including time costs of seeking or obtaining care, should be excluded from the CEA. Where they are believed to be particularly important, they should be measured (rigorously) in physical units (e.g. time gains or losses) and reported separately.

3. Transfer payments should not be included in CEA. However, any related administrative costs should be included.
4. Costs of central administration and the education of health professionals can be regarded as existing or ongoing costs and should not be included in the analysis. This does not include training costs for a specific intervention, which should be included.
5. Shadow pricing should be used to determine the economic costs of goods that have no market price or if market prices are believed to have major distortions.
6. Prices of traded and non-traded goods should, in theory, be expressed in terms of a common numeraire, and we recommend using the world (international) price level to allow for comparability of results.
7. The annual costs of capital investments can be approximated by their rental price where a rental market exists and works relatively well. But because this is often not the case, the preferred approach is to annualized them taking into account purchase value, resale value, interest rate and working life.
8. Costs should be discounted at an annual rate of 3% in the base analysis. The sensitivity of the results to using a 6% rate should also be explored (see Section 6).
9. Analysts should report the capacity utilization that drives their cost-effectiveness estimates. WHO-CHOICE consistently uses 80% capacity utilization to obtain estimates of the cost-effectiveness of interventions if they are undertaken relatively efficiently.
10. Prices should be adjusted to a common year using the GDP deflator where possible. If this is not available, the Consumer Price Index can be used.

Estimating health effects

1. It is recommended that analysts express population effectiveness in terms of DALYs, although measures such as QALYs and HYL could also be used.
2. For interventions that alter life expectancy, years of life saved by an intervention should be estimated from a population model.
3. The counterfactual scenario for estimating population effectiveness is the null set, defined as the lifetime health experience of a defined population in a situation where all related interventions directed against a disease or condition are stopped. The null set can be estimated using natural history models, using trial data or by back-adjusting using coverage rates and effectiveness of currently implemented interventions.
4. Data on the efficacy of interventions ideally comes from systematic reviews of studies. Efficacy can be expressed as relative risks for rates and effects sizes for means. Efficacy should be adjusted to reflect population effectiveness, taking into account factors like coverage,

quality of care, adherence and other local factors, all of which can modify efficacy.

5. The health state valuations derived in the GBD study can be used until regional estimates are available from WHO. Results should be presented with and without age-weighting as part of the sensitivity analysis (see Section 6).

Discounting

1. Costs and effects should be discounted at 3% in the base-case analysis.
2. In the sensitivity analysis (see Section 6) we recommend testing the sensitivity of the results to a 0% discount rate for health effects and a 6% discount rate for costs.

Sensitivity and uncertainty analysis

1. Uncertainty related to variables that carry value judgements should be subjected to one-way, and sometimes multi-way, sensitivity analysis. The base case analysis for WHO-CHOICE involves using a 3% discount rate for both costs and health effects, with age-weighting. We recommend examining the sensitivity of the results to the use of a 0% discount rate for health effects, 6% for costs, and with no age-weighting.
2. Uncertainty related to parameter estimates should be quantified through probabilistic uncertainty analysis using bootstrapping.
3. Stochastic league tables should be used to provide additional information to policy-makers about how to interpret results in the face of uncertainty.

Reporting CEA results

1. Reports on CEA results must provide sufficient information in the public domain to enable independent analysts and policy-makers to critically evaluate the validity of the estimates of the costs and effectiveness of the interventions studied.

REFERENCES

1. Elixhauser A, Luce BR, Taylor WR, Reblando J. Health care CBA/CEA: an update on the growth and composition of the literature. *Medical Care* 1993; 31(7 Suppl):JS1–149.

2. Elixhauser A, Halpern M, Schmier J, Luce BR. Health care CBA and CEA from 1991 to 1996: an updated bibliography. *Medical Care* 1998; 36(5 Suppl):MS1–147.

3. Johannesson M, Weinstein MC. On the decision rules of cost-effectiveness analysis [comment] [see comments]. *Journal of Health Economics* 1993; 12(4):459–467.

4. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *The New England Journal of Medicine* 1977; 296(13):716–721.

5. Dixon J, Welch HG. Priority setting: lessons from Oregon [see comments]. *Lancet* 1991; 337(8746):891–894.

6. Jamison DT, Mosley WH, Measham AR, Bobadilla JL. *Disease control priorities in developing countries*. New York: Oxford University Press; 1993.

7. Tengs TO, Adams ME, Pliskin JS, Safran DG, Siegel JE, Weinstein MC et al. Five-hundred life-saving interventions and their cost-effectiveness [see comments]. *Risk Analysis* 1995; 15(3):369–390.

8. Bobadilla JL, Cowley P, Musgrove P, Saxenian H. Design, content and financing of an essential national package of health services. *Bulletin of the World Health Organization* 1994; 72(4):653–662.

9. Bobadilla JL. *Searching for essential health services in low- and middle-income countries: a review of recent studies on health priorities*. Washington D.C.: Human Development Department, World Bank; 1996.

10. Eddy DM. Oregon's methods. Did cost-effectiveness analysis fail? *JAMA: the Journal of the American Medical Association* 1991; 266(15):2135–2141.

11. Leichter HM. Oregon's bold experiment: whatever happened to rationing? *Journal of Health Politics, Policy and Law* 1999; 24(1):147–160.

12. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? *Social Science and Medicine* 1993; 37(1):33–40.

13. Alban A, Gyldmark M, Pedersen AV, Sogaard J. The Danish approach to standards for economic evaluation methodologies. *PharmacoEconomics* 1997; 12(6):627–636.

14. Canadian Coordinating Office for Health Technology Assessment. *Guidelines for economic evaluation of pharmaceuticals*. 2nd ed. Ottawa: CCOHTA; 1997.

15. Commonwealth of Australia. *Guidelines for pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including economic analyses*. Canberra: Department of Health and Community Services; 1995.

16. Creese A, Parker D. *Cost analysis in primary health care: a training manual for programme managers*. Geneva: World Health Organization; 1994.

17. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 2nd ed. Oxford: Oxford University Press; 1997.

18. *Cost-effectiveness in health and medicine*. Gold MR, Siegel JE, Russel LB, Weinstein MC, eds. New York: Oxford University Press; 1996.

19. Health-Care Economics Consensus Group. Health economic guidelines: the German perspective revisited. *Pharmacoeconomics & Outcomes News* 1996; 56:4–6.

20. Kumaranayake L, Pepperall J, Goodman H, Mills A, Walker D. *Costing guidelines for HIV/AIDS prevention strategies. A companion volume to: cost analysis in primary health care: a training manual for programme managers*. London: Health Economics and Financing Programme, London School of Hygiene and Tropical Medicine; 2000.

21. Langley PC. The November 1995 revised Australian guidelines for the economic evaluation of pharmaceuticals. *PharmacoEconomics* 1996; 9(4):341–352.

22. Medicare Services Advisory Committee. *Funding for new medical technologies and procedures: application and assessment guidelines*. Canberra: Department of Health and Aged Care; 2000.

23. Murray CJ, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Economics* 2000; 9(3):235–251.

24. Ministry of Health. *Ontario guidelines for economic analysis of pharmaceutical products*. Ontario: Ministry of Health. 1994.

25. Oostenbrink JB, Koopmanschap MA, Rutten FFH. Standardisation of costs; the Dutch manual for costing in economic evaluations. *PharmacoEconomics* 2002; 20:443–454.

26. Phillips M, Mills A, Dye C. *Guidelines for cost-effectiveness analysis of vector control*. Geneva: World Health Organization; 1993. PEEM Guidelines Series 3.

27. Shepard DS, Hodgkin D, Anthony YE. *Analysis of hospital costs: a manual for managers*. Geneva: World Health Organization; 2000.

28. World Health Organization. *Expanded Programme on Immunization: Costing guidelines*. Geneva, World Health Organization, 1997 (unpublished document EPI/GEN/790/5).

29. World Health Organization. *Safe blood and blood products: costing blood transfusion services*. Geneva: World Health Organization; 1998.

30. World Health Organization. *Generic protocols for cost and cost-effectiveness analysis of TB diagnosis and treatment services*. Geneva: World Health Organization; 1999.

31. Chen MM, Bush JW. Maximizing health system output with political and administrative constraints using mathematical programming. *Inquiry* 1976; XIII(September):215–227.

32. Murray CJL, Kreuser J, Whang W. Cost-effectiveness analysis and policy choices: investing in health systems. *Bulletin of the World Health Organization* 1994; 72(4):663–674.

33. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *Journal of Health Economics* 1996; 15(5):641–653.

34. Murray CJL, Kreuser J, Whang W. Cost-effectiveness analysis and policy choices: investing in health systems. *Bulletin of the World Health Organization* 1994; 72(4):663–674.

35. Karlsson G, Johannesson M. Cost-effectiveness analysis and capital costs. *Social Science and Medicine* 1998; 46(9):1183–1191.

36. Karlsson G, Johannesson M. The decision rules of cost-effectiveness analysis. *PharmacoEconomics* 1996; 9(2):113–120.

37. Murray CJL, Lopez AD. *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. 1st ed. Murray CJL, Lopez AD, eds. Cambridge, MA: Harvard University Press; 1996.

38. Brouwer WBF, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *Journal of Health Economics* 2000; 19(4):439–459.

39. Garber AM. Advances in CE analysis. In: Culyer AJ, Newhouse JP, editors. *Handbook of health economics, Vol. 1A*. Amsterdam: Elsevier; 2000.

40. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 1997; 16(1):1–31.

41. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* 1997; 16(1):33–64.

42. McGuire A. Theoretical concepts in the economic evaluation of health care. In: Drummond M, McGuire A, eds. *Economic evaluation in health care. Merging theory with practice*. Oxford: Oxford University Press; 2001. p. 1–21.

43. Birch S, Donaldson C. Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism? *Social Science and Medicine* 2003; 56(5):1121–1133.

44. Arrow KJ. Uncertainty and the welfare economics of medical care. *American Economic Review* 1963; 53(5):940–973.

45. Arrow KJ. Agency and the market. In: Arrow KJ, Intrilligator MD, eds. *Handbook of mathematical economics*. Amsterdam: Elsevier; 1986.

46. McGuire AJ, Henderson JB, Mooney GH. *The economics of health care*. London: Routledge and Kegan Paul; 1989.

47. Mooney G. *Key issues in health economics*. London: Harvester; 1994.

48. Atkinson AB. *The economics of inequality*. 2nd ed. New York: Oxford University Press; 1983.

49. Garber AM. Advances in CE analysis. In: Culyer AJ, Newhouse JP, eds. *Handbook of health economics, Vol. 1A*. Amsterdam: Elsevier; 2000.

50. Birch S, Donaldson C. Cost-benefit analysis: dealing with the problems of indivisible projects and fixed budgets. *Health Policy* 1987; 7(1):61–72.

51. Birch S, Gafni A. Cost effectiveness/utility analyses. Do current decision rules lead us to where we want to be? *Journal of Health Economics* 1992; 11(3):279–296.

52. Viscusi WK. Discounting the health effects for medical decisions. In: Sloan F, ed. *Valuing health care: costs, benefits, and effectiveness of pharmaceuticals and other medical technologies*. New York: Cambridge University Press; 1996. p. 125–144.

53. van Hout BA. Discounting costs and effects: a reconsideration. *Health Economics* 1998; 7(7):581–594.

54. Commission on Macroeconomics and Health. *Macroeconomics and health: investing in health for economic development. Report of the Commission on Macroeconomics and Health*. Available at: http://www.cid.harvard.edu/cidcmh/ CMHReport.pdf. 2001.

55. Chadeau A. What is households' non-market production worth? *OECD Economic Studies* 1992; 0(18):85–103.

56. Meltzer D, Weckerle C, Chang LM. Do people consider financial effects in answering quality of life questions? *Medical Decision Making* 1999; 19(4):517.

57. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *The New England Journal of Medicine* 1977; 296(13):716–721.

58. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 1997; 16(1):1–31.

59. Creese A, Parker D. *Cost analysis in primary health care: a training manual for programme managers*. Geneva: World Health Organization; 1994.

60. Adam T, Evans DB, Koopmanschap MA. Cost-effectiveness analysis: can we reduce variability in costing methods? *International Journal of Technology Assessment in Health Care* 2003; 19(2):406–419.

61. Johns B, Baltussen R, Hutubessy RCW. Programme costs in the economic evaluation of health interventions. *Cost-effectiveness and Resource Allocation* 2003; 1(1). Available at: http://www.resource-allocation.com/content/1/1/1.

62. Hutton G, Baltussen RM. *Valuation of goods in cost-effectiveness analysis: notions of opportunity costs and transferability across time and countries*.

Geneva: World Health Organization; 2003. Global Programme on Evidence for Health Policy Discussion Paper (forthcoming).

63. Gittinger J. *Economic analysis of agricultural projects*. John Hopkins Press; 1984.

64. Little IMD, Mirrlees JA. *Project appraisal and planning for developing countries*. London: Heinemann Educational Books; 1974.

65. Adam T, Aikins M, Evans D. *CostIt software*. Available at: www.who.int/ evidence/cea. 2002.

66. Asian Development Bank. *Guidelines for the economic analysis of projects*. Manila: Asian Development Bank; 1997.

67. Creese A. The economic evaluation of immunization programmes. In: Lee K, Mills A, eds. *The economics of health in developing countries*. Oxford University Press; 1983.

68. Adam T, Evans DB, Murray CJL. Econometric estimation of country-specific hospital costs. *Cost-effectiveness and Resource Allocation* 2003; 1(3). Available at: http://www.resource-allocation.com/content/1/1/3.

69. Johns B, Adam T, Tan Torres T, Evans DB. Econometric estimation of country-specific programme costs. *(forthcoming)* 2003.

70. Luce BR, Elixhauser A. *Standards for socioeconomic evaluation of health care products and services*. Culyer AJ, Horisberger B, eds. Berlin: Springer-Verlag; 1990.

71. Kumaranayeke L. The real and the nominal? Making inflationary adjustments to cost and other economic data. *Health Policy and Planning* 2000; 15:230–234.

72. Decision analytic modelling in the economic evaluation of health technologies. A consensus statement. Consensus Conference on Guidelines on Economic Modelling in Health Technology Assessment. *PharmacoEconomics* 2000; 17(5):443–444.

73. Buxton MJ, Drummond MF, van Hout BA, Prince RL, Sheldon TA, Szucs T et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Economics* 1997; 6(3):217–227.

74. World Bank. *World development report 1993: investing in health*. New York: Oxford University Press; 1993.

75. Fox-Rushby JA, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. *Health Policy and Planning* 2001; 16(3):326–331.

76. Murray CJ, Lauer JA, Hutubessy RC, Niessen L, Tomijima N, Rodgers A et al. Effectiveness and costs of interventions to lower systolic blood pressure and cholesterol: a global and regional analysis on reduction of cardiovascular-disease risk. *Lancet* 2003; 361(9359):717–725.

77. Ranson MK, Jha P, Chaloupka FJ, Nguyen SN. Global and regional estimates of the effectiveness and cost-effectiveness of price increases and other tobacco control policies. *Nicotine & Tobacco Research* 2002; 4(3):311–319.

78. Marseille E, Kahn JG, Mmiro F, Guay L, Musoke P, Fowler MG et al. Cost effectiveness of single-dose nevirapine regimen for mothers and babies to

decrease vertical HIV-1 transmission in sub-Saharan Africa. *Lancet* 1999; 354(9181):803–809.

79. Goodman CA, Coleman PG, Mills AJ. Cost-effectiveness of malaria control in sub-Saharan Africa. *Lancet* 1999; 354(9176):378–385.

80. Politi C, Carrin G, Evans D, Kuzoe FA, Cattand PD. Cost-effectiveness analysis of alternative treatments of African gambiense trypanosomiasis in Uganda. *Health Economics* 1995; 4(4):273–287.

81. Murray CJL, Lopez AD. *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. 1st ed. Murray CJL, Lopez AD, eds. Cambridge, MA: Harvard University Press; 1996.

82. Murray CJL, Lopez AD. *Global health statistics: a compendium of incidence, prevalence and mortality estimates for over 200 conditions*. 2nd ed. Cambridge, MA: Harvard University Press; 1996.

83. Fox-Rushby JA. *Disability adjusted life years (DALYs) for decision-making? An overview of the literature*. London: Office of Health Economics; 2002.

84. Mathers CD, Salomon JA, Murray CJL, Lopez AD. Alternative summary measures of average population health. In: Murray CJL, Evans DB, eds. *Health systems performance assessment: debate, methods, and empiricism*. Geneva: World Health Organization; 2003.

85. Salomon JA, Murray CJL, Ustun TB, Chatterji S. Health state valuations in summary measures of population health. In: Murray CJL, Evans DB, eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva: World Health Organization; 2003.

86. Salomon JA, Murray CJL. A multi-method approach to measuring health state valuations. *Health Economics* 2003; (in press).

87. Murray CJL, Salomon JA, Mathers CD. A critical examination of summary measures of population health. *Bulletin of the World Health Organization* 2000; 78(8):981–994.

88. Salomon JA, Mathers CD, Chatterji S, Sadana R, Ustun TB, Murray CJL. Quantifying individual levels of health: definitions, concepts, and measurement issues. In: Murray CJL, Evans DB, eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva: World Health Organization; 2003.

89. Broome D. Measuring the burden of disease by aggregating well-being. In: Murray CJL, Lopez AD, Salomon JA, Mathers CD, eds. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002.

90. Gafni A, Birch S. QALYs and HYEs (healthy years equivalent). Spotting the differences. *Journal of Health Economics* 1997; 16(5):601–608.

91. Dachs N. An equity motivated indicator of population health. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, eds. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002. p. 161–169.

92. Brock D. Separate spheres and indirect benefits. *Cost-effectiveness and Resource Allocation* 2003.

93. Brock DW. The separability of health and well-being. In: Murray CJL, Lopez AD, Salomon JA, Mathers CD, eds. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002.

94. Ubel PA, Loewenstein G. Distributing scarce livers: the moral reasoning of the general public. *Social Science and Medicine* 1996; 42(7):1049–1055.

95. Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. *Health Economics* 2000; 9(2):137–148.

96. Anand S, Hanson K. Disability-adjusted life years: a critical review. *Journal of Health Economics* 1997; 16(6):685–702.

97. Murray CJL. Rethinking DALYs. In: Murray CJ, Lopez AD, eds. *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard University Press; 1996. p. 1–98.

98. Tsuchiya A. Age-related preferences and age weighting health benefits. *Social Science and Medicine* 1999; 48(2):267–276.

99. Jelsma J, Shumba D, Kristian H, De Weerdt W, De Cock P. Preferences of urban Zimbabweans for health and life lived at different ages. *Bulletin of the World Health Organization* 2002; 80(3):204–209.

100. Preston SH. Health indices as a guide to health sector planning: a demographic critique. In: Gribble JN, Preston SH, eds. *The epidemiological transition, policy and planning implications for developing countries*. Washington DC: National Academy Press; 1993.

101. Hrobjartsson A, Gotzsche PC. Placebo treatment versus no treatment. *Cochrane Database of Systematic Reviews* 2003;(1):CD003974.

102. Lauer JA, Rohrich K, Wirth H, Charette C, Gribble S, Murray CJL. PopMod: a longitudinal population model with two interacting disease states. *Cost-effectiveness and Resource Allocation 2003*; 1(6). Available at: http://www.resource-allocation.com/content/1/1/6.

103. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *British Medical Journal* 2003; 326(7387):472.

104. Mulrow CD. Rationale for systematic reviews. *British Medical Journal* 1994; 309(6954):597–599.

105. Schmid CH, McIntosh M, Cappelleri JC, Lau J, Chalmers TC. Measuring the impact of the control rate in meta-analysis of clinical trials. *Controlled Clinical Trials* 1995; 16:66s.

106. Coe R. *What is an effect size? A brief introduction*. Available at: http://www.cem.dur.ac.uk/ebeuk/research/effectsize/ESbrief.htm. 2001.

107. Acharya A, Murray CJL. Rethinking discounting of health benefits in cost-effectiveness analysis. In: Wikler D, Murray CJL, eds. *Fairness and goodness:*

ethical issues in health resource allocation. Geneva: World Health Organization, (forthcoming); 2003.

108. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Economics* 2001; 10(7):587–599.

109. Cairns J. Discounting in economic evaluation. In: Drummond M, McGuire A, eds. *Economic evaluation in health care. Merging theory with practice*. Oxford: Oxford University Press; 2001. p. 236–255.

110. HM Treasury. *The green book. Appraisal and evaluation in central government. Annex 7 discount rate*. London: HM Treasury; 2003.

111. Sen AK. Approaches to the choice of discount rates for social benefit-cost analysis. In: Lind RC, ed. *Discounting for time and risk in energy policy*. Baltimore: Johns Hopkins University Press; 1982. p. 325–350.

112. Lazaro A, Barberan R, Rubio E. The economic evaluation of health programmes: why discount health consequences more than monetary consequences? *Applied Economics* 2002; 34(3):339–350.

113. Evans DB, Hurley SF. The application of economic evaluation techniques in the health sector: the state of the art. *Journal of International Development* 1995; 7(3):503–524.

114. van Hout BA. Discounting costs and effects: a reconsideration. *Health Economics* 1998; 7(7):581–594.

115. Reinschmidt KF. Aggregate social discount rate derived from individual discount rates. *Management Science* 2002; 48(2):307–312.

116. Poulos C, Whittington D. *Individuals' time preferences for life-saving programs: results from six less developed countries*. Available at: http://idrinfo.idrc.ca/archive/corpdocs/118114/Individuals.htm. 1999.

117. Viscusi WK, Moore MJ. Rates of time preference and valuations of the duration of life. *Journal of Public Economics* 1989; 38:297–317.

118. Acharya A, Murray CJL. Rethinking discounting of health benefits in cost-effectiveness analysis. In: Wikler D, Murray CJL, eds. *fairness and goodness: ethical issues in health resource allocation*. Geneva: World Health Organization, (forthcoming); 2003.

119. Murray CJ, Acharya AK. Understanding DALYs (disability-adjusted life years). *Journal of Health Economics* 1997; 16(6):703–730.

120. Keeler EB, Cretin. Discounting of life saving and other nonmonetary effects. *Management Science* 1983; 29(3):300–306.

121. Parsonage M, Neuburger H. Discounting and health benefits. *Health Economics* 1992; 1(1):71–76.

122. Cropper M, Aydede S, Portney P. Preferences for life saving programs: how the public discounts time and age. *Journal of Risk and Uncertainty* 1994; 8:243–265.

123. Dasgupta P. *Human well-being and the natural environment*. New York: Oxford University Press; 2001.

124. Newell R, Pizer W. *Discounting the distant future: how much do uncertain rates increase valuations?* Washington, D.C: Resources for the Future; 2000. Discussion Paper 00–45.

125. Reinschmidt KF. Aggregate social discount rate derived from individual discount rates. *Management Science* 2002; 48(2):307–312.

126. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmaco-Economics* 2000; 17(5):479–500.

127. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Economics* 1994; 3(2):95–104.

128. King G, Tomz M, Wittenberg J. Making the most of statistical analyses: improving interpretation and presentation. *American Journal of Political Science* 2000; 44:341–355.

129. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics* 2001; 10(8):779–787.

130. Hutubessy RC, Baltussen RM, Evans DB, Barendregt JJ, Murray CJ. Stochastic league tables: communicating cost-effectiveness results to decision-makers. *Health Economics* 2001; 10(5):473–477.

131. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technology Assessment* 1999; 3(2):1–134.

132. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* 1997; 6(4):327–340.

133. McNeil BJ, Dudley RA, Hoop B, Metz C, Thompson M, Adelstein SJ. A cost-effectiveness analysis of screening for hepatitis B surface antigen in India. *Medical Decision Making* 1981; 1(4):345–359.

134. Goodman CA, Mills AJ. The evidence base on the cost-effectiveness of malaria control measures in Africa. *Health Policy and Planning* 1999; 14(4):301–312.

135. Baltussen RM, Hutubessy RC, Evans DB, Murray CJ. Uncertainty in cost-effectiveness analysis. Probabilistic uncertainty analysis and stochastic league tables. *International Journal of Technology Assessment in Health Care* 2002; 18(1):112–119.

136. Murray CJ, Lauer JA, Hutubessy RC, Niessen L, Tomijima N, Rodgers A et al. Effectiveness and costs of interventions to lower systolic blood pressure and cholesterol: a global and regional analysis on reduction of cardiovascular-disease risk. *Lancet* 2003; 361(9359):717–725.

137. Murray CJ, Frenk J. A framework for assessing the performance of health systems. *Bulletin of the World Health Organization* 2000; 78(6):717–731.

138. *Fairness and goodness: ethical issues in health resource allocation.* Wikler D, Murray CJL, eds. Geneva: World Health Organization, (forthcoming); 2003.

139. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 1997; 349(9063):1436–1442.

140. Murray CJL, Salomon JA, Mathers CD. A critical examination of summary measures of population health. In: Murray CJL, Lopez AD, Salomon JA, Mathers CD, eds. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002.

141. Evans DB, Tandon A, Murray CJ, Lauer JA. Comparative efficiency of national health systems: cross national econometric analysis. *British Medical Journal* 2001; 323(7308):307–310.

142. Clewer A, Perkins D. *Economics for Health Care Management*. Harlow, England: Pearson Education Limited; 1998.

143. Atkinson AB. *The economics of inequality*. 2nd ed. New York: Oxford University Press; 1983.

144. The discipline of cost-benefit analysis. Sen A, ed. *Rationality and freedom*. Cambridge, MA: The Belknap Press of Harvard University Press; 2002. pp. 553–577.

145. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR. Assessing the effectiveness of health interventions for cost- effectiveness analysis. Panel on Cost-Effectiveness in Health and Medicine. *Journal of General Internal Medicine* 1997; 12(9):551–558.

146. *Cost-effectiveness in health and medicine*. Gold MR, Siegel JE, Russel LB, Weinstein MC, eds. New York: Oxford University Press; 1996.

147. Lothgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Economics* 2000; 9(7):623–630.

Annexes

Annex A. WHO-CHOICE activities on Generalized Cost-Effectiveness Analysis

1. WHO-CHOICE regional analyses

Individual analysts will usually undertake studies of specific interventions in a particular country. Ideally, WHO-CHOICE would also like to be able to provide evidence of the cost-effectiveness of a wide range of interventions in each of the 192 countries that are WHO Member States since each has a different starting point in terms of training, infrastructure and history of health promotion, a different epidemiological situation, and a different cost structure. Within large countries it would be better still to undertake the analysis at a sub-national level (e.g. provinces of China or states of India). However, it is not practical for WHO-CHOICE to attempt such a country-specific exercise in the short-term (indeed, no country has yet been able to undertake the volume of work necessary to assess a very wide range of interventions in their own settings).

There is consequently a pragmatic need for policy-makers to borrow and adapt results obtained in other settings and to generalize these to their own settings. Global estimates, however, have limited credibility among policy-makers in individual countries because of the diversity of cost structures, epidemiological profiles and starting conditions. It is therefore necessary to compromise between specificity and the practicality of undertaking the necessary work. To facilitate this process, WHO-CHOICE reports costs and effects of a large number of interventions by 14 epidemiological subregions, with the regions being derived with reference to geographical location, epidemiological status and mortality stratum, which are relatively homogenous (details of the subregions chosen for the initial analysis are found in Annex G). Analysts

should still undertake their own studies in particular settings. However, analysts are strongly urged to report their results in a way that allows WHO-CHOICE to add them to the regional database, and which allows analysts in other countries to adapt them to their own settings.

2. Reporting WHO-CHOICE results

WHO-CHOICE reports results of the cost-effectiveness of clusters of mutually exclusive interventions for 14 epidemiological subregions at www.who.int/evidence/cea. WHO-CHOICE requires all interventions to be evaluated in a consistent and comparable manner following the methods and recommendations presented in these guidelines. The following is a partial list of the information available on the Web:

- the interventions studied and their definitions;
- background papers on the methods and assumptions used in the GCEA analysis;
- detailed region-specific lists of the variables used in the analysis, including: quantities and prices of factor inputs, useful life of capital items, hospital and health centre unit costs, PPP exchange rates and price multipliers for different coverage levels, detailed documentation of the epidemiological models and the assumptions made, and disaggregated analysis spreadsheets of costs and effectiveness results; and
- tools used in the WHO-CHOICE analysis and their user manuals, available for download.

3. Contextualizing WHO-CHOICE results

The WHO-CHOICE database provides estimates of the cost-effectiveness of interventions in the 14 epidemiological subregions under study, and is useful for decision-makers to distinguish between "good buys" and "bad buys" in health. As indicated in previous sections, results should not be used in a formulaic approach, but should be analysed to identify order of magnitude differences in cost-effectiveness of different interventions.

Rather than using results that are relevant at the regional level, decision-makers may wish to adapt this information to their local decision-making context. The ingredient approach as utilized in WHO-CHOICE allows analysts to contextualize costs and effects. As a first step in this process, analysts should asses the results of sensitivity/uncertainty analysis to determine which variables have the largest impact on the resulting cost-effectiveness ratios. As a second step, to the extent possible regional values for those variables should be substituted by values that are more relevant for the context under study. This may involve a number of issues for both costs and effects.

Costs of interventions are broken down in quantities and prices, which means that analysts should collect relevant quantities and unit costs for the intervention(s) in their setting. WHO-CHOICE reports estimates on

costs on the basis of technical efficiency, i.e. capacity utilization is assumed to be 80% and prices are assumed to be the lowest achievable. This ensures that the observed differences in cost-effectiveness of interventions are due to the intrinsic characteristics of the intervention rather than the extent to which capital and labour have been utilized in the environment in which the interventions were evaluated, and the extent to which negotiations on prices of goods (especially drugs) have resulted in low prices. Analysts may wish to use alternative assumptions in their estimates, but should do so consistently to maintain comparability of estimates. Furthermore, to estimate non-linearities in costs when scaling up interventions, analysts should review the WHO-CHOICE assumptions that estimate the increased marginal costs of reaching more remote areas, including the definition of catchment areas of health centres, and adapt this to their context *(61)*.

The effects of interventions at the population level have been estimated using the population model PopMod, based on a set of regional demographic and epidemiological parameters, and assumptions on the (clinical) effectiveness of interventions. Intervention effects can be adapted to the local context by (i) scaling down the results to the local level, i.e. by analysing the relative population size of the target population, and (ii) by using a local set of demographic and epidemiological parameters, and effectiveness assumptions (including those on non-compliance for example). In the second option, population models need to be rerun to estimate local effectiveness at the population level.

WHO-CHOICE is currently undertaking a number of country studies to adapt the regional cost-effectiveness results to the country level. This will provide more detailed information on the process involved and information required.

ANNEX B. DRAFT LIST OF INTERVENTION CLUSTERS FOR EVALUATION BY WHO-CHOICE*

A. Respiratory infections

B. Diarrhoeal diseases

C. Malnutrition

D. Vaccine-preventable diseases

E. Antenatal/perinatal care and other reproductive health services

F. Musculoskeletal diseases

G. Cancers

H. Cardiovascular disease including stroke

I. Diabetes mellitus

J. Neuro-psychiatric disorders

K. HIV/AIDS/STD/TB and other infectious diseases

L. Motor vehicle accident and other injuries

M. Malaria and other tropical diseases

N. Blindness and other sense organ disturbances

O. Genetic and other disorders with a hereditary basis

P. Health system interventions

  1. Facilities

  2. Pharmaceuticals

 * 1. *The World Health Report 2002* focused on identifying risk factors to health; this cuts across intervention clusters.

   2. Criteria used in prioritizing interventions for analysis include: those that are not widely used and are probably cost-effective, those that are widely used and probably cost-ineffective, and those that are widely used and probably cost-effective (as benchmarks).

ANNEX C. AN ILLUSTRATION OF THE TYPES OF COSTS INCLUDED IN A SELECTION OF INTERVENTION ACTIVITIES AT CENTRAL LEVELS

Name of the activity	Description
Administration	Includes overhead costs (e.g. space, furniture, equipment, utilities, maintenance, etc.) used by the programme and personnel in preparation for meetings or training and other administration activities.
Planning	Includes costs associated with planning. For example, per diem allowances for meetings for the endorsement and implementation of the intervention as well as other meeting costs, such as venue, supplies, transportation, etc.). It also includes payments to consultants who participated in the planning phase of the intervention.
Training	Includes special training to develop health workers' skills to deliver the intervention. It should not include under-graduate or post-graduate training as well as that occurring during the residency period. Only training that had to be provided to deliver the intervention should be included, i.e. it should be specific to the intervention studied. This will depend largely on the extent of training facilities that are available in each country. For example, in countries where most radiologists are already trained to perform this service, to add an intervention that does not require them to learn more skills but just to modify the optimum doses of radiotherapy given to cancer patients will not require prior training. In other countries where this initial experience was not available, training of providers should be included. Other examples include training of health workers to administer a new vaccine or to use a new guideline for case management. This will include costs of adaptation of the guidelines and training materials as well as the translation of the training materials to the national language if required. This should not include costs incurred at an international level, e.g. where international organizations develop guidelines for international consumption such as the development of the WHO guidelines for case management of acute lower respiratory infections.
Media and IEC	Development and production of information, education and communication (IEC) materials. This includes costs of developing the IEC materials in terms of designing the message, testing, revision and re-testing. It also includes the cost of printing those materials and/or the radio or television time to air them.
Monitoring and supervision	This includes supervision visits to health facilities in terms of per diem allowances, travel allowances and personnel salaries, if the latter is not already included in one of the activities listed above (e.g. in administration activities).
Social mobilization	This includes motivating and educating the public, and marketing health-related interventions through local markets. For example, this might involve retailers receiving some guidance on the correct use of items such as insecticide-impregnated bednets for malaria prevention. All advertising and promotion activities, seminars, technical support to retailers should be included.

ANNEX D. INTERPRETING INTERNATIONAL DOLLARS

WHO-CHOICE reports on costs and effects of interventions at the subregional level. To adequately represent subregional costs, the use of one national currency, such as US dollars, would be inadequate: the costs of non-traded goods are likely to differ between countries within a certain subregion, and it would not be possible to assign a single US$ value that would properly represent the costs of non-traded goods in all countries in the subregion. For example, assuming that an intervention consists only of the salary of a nurse, the US$ cost equivalent of the intervention would differ considerably between countries in a particular subregion.

To adequately summarize the costs of interventions in a common currency, WHO-CHOICE reports its cost estimates in terms of international dollars (I$). The basic concept of I$ is that it represents the same value in every country, i.e. the purchasing power of 1 I$ is similar around the world. In other words: one can buy the same things in any country with the same amount of I$.

How does this work in practice? To interpret the cost-effectiveness results of WHO-CHOICE, I$ costs should be converted into local currency units. A distinction needs to be made between non-traded and traded goods. To convert costs of *non-traded* goods, the I$ amount needs to be multiplied by the purchasing power parity (PPP) of a country. For example, in 1999, the costs of a bus ticket in the AFRO-D Region may be valued at 5 I$. Given that the PPP in Benin in 1999 equalled 302, its value in local currency unit (CFA in this case) would be 5 x 302 = CFA 1510. Since PPP values differ across countries within regions, the regional costs of a non-traded good may be similar in terms of I$ but not in terms of the local currency units.

To convert costs of *traded* goods (expressed in I$) into local currency units, one only needs to divide by the official exchange rate (OER) because prices of traded goods are similar across countries. For example, if a drug costs I$ 0.28, and the OER is 0.0014, the CFA equivalent would be 200.

Estimating I$ involves the reverse process. Costs of non-traded goods (as expressed in local currency units) should be divided by the PPP to obtain the I$ equivalent, while cost of traded goods (as expressed in local currency units) should be multiplied by the official exchange rate.

ANNEX E. DALYs TO MEASURE BURDEN OF DISEASE

DALYs are the sum of years of life lost (YLLs) and years of life lived with disability (YLDs). A variety of measures have been developed to measure the stream of life lost due to death at different ages. These measures can be divided into four families: potential years of life lost, period expected years of life lost, cohort expected years of life lost and standard expected years of life lost *(139)*.

- **Potential years of life lost (PYLL)** is the simplest measure of time lost due to premature death. A potential limit to life is chosen arbitrarily and the duration of life lost due to a death is simply the potential limit to life minus the age at death. PYLL are criticized because deaths averted for people older than the arbitrarily chosen potential limit of life do not contribute to the burden of premature mortality. Using it as an indicator for CEA implies that there is no benefit to health interventions that reduce mortality over the potential limit to life. This is at odds with the values of most societies.
- **Period expected years of life lost.** A popular alternative to PYLL is to calculate period expected years of life lost (PEYLL), where the duration of life lost is the local period life expectancy at each age. In a period life table, life expectancy at each age is the estimated duration of life expected at each age if the current age-specific mortality patterns were to hold in the future. In PEYLL, a population's current mortality level is being used as the "ideal" against which it is compared in order to calculate the burden of disease. Over time and across communities, local life expectancies vary and thus the reference standards vary, creating at times, peculiar findings for burden comparisons.
- **Cohort expected years of life lost.** Given past secular trends in mortality, the average individual alive today at any given age is likely to live substantially longer than period life expectancy at that age. As distinct from period life expectancy, cohort life expectancy is the estimated average duration of life a cohort would actually experience. Cohort life expectancy is substantially higher than period life expectancy. However, a disadvantage is that if expected years of life lost are used as a measure of the burden of disease, a death in a rich country where life expectancy at each age is higher would be considered a greater burden than a death in a poor country with a lower life expectancy. If burden of disease assessments were to influence resource allocation this could lead to counter-intuitive and inequitable conclusions.
- **Standard expected years of life lost.** The advantages of an expectation approach where every death contributes to the burden of disease, and the equitable approach of PYLL where every death of a given age contributes equally to the calculation of the burden of disease, can be combined by using a standard expectation of life at each age as the

reference norm. For measuring the global burden of disease due to premature mortality, the SEYLL method has been adopted. To define the standard, the highest national life expectancy observed was taken. Based on the observation that Japanese females achieve a period life expectancy at birth higher than 82 years, the standard expectations were based on model life table which has a life expectancy at birth for females of 82.5 years. Note that this is not the approach used to measure DALYs averted by interventions which requires a different calculus. Details are found in Section 4.

ANNEX F. MEASURING INTERVENTION BENEFIT AT THE
               POPULATION LEVEL

In Section 4.1.7, it was claimed that a population model is often
necessary to measure intervention benefit accurately. It is therefore
necessary to relate the kind of measure deriving from such a model (see
Section 4.2) to the other standard measures of benefit that satisfy the
general criteria established in the foregoing parts of Section 4.1.

   For example, under appropriate assumptions, changes in healthy years
lived (HYL) are equal to changes in DALYs. To see this, consider Figure
F.1. Area A (dark grey) represents the population number surviving in
equivalent full health under the baseline scenario, i.e. where there is no
intervention. This area is analogous to the area under the lower line in
Figure 4.1 (Section 4.1.7), except that here non-fatal health effects are
also considered: "equivalent full health" means that the survivorship
curve forming the upper boundary of Area A has been adjusted for time
spent in states less than perfect health (see Section 4.1.2). While a
standard survivorship curve typically represents the percentage surviving
at a given *age*, here the absolute number surviving at a given *time* is
shown. For simplicity, only the population alive at time $t = 0$ is depicted,
i.e. there are no births or other entrances.

   Now suppose that an intervention is introduced and that Area B (light
grey) represents the increment experienced by the population when that
intervention is implemented at time $t = 0$. It is clear that this area
represents the intervention effect, or intervention benefit, measured at
population level. On the stated assumptions, Area B in the above diagram

Figure F.1   The area below the health-adjusted survival curve (A),
                   the benefit (B) resulting from a given intervention, and the
                   loss (C) relative to a normative goal
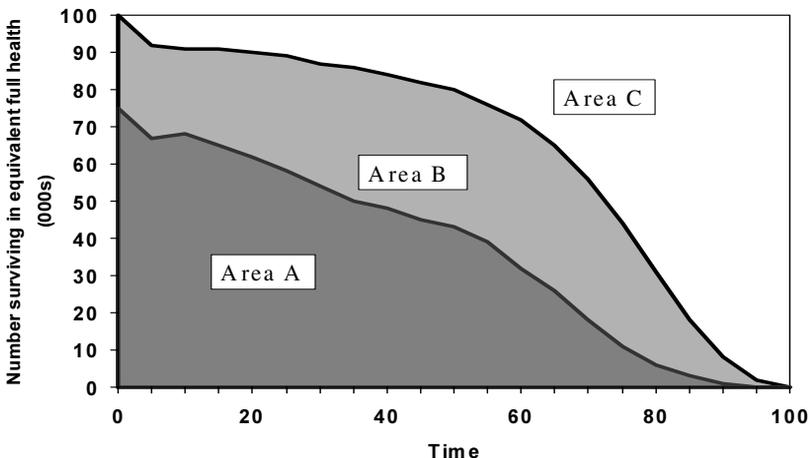
Figure F.2   The area below the health-adjusted survival curve (A)
                  and the loss (B + C) relative to a normative goal when no
                  intervention is implemented
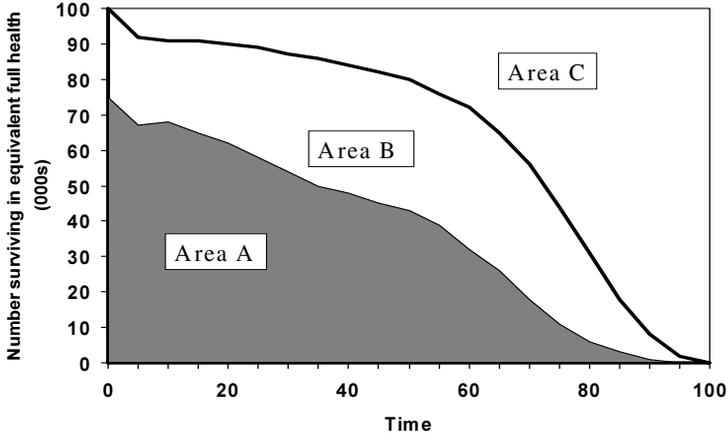
is denominated in units of HYL. The area under the survivorship curve
after the intervention consists of the sum of Area A and Area B. Area B
can be measured as the difference of the integrals of the two survivorship
curves *(100)*. Both Area A and Area B belong to the type of measure that
demographers denote as belonging to the "health expectancy" family
(140).

   Now assume that in Figure F.1 the population members alive at a given
time have different ages, i.e. they do not all belong to the same birth
cohort, and further assume that the survivorship curves shown in Figure
F.1 are adjusted for the average societal values attached to life lived at
different ages, using a system of age-specific weights (see Section 4.1.6).
Finally, assume that time discounting is also represented in the figure at a
constant rate of 3%. On these assumptions, Area B is denominated in
age-weighted, discounted HYL.

   Area C (white) represents a loss in population health, where "loss" is
measured relative to a particular reference standard. Here the reference
standard is the vertical line drawn at time *t* = 100. A vertical line is used
for purposes of illustration, although another reference standard could
be used, such as the age-specific life expectancy of a particular
population, which is the approach is used for calculating the YLL
component of DALYs (see Section 4.1.7 and Annex E). Note that "loss"
is by definition measured above the survivorship curve, and is therefore
is a "health gap" measure *(2)*. Although the reference standard in Annex
Figure F.1 is only a vertical line representing death at an arbitrary point
in time, Area C can be thought of as measuring a particular kind of
DALY (i.e. one in which the reference standard is simpler than an
idealized survivorship curve).

In any case, once a reference standard is chosen, Area C is fully determined by the curve forming the upper boundary of Area B. Area C represents DALYs in a population in which the intervention was implemented at time t = 0; if the intervention is not implemented, DALYs are measured by the sum of Area B and Area C, as shown in Figure F.2.

It is therefore evident that Area B represents the same quantity of intervention benefit, whether benefit is measured in terms of DALYs or HYLs. This is because intervention benefit is not measured "above" or "below" the curve but is measured as the difference between two survivorship curves. For the equivalence to hold exactly, it is only necessary to ensure that changes in DALYs/HYLs are calculated using the same assumptions, namely, with the same discount rate and the same set of age weights and health state valuations.

ANNEX G. EPIDEMIOLOGICAL SUBREGIONS AS APPLIED IN WHO GENERALIZED CEA

WHO region	Mortality stratum	WHO Member States
AFRO	D	Algeria, Angola, Benin, Burkina Faso, Cameroon, Cape Verde, Chad, Comoros, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Niger, Nigeria, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Togo
AFRO	E	Botswana, Burundi, Central African Republic, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Eritrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Uganda, United Republic of Tanzania, Zambia, Zimbabwe
AMRO	A	Canada, United States Of America, Cuba
AMRO	B	Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Dominica, Dominican Republic, El Salvador, Grenada, Guyana, Honduras, Jamaica, Mexico, Panama, Paraguay, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela
AMRO	D	Bolivia, Ecuador, Guatemala, Haiti, Nicaragua, Peru
EMRO	B	Bahrain, Cyprus, Iran (Islamic Republic of), Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates
EMRO	D	Afghanistan, Djibouti, Egypt, Iraq, Morocco, Pakistan, Somalia, Sudan, Yemen
EURO	A	Andorra, Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Sweden, Switzerland, United Kingdom
EURO	B	Albania, Armenia, Azerbaijan, Bosnia And Herzegovina, Bulgaria, Georgia, Kyrgyzstan, Poland, Romania, Slovakia, Tajikistan, The Former Yugoslav Republic of Macedonia, Turkey, Turkmenistan, Uzbekistan, Yugoslavia
EURO	C	Belarus, Estonia, Hungary, Kazakhstan, Latvia, Lithuania, Republic of Moldova, Russian Federation, Ukraine
SEARO	B	Indonesia, Sri Lanka, Thailand
SEARO	D	Bangladesh, Bhutan, Democratic People's Republic of Korea, India, Maldives, Myanmar, Nepal
WPRO	A	Australia, Japan, Brunei Darussalam, New Zealand, Singapore
WPRO	B	Cambodia, China, Lao People's Democratic Republic, Malaysia, Mongolia, Philippines, Republic of Korea, Viet Nam
		Cook Islands, Fiji, Kiribati, Marshall Islands, Micronesia (Federated States of), Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu

ENDNOTES

1 There is an important distinction to make between allocative and technical efficiency *(141)*. Allocative efficiency refers to the optimal choice of input mix, given their respective prices. If interventions are regarded as inputs, it is defined in terms of whether the health system provides the most cost-effective set of interventions for the given level of expenditure. Technical efficiency is defined as the ability to produce a given output at the lowest possible cost.

2 For example, CEA may be used by pharmaceutical firms to set price levels for pharmaceuticals.

3 Interventions are said to be independent if choosing one does not prevent the choice of any other intervention. Interventions are said to be mutually exclusive if only one alternative can be selected. Mutually exclusive interventions have also been called "competing" or "incompatible" interventions. This situation frequently occurs in health care, e.g. in screening or vaccination programmes that have various alternatives according to intensity or age limit.

4 This is often called the average CER.

5 An intervention is said to be weakly dominated by other interventions if a combination of these other interventions has a more favourable cost-effectiveness ratio. Weakly dominated interventions can be identified by calculating the incremental cost-effectiveness ratios for each successively more costly intervention: if any of these incremental ratios turns out to be less than the previous one in the sequence of increasingly costly mutually exclusive interventions, then the previous one is ruled out by weak dominance.

6 Lack of information is one example of the "market failures" that typify the health sector. Others include the public goods nature of some health interventions and the existence of externalities. Taken together, these mean that the traditional way of valuing the benefits of an intervention in economics, using willingness to pay, is not appropriate. More information on market failures in health can be found in microeconomics textbooks such as Clewer and Perkins *(142)*.

7 This analysis assumes no uncertainty. The impact of uncertainty on this expansion path is considered in Section 6.

8   Ideally it would be better to estimate full cost functions for all levels of coverage, joint cost functions for all feasible levels of coverage of interventions provided jointly, and effectiveness functions for expanding coverage or joint production. In the short to medium term this is not feasible so we focus here on a parsimonious set of combinations.

9   Yet another option would be to tailor the time horizon to the specific characteristics of interventions, i.e. to represent the time required until the intervention reaches its full effectiveness. However, as discount rates for costs and health effects may differ, the definition of the implementation period—which is somewhat arbitrary—is likely to affect the cost-effectiveness of the intervention.

10 Inequality aversion reflects people's dislike of differences between the better-off and the worse-off people in society and their preference for a redistribution of resources to reduce inequality *(143)*.

11 Willingness to pay methods of valuing these changes such as contingent valuation are also problematic partly because of the market failures argument described in Section 2.1 *(144)*.

12 Informal care-giver time is time spent caring for a patient by non-professionals such as family, friends, acquaintances, or neighbours for which they are not financially compensated

13 A Cobb-Douglas production function takes the form of:

$$Y = AK^{\alpha}L^{1-\alpha}$$

where α is the output elasticity of capital, K is capital, L is labour, A is a technology parameter, and constant returns to scale are assumed. From the Cobb-Douglas production function, the following relationship can be derived:

$$\Delta Y/\Delta L = (1-\alpha) \, Y/L$$

This implies that, on the aggregate, the marginal product of labour equals $(1-\alpha)$ times the average product of labour. Hence, the change in output from a unit change in labour input equals $(1-\alpha)$ times the output per worker in a given country. So only if $\alpha = 0$ will the GDP per person equal the marginal product of labour. If the shadow price of labour is assumed to be Y/L, it will seriously overstate the true marginal cost. If it is assumed to be Y divided by population, it will be only by chance that it equals the marginal product of labour.

14 An infinite number of combinations could be defined by varying x. It is necessary to be pragmatic in practice and define a parsimonious set of combinations by identifying critical levels of coverage at which the slope of the expansion path is likely to change. See Murray et al. *(23)* for further information.

15 Another example is that some international prices may include transfer payments, such as export taxes or subsidies.

16 For example, the data source "World Development Indicators" (from the World Bank) uses a Laspeyres index formula.

17 In some countries health-specific GDP deflators and CPI are available. But because this is not true for most of the countries of the world, we are not recommending that they be used for GCEA.

18 Remember that the health state valuations used in the GBD study are for calculating the burden of disease, so their complements (1-health state valuation) should be used for CEA.

19 Alternatively, when no population model is used this can be considered as the difference in DALYs averted in the intervention scenario and in the null scenario.

20 Population models such as PopMod make use of instantaneous transition rates or hazard rates to express rates of incidence or remission, for example. The use of person time in the denominator is a more accurate representation of transitions between health states than the use of proportions (persons at risk in the denominator), often applied in decision-tree analysis (145).

21 The original version of DisMod (DisMod v1.0. President and Fellows of Harvard College. All rights reserved, 1994) is available from the WHO web site at http://www.who.int/whosis, under Burden of Disease activities. Installation instructions are also provided on the web site. Also available is DisMod II, a new software system developed to provide a full graphical interface, database storage capabilities and substantially enhanced features and options.

22 In the basic 5-box PopMod, additional states can be added by doing weighted averages of the hazard rates or health state valuations of two or more (similar) states within a single box.

23 A range of assumptions can be made for co-morbidity, ranging from independence of the conditions of interest to co-occurrence beyond that expected purely by chance. The analyst starts with what is known about the disease and its associated risk factors (e.g. diabetes and cardiovascular disease) and PopMod provides guidance on how to derive these rates depending on the data that the analyst has available.

24 A more general "continuous-time" version of Equation 5.1 can be expressed as:

$$C_{presentvalue} = \int_0^T \Psi(t) e^{-rt} dt$$

where the integration is bounded by the beginning period 0 and the end period T (146).

25 The Health Department is now considering moving to a 3.5% discount rate for costs.

26 Health state valuations are sometimes considered to be social choice variables and sometimes measurable variables with a probability distribution. CHOICE applies probabalistic uncertainty analysis to account for uncertainty in health state valuations rather than sensitivity analysis.

27 A 90% uncertainty interval is used for illustrative purposes. Any level of significance can be chosen at the analysts discretion. In WHO-CHOICE 90% limits are used.

28 Usually it is not realistic to assume that costs and effects have zero covariance. A high cost drug, for example, might be more effective than a low cost drug. If adherence to medicines is lower than expected, benefits will be lower and so

will costs because less of the medicine is consumed. For the ease of exposition here, however, no covariance structure is assumed.

[29] It would be possible for the price to fall with increasing coverage due to the ability to negotiate a lower price, but that is a different point.

[30] In the current hypothetical example, correlations in costs and effects are dealt with by estimating the correlation between total costs (and benefits) of one intervention with the next. The ideal way of dealing with these interactions, however, is by linking the uncertainty for variables that that are common to those interventions, such as unit price or quantity. The decision model described subsequently (MCLeague) allows for this more sophisticated way of dealing with correlations between the costs of different interventions.

[31] Another reason why it is difficult to interpret ICERs and their uncertainty intervals derived using bootstrapping methods is that the ICER can be negative in some draws, because either the incremental costs or the incremental effectiveness of an intervention is negative. Indeed, a positive ICER can occur where incremental costs and effects are both negative and this has very different policy implications from a positive ICER where incremental costs and effects are both positive . This is one reason why CE acceptability curves has been developed to aid decision-making *(147)*. The decision model (MCLeague) that is described below avoids this problem so acceptability curves are not discussed further.

[32] Analysis of the paired simulation data shows that b1-ab1 is more cost-effective than b1–b2 83% of the time. Inspection of the uncertainty intervals alone would suggest, however, that we are less than 83% confident that b1–ab1 is more cost-effective.

[33] There are several other examples of this problem in the current example, e.g. in order to move from a starting budget of $400 000 (optimal mix B2) to $700 000 (optimal mix ab1), funding would need to be cut for BCG to b1 in order to purchase a, and thereby move to ab1.

[34] The cost penalty of shifting resources may outweigh the efficiency gains or may no longer make the shift between b1-a affordable.

[35] This is a preliminary presentation, using the available data on the sets of interventions which have already been analysed. With analysis of other interventions, e.g. immunization, the frontier would probably move upward.

[36] Ideally, independent analysts should be able to rerun the analysis including the ingredients that are relevant to their context. However, this is often not possible because modelling tools are typically unavailable and/or their use may be complex. By providing its modelling tools with manuals, WHO-CHOICE tries to overcome this problem.

Part Two: Background papers and applications

1   Development of WHO guidelines on generalized cost-effectiveness analysis[1]

Christopher J.L. Murray, David B. Evans, Arnab Acharya and Rob M.P.M. Baltussen

Summary

The growing use of cost-effectiveness analysis (CEA) to evaluate specific interventions is dominated by studies of prospective new interventions compared with current practice. This type of analysis does not explicitly take a sectoral perspective in which the costs and effectiveness of all possible interventions are compared, in order to select the mix that maximizes health for a given set of resource constraints.

WHO guidelines on generalized CEA propose the application of CEA to a wide range of interventions to provide general information on the relative costs and health benefits of different interventions in the absence of various highly local decision constraints. This general approach will contribute to judgements on whether interventions are highly cost-effective, highly cost-ineffective, or something in between. Generalized CEAs require the evaluation of a set of interventions with respect to the counterfactual of the null set of the related interventions, i.e. the natural history of disease.

Such general perceptions of relative cost-effectiveness, which do not pertain to any specific decision-maker, can be a useful reference point for evaluating the directions for enhancing allocative efficiency in a variety of settings. The proposed

framework allows the identification of current allocative inefficiencies as well as opportunities presented by new interventions.

Key words: cost-effectiveness analysis, guidelines, resource allocation

Introduction

The growing use of cost-effectiveness analysis (CEA) to evaluate the efficiency of specific interventions is dominated by studies of prospective new interventions compared with current practice *(1–11)*. This type of analysis does not explicitly take a sectoral perspective in which the costs and effectiveness of all possible interventions are compared, in order to select the mix that maximizes health for a given set of resource constraints. The estimated cost-effectiveness of a single proposed new intervention is compared either with the cost-effectiveness of a set of existing interventions derived from the literature *(12–17)* or with a fixed price cut-off point representing the assumed social willingness to pay for an additional unit of benefit *(18–21)*. The implicit assumption that, to improve overall efficiency, resources would need to be transferred to the more efficient intervention either from another health intervention or from another sector, is rarely discussed.

On the other hand, much of the theoretical literature has taken a broader view of cost-effectiveness, exploring its use in allocating a fixed health budget between interventions in such a way as to maximize health in a society *(22–34)*. This we call sectoral CEA. Only a few applications of this broader use—in which a wide range of preventive, curative and rehabilitative interventions that benefit different groups within a population are compared in order to derive implications for the optimal mix of interventions—can be found. Examples include the work of the Oregon Health Services Commission *(35–40)*, the World Bank Health Sector Priorities Review (41) and the Harvard Life Saving Project *(42,43)*. Of these, only the World Bank attempted to make international or global comparisons of sectoral cost-effectiveness.

At the heart of this broadened policy use is the notion that resources in the health sector should be allocated across interventions and population groups to generate the highest possible overall level of population health. If the calculations show that some current interventions are relatively cost-ineffective, and that some which are not undertaken fully are relatively cost-effective, resources could be reallocated across interventions to improve population health. In other words, the allocative efficiency of the health sector could be enhanced by moving resources from cost-ineffective interventions to cost-effective

ones.[a] Interest in the promise of enhancing allocative efficiency of health systems has led to analytical efforts to study the cost-effectiveness of a broad range of interventions in a number of countries *(44,45)*.

Several challenges have emerged to this wider use of CEA. First, analysts and decision-makers have correctly noted that resource allocation decisions affecting the entire health sector must also take into account social concerns, such as a priority for the sick *(46–49)*, reducing social inequalities in health *(50–53)*, or the well-being of future generations *(54,55)*. Vociferous debate on the use of CEA to prioritize the use of Medicaid resources in Oregon State is one indication of these concerns in the political arena *(35–40)*. So far there have been two proposed responses to this challenge: abandon the practice of using CEA to inform resource allocation decisions entirely or to progressively incorporate more of these social concerns into the methods of CEA *(56)*.

Second, current CEA practice *(57,58)* often fails to identify existing misallocation of resources by focusing on the evaluation of new technologies or strategies. The very wide range of cost-effectiveness ratios found in the compendia of CEAs listed above suggest that addressing current allocative inefficiencies in many countries may yield substantial health gains, possibly more than identifying new technologies that will make small improvements in health.

Third, for all but the richest societies, the cost and time required to evaluate the large set of interventions required to use CEA to identify opportunities to enhance allocative efficiency may be prohibitive. The results of many, if not most, CEA studies are so context-specific that they cannot be used to inform policy debate in another population—as reflected in the debate about the use of league tables, which include the results of studies using a variety of methods and which were undertaken to answer a variety of context-specific questions *(12,14–17,59–68)*. For low- and middle-income countries and smaller high-income countries, there has been little progress towards the goal of affordable and timely information on the costs and effects of a wide array of interventions to inform policy.

Fourth, the difficulties of generalizing context-specific CEA studies have been institutionalized by the proliferation of multiple national or subnational guidelines for CEA practice, all using slightly different methods *(69–91)*. International guidelines have not to date been developed. As part of the reorganization of the World Health Organization (WHO) following the election of Dr Gro Harlem Brundtland as the Director-General in May 1998, a new programme, Choosing Interventions: Effectiveness Quality, Costs, Gender and Ethics, part of the Global Programme on Evidence for Health Policy, has been established. This group is attempting to address some of the challenges of providing decision-makers with timely information on the technical and ethical characteristics of different interventions to inform health policy debates. It is collaborating with other international organizations to develop

international guidelines for CEA intended in part to address some of the challenges listed here. In this paper, we outline some of the uses of CEA, the limitations of current methods, directions for revising these methods and some of the remaining technical challenges facing this revision.

Two sectoral uses of cea

The appropriate methods, transferability of results and policy applicability of CEA depend critically on the intended use. CEA can have many applications beyond informing health sector resource allocation decisions across interventions, however, the focus of this paper is on two potential applications. They will be outlined briefly, after which the strengths and weaknesses of current methods of undertaking CEA will be discussed in relation to the two uses.

First, *CEA of a wide range of interventions can be undertaken to inform a specific decision-maker. This person faces a known budget, a set of options for using the budget, and a series of other (resource, ethical or political) constraints.* The set of constraints in this highly context-specific use of CEA for sectoral decision-making will vary tremendously from setting to setting. A decision-maker may be able to reallocate an entire budget or only allocate a budget increase; the decision-maker might be a donor, a minister of health, a district medical officer, or a hospital director. The choices available, at least in the short- to medium-term, might be limited by factors such as the currently available physical infrastructure, human resources or political considerations—for example, in systems with substantial public provision there is a relatively fixed stock of hospital beds that cannot be increased or decreased easily. Decisions could also be constrained by the current mix of interventions that are delivered; perhaps for political reasons specific interventions may not be reduced or eliminated without providing some alternative for that class of health problem. The set of constraints facing a decision-maker defines the decision space or the set of possible options from which choices can be made *(92)*.

Second, *CEA of a wide range of interventions can be undertaken to provide general information on the relative costs and health benefits of different technologies or strategies that are meant to contribute through multiple channels to a more informed debate on resource allocation priorities.* Such general information should be seen as only one input into the policy debate on priorities. Because it is not meant to provide a formulaic solution to resource allocation problems, it need not be highly contextualized. This general approach will contribute to judgements on whether interventions are highly cost-effective, highly cost-ineffective, or something in between. Such general perceptions of relative cost-effectiveness can have far-reaching and constructive influence on policy formulation, defining the set of options that are debated without defining

the allocation of resources in a precise or mechanical fashion. An alternative way to conceptualize this more general use of sectoral CEA is that the results define the mix of interventions that would be health maximizing in the absence of any constraints on possible decisions, except a finite budget. That health maximizing mix of interventions, which does not pertain to any specific decision-maker, can be a useful reference point for evaluating the directions for enhancing allocative efficiency in a variety of settings.

Although all CEA runs the risk of being used in a formulaic way, we believe that the first use of sectoral CEA—to inform a given decision-maker in a specific context—is more likely than the second to be used in this way to determine resource allocation. In context-specific CEA, the challenges of incorporating explicitly other social concerns are more pressing, but efforts to incorporate legitimate context-specific social concerns into the calculation of cost-effectiveness through devices such as equity weights inevitably make the results more difficult to communicate to some decision-makers and to the public. Such efforts also decrease the transferability of results. At some point in the continuum of complexity, the goal of informing a given decision-maker in a specific context may become impossible because of the cost and time required to generate the information *(18)*. We believe that the more general use of CEA, to inform sectoral debates on resource allocation, is where CEA can make the greatest contribution to health policy formulation. Such analysis indicates the general directions for resource reallocation required to enhance allocative efficiency. The results can be weighed alongside other social goals and considered together with the other constraints on decision-makers, which are inevitable in specific contexts. The more generalized approach will enhance transferability and will make it possible to provide useful, timely and affordable information on the health generating characteristics of interventions.[b] In some sense, there is a trade-off between making CEA information precisely relevant to a given context and the time and resources required for that contextualization. Our preference for the more general use of CEA is an indication of how we see the outcome of that trade-off.

Intervention Mix Constrained Cost-Effectiveness

Various attempts have been made to codify a standard practice for CEA *(14,57,58,93–125)*. These guidelines differ for certain technical assumptions, such as standard discount rates, the treatment of unrelated medical costs or the valuation of health outcomes. The broad approach, however, is similar. Intervention costs and health benefits are evaluated with respect to current practice, so that the numerator in the cost-effectiveness ratio is the change in cost due to the application of an intervention compared with the change in health benefit. For the

development of league tables, decision rules have been developed for both independent and mutually exclusive interventions to be ranked in a single league table *(22,28)*. When applied to a wide range of interventions in a population, the results inform decision-makers faced with a single constraint, the budget. The results of this type of analysis do not lead to recommendations to change the current mix of interventions unless the new intervention is accepted over current practice. For this reason, we will refer to this standard practice as intervention mix constrained CEA or IMC-CEA. Interestingly, IMC-CEA as currently practised does not consider other possible constraints on decision making. It is worth noting that the policy environment in which decision-makers come closest to facing a constraint to continue current practice (or expand benefits in areas where there are existing interventions) but face no physical infrastructure, human capital or other constraints, is the United States, where most provision of interventions is in the private sector and ethical guidelines on standards of care tend to automatically adopt all health enhancing interventions.

Figure 1   Costs and benefits of six mutually exclusive interventions
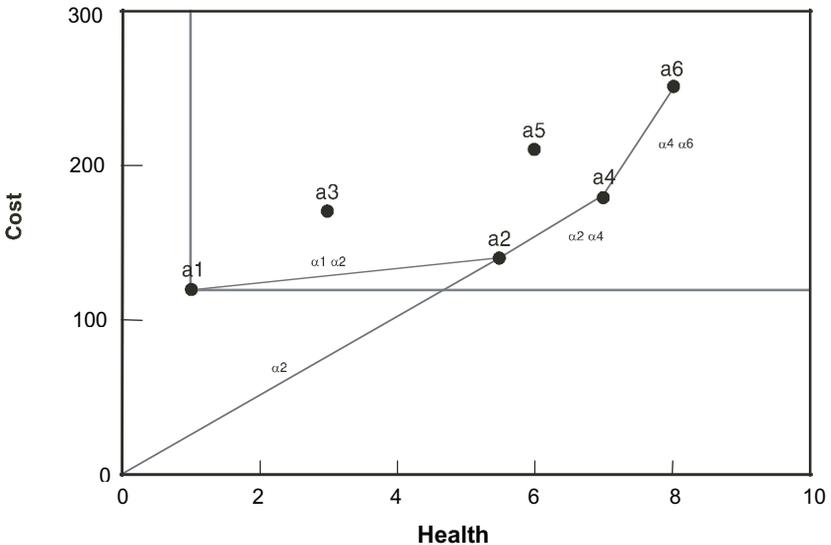
To further explicate the advantages and disadvantages of standard cost-effectiveness methods, consider Figure 1, which depicts the costs and benefits of six mutually exclusive interventions. Following standard practice *(58)*, intervention costs are on the *y*-axis and health benefits on the *x*-axis. In this and subsequent diagrams, each intervention should be thought of as a national programme or policy, which can be purchased at only the point on the figure shown.[c] If a population has purchased intervention a1, then IMC-CEA would evaluate the cost-effectiveness of interventions a2–a6 with respect to the origin set equal to a1—indicated

by the light grey axes. Average cost-effectiveness for each intervention is equal to the slope of the line joining the point to the currently delivered intervention a1, illustrated for intervention a2—this slope is labelled as α1α2. Incremental cost-effectiveness for moving from a2 to a4 is shown as the slope α2α4. For reasons that will be discussed in detail below, the origin in Figure 1 has been set as the costs and health benefits in the absence of any of the interventions a1–a6. The line joining intervention a2 to the origin is the average cost-effectiveness with respect to the null

Figure 2   Costs and benefits of three sets of mutually exclusive interventions
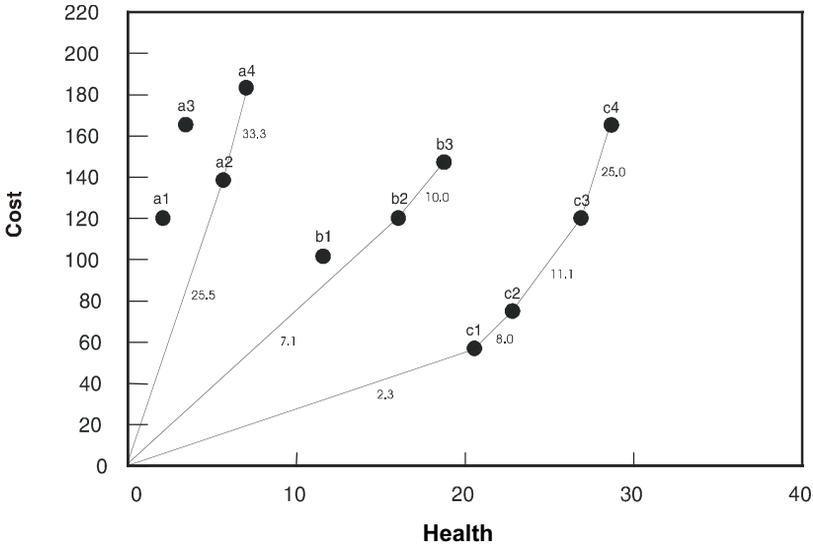
set of interventions a1–a6, labelled simply α2. This format follows standard practice in the literature.

Figure 2 will be used to illustrate one of the main limitations of IMC-CEA. Eleven different interventions to those of Figure 1 are divided into

Table 1   Average cost effectiveness for 11 interventions

Intervention	Costs	Health benefits	Average cost-effectiveness
a1	120	1	120
a2	140	5.5	25.45
a3	170	3	56.67
a4	190	7	27.14
b1	100	12	8.33
b2	120	17	7.06
b3	150	20	7.5
c1	50	22	2.27
c2	70	24.5	2.86
c3	120	29	4.14
c4	170	31	5.48

three sets of mutually exclusive interventions, a1–a4, b1–b3 and c1–c4. Costs and health benefits for each intervention are shown with respect to the null set of this set of 11 interventions—health benefits could be denominated in QALYs gained, DALYs averted or some other general measure of health. In other words, costs and benefits are shown compared with the costs and benefits in the absence of any of these interventions. Table 1 provides the costs and benefits for each intervention and the average cost-effectiveness of each with respect to the null set.

Consider a population where a budget of 170 is currently spent to purchase a1 and c1 producing 23 units of health. Next, consider an increase in the budget from 170 to 190. The remaining set of mutually exclusive interventions with respect to a1 would be evaluated. It shows that a3 is dominant and yields the incremental cost-effectiveness ratios in Table 2, which also shows similar calculations for the independent sets of interventions. A decision-maker would choose to purchase a2 instead of a1 because moving from a1 to a2 has the lowest incremental cost-effectiveness ratio. The final combination of a2 and c1 yields 27.5 units of health.

Table 2   Sequential incremental cost-effectiveness ratios starting from a1–c1

Category A $\Delta C/\Delta E$		Category B $\Delta C/\Delta E$		Category C $\Delta C/\Delta E$	
a2	4.4	b1	8.3	c2	8.0
a3	Dominated	b2	7.1	c3	11.1
a4	33.3	b3	10.0	c4	25.0

Consider another population where a budget of 170 is currently spent on a3 yielding only 3 health units. In this population, incremental CEA of the remaining interventions with respect to the starting point of a3 would yield the ratios in Table 3. If the budget now increases from 170 to 190, the decision-maker would first choose to save money and increase health output by moving to a2. With the savings of 30 and the increased budget of 20, the next most attractive intervention would be to purchase c1, with the resulting allocation of resources being a2 and c1 yielding 27.5 units of health.

Table 3   Sequential incremental cost-effectiveness ratios starting from a3

Category A $\Delta C/\Delta E$		Category B $\Delta C/\Delta E$		Category C $\Delta C/\Delta E$	
a2	-12	b1	8.3	c2	2.3
a4	33.3	b2	7.1	c3	8.0
		b3	10.0	c3	11.1
				c4	25.0

In both examples, IMC-CEA identified health enhancing resource allocations but the basic fact that the C and B category interventions are

much more cost-effective than the A category interventions does not emerge from the analysis. This is because the cost-effectiveness of the starting point is not evaluated in current practice. As detailed below, it is relatively straightforward to identify the health maximizing combination of interventions for a budget of 170 as c1 and b2, which yields 39 health units and the health maximizing combination of interventions for a budget of 190 is c2 and b2 yielding 41.5 health units.[d] In reality there is likely to be substantial allocative inefficiency in current allocations of health resources in many settings, and this example demonstrates that the application of IMC-CEA may fail to identify major opportunities for enhancing the overall cost-effectiveness of the health system.[e]

The intervention mix constraint on CEA means that major allocative inefficiencies may not be evaluated and thus identified. If the current intervention mix is an unavoidable constraint on decision-makers in a given context, then this is appropriate for context-specific CEA analyses. In most situations, however, other constraints on decision-makers may be more pervasive. As described above, in many health systems with a large share of public provision there is a fixed stock of community and referral hospitals, which cannot be modified in the short- to medium-term for powerful political reasons. Likewise, in many countries the supply of different types of health providers (nurses, general practitioners, specialists or community health workers) may limit the set of interventions that can be delivered. These decision constraints may be more common than the strict commitment to the current mix of interventions assumed in current practice—it may be easier to shift spending from the treatment of ischaemic heart disease to childhood immunization programmes than to shut district hospitals or import ophthamologists.

Table 4   Optimal solutions with two constraints

Interventions	Total cost	Current budget = 70	Infrastructure = 100	Benefit	Average cost-effectiveness	Benefit at current use
a1	120	60	60	1	120.00	
a2	140	80	60	5.5	25.45	
a3	170	90	80	3	56.67	
a4	190	110	80	7	27.14	
b1	100	35	65	12	8.33	12
b2	120	60	60	17	7.06	
b3	150	75	75	20	7.50	
c1	50	15	35	22	2.27	
c2	70	35	35	24.5	2.86	24.5
c3	120	50	70	29	4.14	
c4	170	85	85	31	5.48	
				Total benefit		36.5
				Slack in the current budget		0
				Slack in the infra budget		0

If the focus of sectoral CEA is to inform context-specific decision making, then methods need to be developed to incorporate these and

other constraints on the set of possible decisions. This can be achieved relatively easily through the use of optimal resource allocation planning models adapted to the health sector *(22,26–34)*. For example, Table 4 illustrates using a simple resource allocation model that the health maximizing resource allocation in the setting of two binding constraints (physical capacity of health facilities and fungible dollars[f]) is substantially different than the health maximizing resource allocation in the setting of only a dollar constraint. Using the data from Table 1, the total budget is set at 170, 70 of which is fungible dollars and the rest is the constraint on infrastructure or the physical capacity of health facilities valued at 100. For each intervention, we have divided the costs of Table 1 into two components—fungible dollars and infrastructure. With a single budget constraint of 170, optimal allocation required provision of b2 and c1 with a benefit of 39. The dual constraints of Table 4 now require b1 and c2 to be carried out at a benefit of 36.5, because the two constraints must be met. With multiple constraints, there is no easy way of developing a cost-effectiveness league table and more complex programming models should be used to allocate resources. In this case, the solution was obtained with 0–1 linear programming solved using the programming language LINGO®.

Generalized CEA

For some decision-makers, the development of complex resource allocation models that explicitly incorporate a range of decision constraints and multiple objectives may be very useful. However, such efforts are information intensive, time consuming, costly and very often difficult to communicate to the full set of actors in any health policy dialogue *(18)*. We believe that CEA can be most useful with more modest goals by focusing on the more general use of cost-effectiveness information to inform health policy debates without being completely contextualized. Moreover, sectoral CEA should identify current allocative inefficiencies as well as opportunities presented by new interventions. For this reason, WHO will propose a modification of the standard ICM-CEA lifting the constraint on the current mix of interventions to evaluate the cost-effectiveness of all options including currently funded interventions.

In brief, the basic modification can be summarized in two propositions.

1. The costs and benefits of a set of related interventions should be evaluated with respect to the counterfactual of the null set of the related interventions. This is illustrated in Figure 2 for the 11 interventions. This provides the complete set of information for evaluating both independent and mutually exclusive options to identify the health maximizing combination of interventions for any given budget.

2. Results of CEA should be presented in a single league table. For each set of mutually exclusive interventions, the intervention with the lowest average cost-effectiveness ratio (the lowest slope in the figure of cost versus benefit) with respect to the null set appears first in the league table. The second intervention from the set (if there are at least two) that appears in the league table is the one with the lowest slope with respect to the intervention with the lowest CE ratio that already appeared in the table. The third intervention is the one with the lowest slope with respect to the second intervention, etc. Weakly dominated interventions do not appear in the league table. The results for all sets of mutually exclusive interventions are shown in the same league table according to the same principles. The application of this simple approach to the 11 interventions example in Figure 2 is shown in Table 5. Interventions a1, a3 and b1 are weakly dominated and do not appear. For heuristic purposes, the health maximizing combination for any budget level can be selected from the table. These decision rules are similar to those that have been derived for IMC-CEA but the analysis starts from the origin (*18,68,126–133*).

Table 5   Generalized cost-effectiveness league table

Intervention	Cost-effectiveness ratio
c1	2.3
b2	7.1
c1–c2	8.0
b2–b3	10.0
c2–c3	11.1
c3–c4	25.0
a2	25.5
a2–a4	33.3

A key issue in this or any other approach to CEA is defining an intervention. If the comparator for a set of related interventions is the null set then each intervention must be defined with respect to that null set. Thus, if a new drug shortens the length of stay and reduces complication rates post coronary artery bypass graft operations, the drug is not the intervention. The intervention is coronary artery bypass graft plus the new drug. This logic in defining interventions allows for complex interactions in costs and health benefits to be easily captured and represented in a league table.

Figure 3 illustrates such an evaluation for four interventions for tuberculosis: passive case detection and treatment with directly observed short course therapy (DOTS), BCG vaccination at 50% coverage, BCG at 75% coverage and BCG at 100% coverage. In addition, three other mutually exclusive options are presented: passive case detection and treatment with DOTS combined with the three different levels of BCG

Figure 3   Costs and benefits of interventions with cost and
effectiveness interactions

	Costs	Benefits	CE
A	550	500	110
b1	180	200	90
b2	325	300	108.3
b3	600	400	150
Ab1	631	600	105.2
Ab2	726.5	650	111.8
Ab3	952	700	136

coverage. Costs interact, in that, if BCG is delivered, the number of cases of tuberculosis that will occur, be detected and accept treatment will decline so that the variable cost component of the treatment programme will decline but the fixed cost component will not. Likewise, the health benefits of BCG in the presence of a treatment programme will be less because many of the deaths from tuberculosis expected in the absence of treatment will be avoided.

Using a multiplicative model, the interaction of the benefits of the two programmes can be estimated. The lines in Figure 3 indicate graphically the league table for this set of mutually exclusive interventions, in order: BCG at 50% coverage, BCG at 50% coverage combined with passive detection and treatment, BCG 75% with detection and treatment and BCG 100% with detection and treatment. BCG 75%, BCG 100% and passive detection and treatment alone do not appear in the list as they are dominated by the other alternatives.

In the literature on cost-effectiveness *(23,30,34)* there has been considerable concern about nonlinear cost-effectiveness functions; for example, the cost per DALY averted through the expansion of measles coverage from 50% to 90% is likely to be much lower than the cost per DALY averted through the expansion of coverage from 90% to 99%. Because interventions at different levels of coverage are clearly mutually exclusive at the population level, then the same approach outlined above can be used to capture in a series of discrete points a non-linear cost-effectiveness function. In Figure 2, the set of interventions c1–c4 could be different strategies or different levels of coverage for the same strategy. By picking a parsimonious set of coverages, a set of indivisible and mutually exclusive interventions can be defined and the key consequences of nonlinear cost-effectiveness functions captured in a single league table. The tuberculosis example of Figure 2 clearly does not by itself represent an example of generalized CEA, but would be part of the larger league table used to inform the policy debate.

By analysing the costs and benefits of sets of related interventions with respect to the null set of those interventions, the results are likely to be

more transferable from one population to another —though only through experience will we learn if this is true. Clearly, the costs of different resource inputs to the production of a given intervention vary across populations as do some of the determinants of effectiveness *(15,59,60,63–68)*.ᵍ However, one major factor limiting the relevance of ICM-CEA results in one population to another population, namely different current mixes of interventions, can be removed by using the generalized CEA approach. To put it another way, the null set for a group of related interventions is more comparable across populations (or at least sets of populations) than the current mix of interventions. Nevertheless, there are clear limits to the comparability across populations of the counterfactual null set. It will depend on the development of the health system and on the epidemiological pattern. Clearly, global comparisons of the cost effectiveness of interventions with respect to the null set even if input costs and effectiveness determinants are adjusted is unlikely to be useful.

The strategy for the development of this idea will be to define a limited set of average health system and epidemiological contexts within which null set comparisons are likely to be informative. Many groupings of countries or communities could be developed, on the basis of income per capita, region, public/private splits in health care finance or provision, burden of disease, etc. This will be one major challenge for the development of this approach.

The benefits of analysing the costs and health benefits of interventions with respect to the null set for a group of related interventions appears to be greater but the technical challenge of estimating the conditions in the null set counterfactual need to be addressed. In theory, in ICM-CEA, costs and benefits of each intervention are evaluated with respect to the current mix of interventions but many studies are based on retrospective analysis where the intervention cost and benefits are evaluated with respect to a past mix of interventions not necessarily the current mix *(134,135)*. Likewise, estimates of benefits of interventions that involve a time lag between purchase and benefit, such as hepatitis B immunization, are based on relatively implausible assumptions that the current mix of interventions will apply in the future *(136–139)*. A symptom of this problem is demonstrated by the standard practice in ICM-CEA of estimating the benefits of life saving interventions using period life tables when in fact the cohort life expectancy at each age would be a more accurate (but more difficult to estimate) estimate of the years of life gained. Historically, cohort life expectancy has been 10–20 years higher at birth than period life expectancy *(140)* so that this is not a minor bias.

Estimating the null set conditions for a group of related interventions will require the development of natural history models. Some have already been developed and some have been used in cost-effectiveness studies *(137,141–149)*. De Koning et al. *(150,151)* have developed a natural history model for breast cancer in the Netherlands as part of an in-depth

analysis of intervention options for breast cancer. To implement this generalized approach to CEA, clear guidelines and standards on the development of natural history models will need to be developed as a priority.

DISCUSSION

Broader use of cost-effectiveness studies to analyse the allocative efficiency of health systems and recommend resource allocations has led to a number of challenges. It appears that the field can develop in two distinct directions, towards increasingly contextualized analyses or towards more generalized assessments. Cost-effectiveness studies and the sectoral application of CEA to a wide range of interventions can become increasingly context specific; at the individual study level by incorporating directly other social concerns, such as distributional weights or a priority to treating the sick, and at the sectoral level by developing complex resource allocation models that capture the full range of resource, ethical and political constraints facing decision-makers. We fear that this direction will lead ultimately to less use of cost-effectiveness information in health policy dialogue. Highly contextualized analyses must by definition be undertaken in each context, the cost and time involved as well as the inevitable complexity of the resource allocation models will limit their practical use.

The other direction for sectoral cost-effectiveness, the direction that we are suggesting, is to focus on the general assessment of the costs and health benefits of different interventions in the absence of various highly variable local decision constraints. A general league table of the cost effectiveness of interventions for a group of populations with comparable health systems and epidemiological profiles can make the most powerful component of CEA readily available to inform health policy debates. Judgements on the relative cost-effectiveness of interventions such as DOTS for tuberculosis is highly cost-effective and liver transplants for alcoholic cirrhosis are highly cost ineffective, can have wide ranging influence—as one input to an informed policy debate they can enhance the allocative efficiency of many health systems. Information on generalized cost-effectiveness can be used alongside consideration of the effect of different resource allocations on other important social goals, such as equity. Because we believe this is the most constructive use of cost-effectiveness information, we would like to open for debate the proposal to modify standard cost-effectiveness methods. The modifications proposed, to remove the current intervention mix decision constraint, will expose current allocative inefficiencies to analysis and at the same time enhance the transferability of results from one population to another.

For many narrower applications of CEA, such as the appraisal of new drugs in a specific country, the currently practised ICM-CEA remains the most appropriate method. Nevertheless, even in these circumstances it

would be useful for authors to also estimate the costs and health benefits of interventions with respect to the null set. This would substantially improve the world's body of knowledge on the cost-effectiveness of different interventions. In this way, each new study would add to our collective knowledge of the relative costs and effectiveness of different interventions.

Notes

a. The term allocative efficiency can be used in many ways. Here, we strictly use it to refer to whether resources are allocated across different health interventions (specific public health, curative, promotive, rehabilitative, or palliative interventions) so as to maximize population health status.

b. Some of the problems of international transferability of results even for generalized CEA are discussed later in the paper.

c. Issues of divisibility of interventions are at the heart of many of the theoretical issues in CEA. For example, the definition of extended dominance *(23,24,34)* depends on the assumption of divisibility. In fact, divisibility of interventions is only required because of the problems of the choice of the last intervention with a hard budget constraint. If the most cost-effective intervention is indivisible and costs more than the available slack in the budget, then other interventions, including some that may be weakly dominated, may be in the optimal resource allocation. The issue of divisibility of interventions often plagues simple illustrations of optimal resource allocation across a small set of interventions *(22,123,124,141,142)*. For these graphical representations to provide clear and correct answers, it is necessary that each possible combination of cost and benefit that could be implemented be represented as a specified point. In reality, for most programmes, one cannot purchase any level of coverage for technical or political reasons. For example, in implementing short-course chemotherapy for smear-positive tuberculosis using passive case detection, by the nature of the case detection modality only one level of coverage can be achieved with that strategy. To change the coverage would require an explicit change in the case detection strategy, such as active screening or public awareness campaigns, which would have different costs and benefits and thus should be seen as another incompatible intervention. Alternatively, while it is theoretically possible to envisage a vaccination strategy that targets only a quarter or a half of the population, it would be impossible to implement for political reasons in most countries. In reality, there would be a few mutually exclusive combinations of costs and coverage for most programmes. The decision rules developed in this paper apply to this situation.

As has been argued above, faced with a budget constraint and a series of indivisible interventions, the health maximizing allocation of the budget is complicated by issues of slack—close to the budget constraint, it might not be possible to fully implement the preferred intervention. The examples in this paper have been designed to avoid these problems, but we do not believe that slack is a critical issue in any real allocation decision. First, the size of any slack *vis-à-vis* the total budget is likely to be very small *(13)*. Slack problems are exaggerated in the practical examples in the literature, where the number of

interventions purchased is always small, which means that slack may be a large per cent of the budget. In any real health system, slack related to the last intervention selected is likely to be very small. Second, in any real health system, budget constraints are never so firmly fixed that issues of slack become an issue in actual debates on resource allocation. In fact, we strongly believe that results of the type of CEA proposed in this paper should not be used with such precision.

A more important issue concerns the situation concerning the indivisibility of a capital investment, where the investment can be used for several patient or population groups, such as a hospital. Such problems can only be addressed with resource allocation models *(13)*. For example, Murray *et al. (24)* developed a resource allocation model in which expansion of capital infrastructure was evaluated as a separate type of intervention, which relaxed the physical infrastructure constraint in the resource allocation model.

d. A simpler approach to allocating resources across a set of interventions might be to rank all independent and mutually exclusive interventions by their average cost-effectiveness and then fund down the list of interventions until the budget is exhausted. In this example, for a budget of 170, the average cost-effectiveness rank list approach would choose intervention c4 producing 31 health units. This is substantially less than the health maximizing combination of c1 and b2 yielding 39 units. Average cost-effectiveness rank lists that ignore the issues related to mutually exclusive interventions will in general yield sub-optimal resource allocations.

e. This point has been made in various forms in the literature, e.g. see Drummond *et al. (58)*.

f. We use the term 'fungible dollars' to describe the assumption that no constraints other than physical capital and the total budget are binding. The total budget can be moved between all inputs other than capital with no restrictions.

g. A challenge to our approach will be to separate out technical inefficiencies in production of a given intervention from the allocative efficiency questions described here. For example, it has also been shown that the physical quantities of resources used for a given intervention can vary from place to place according to practice patterns *(152,153)*. If by chance the cost-effectiveness of an intervention has been evaluated in a setting that is technically inefficient and another is evaluated in a setting that is technically efficient, conclusions on relative cost-effectiveness may be biased. The confounding effect of variation in technical efficiency across study locations for the development of generalized cost-effectiveness league tables needs to be minimized. At the same time, systematic regional variation in technical efficiency due to health system characteristics or epidemiological patterns should be incorporated into regional league tables of generalized cost-effectiveness.

References

1. al Umran K, Yaseen H. Cost-effectiveness of surfactant replacement therapy in a developing country. *Journal of Tropical Pediatrics* 1997; 43: 167–169.

2. Avants SK, Margolin A, Sindelar JL. *et al.* Day treatment versus enhanced standard methadone services for opioid-dependent patients: a comparison of clinical efficacy and cost. *American Journal of Psychiatry* 1999; 156: 27–33.

3. Mauskopf J, Lacey L, Kempel A. *et al.* The cost-effectiveness of treatment with lamivudine and zidovudine compared with zidovudine alone: a comparison of Markov model and trial data estimates. *American Journal of Management Care* 1998; 4: 1004–1012.

4. Morris S, Gray A, Noone A. *et al.* The costs and effectiveness of surveillance of communicable disease: a case study of HIV and AIDS in England and Wales. *Journal of Public Health Medicine* 1996; 18: 415–422.

5. Platis H, Liaropoulos L. Comparative cost analysis of two different medical interventions: educational implications. *Studies in Health Technology and Information* 1998; 51: 50–64.

6. Porta M, Rizzitiello A, Tomalino M. *et al.* Comparison of the cost-effectiveness of three approaches to screening for and treating sight-threatening diabetic retinopathy. *Diabetes Metabolism* 1999; 25: 44–53.

7. Streitz JMJ, Ellis FHJ, Tilden RL. *et al.* Endoscopic surveillance of Barrett's esophagus: a cost-effectiveness comparison with mammographic surveillance for breast cancer. *American Journal of Gastroenterology* 1998; 93: 911–915.

8. Szucs TD. Pharmaco-economic aspects of lipid-lowering therapy: is it worth the price? *European Heart Journal* 1998; 19(Suppl M): 22–28.

9. Willke RJ, Glick HA, Polsky D. *et al.* Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Economics* 1998; 7: 481–493.

10. Elixhauser A, Luce BR, Taylor WR. *et al.* Health care CBA/CEA: an update on the growth and composition of the literature. *Medical Care* 1993; 31(Suppl): 1–149.

11. Elixhauser A, Halpern M, Schmier J. *et al.* Health care CBA and CEA from 1991 to 1996: an updated bibliography. *Medical Care* 1998; 36(Suppl): 1–147.

12. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? *Social Science Medicine* 1993; 37: 33–40.

13. Hoerger TJ, Bala MV, Rowland C. *et al.* Cost effectiveness of pramipexole in Parkinson's disease in the US. *PharmacoEconomics* 1998; 14: 541–557.

14. Mason J, Drummond M, Torrance G. Some guidelines on the use of cost effectiveness league tables. *British Medical Journal* 1993; 306: 570–572.

15. Schulman KA, Lynn LA, Glick HA. *et al.* Cost effectiveness of low-dose zidovudine therapy for asymptomatic patients with human immunodeficiency virus (HIV) infection. *Annals of Internal Medicine* 1991; 114: 798–802.

16. Torrance GW, Zipursky A. Cost-effectiveness of antepartum prevention of Rh immunization. *Clinical Perinatology* 1984; 11: 267–281.

17. Williams A. Economics of coronary artery bypass grafting. *British Medical Journal, Clinical Research Edition* 1985; 291: 326–329.

18. Karlsson G, Johannesson M. Cost-effectiveness analysis and capital costs. *Social Science Medicine* 1998; 46: 1183–1191.

19. Ackerman SJ, Sullivan EM, Beusterien KM. *et al.* Cost effectiveness of recombinant human insulin-like growth factor I therapy in patients with ALS. *PharmacoEconomics* 1999; 15: 179–195.

20. Hay JW, Yu WM, Ashraf T. Pharmacoeconomics of lipid-lowering agents for primary and secondary prevention of coronary artery disease. *PharmacoEconomics* 1999; 15: 47–74.

21. Patel ST, Haser PB, Bush HLJ. *et al.* The cost-effectiveness of endovascular repair versus open surgical repair of abdominal aortic aneurysms: a decision analysis model. *Journal of Vascular Surgery* 1999; 29: 958–972.

22. Johannesson M, Weinstein MC. On the decision rules of cost-effectiveness analysis. *Journal of Health Economics* 1993; 12: 459–467.

23. Birch S, Gafni A. Cost effectiveness:utility analyses. Do current decision rules lead us to where we want to be? *Journal of Health Economics* 1992; 11: 279–296.

24. Birch S, Gafni A. Changing the problem to fit the solution: Johannesson and Weinstein's (mis)application of economics to real world problems. *Journal of Health Economics* 1993; 12: 469– 476.

25. Karlsson G, Johannesson M. The decision rules of cost-effectiveness analysis. *PharmacoEconomics* 1996; 9: 113–120.

26. Nord E. Health status index models for use in resource allocation decisions. A critical review in the light of observed preferences for social choice. *International Journal of Technology Assessment in Health Care* 1996; 12: 31–44.

27. Ubel PA, DeKay ML, Baron J. *et al.* Cost-effectiveness analysis in a setting of budget constraints—is it equitable? *New England Journal of Medicine* 1996; 334: 1174–1177.

28. Weinstein MC. Principles of cost-effective resource allocation in health care organizations. *International Journal of Technology Assessment in Health Care* 1990; 6: 93–103.

29. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine* 1977; 296: 716–721.

30. Birch S, Donaldson C. Cost-benefit analysis: dealing with the problems of indivisible projects and fixed budgets. *Health Policy* 1987; 7: 61–72.

31. Birch S, Leake JL, Lewis DW. Economic issues in the development and use of practice guidelines: an application to resource allocation in dentistry. *Community Dental Health* 1996; 13: 70–75.

32. Murray CJ, Kreuser J, Whang W. Cost-effectiveness analysis and policy choices: investing in health systems. *Bulletin of the World Health Organization* 1994; 72: 663–674.

33. Chen MM, Bush JW. Maximizing health system output with political and administrative constraints using mathematical programming. *Inquiry* 1976; XIII: 215–227.

34. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *Journal of Health Economics* 1996; 15: 641–653.

35. Oregon Health Services Commission. *Oregon Medicaid priority setting project.* Oregon State Government, Portland, OR, 1991.

36. Blumstein JF. The Oregon experiment: the role of cost-benefit analysis in the allocation of Medicaid funds. *Social Science Medicine* 1997; 45: 545–554.

37. Brown LD. The national politics of Oregon's rationing plan. *Health Affiliation Millwood* 1991; 10: 28–51.

38. Dixon J, Welch HG. Priority setting: lessons from Oregon. Lancet 1991; 337: 891–894.

39. Fox DM, Leichter HM. Rationing care in Oregon: the new accountability. *Health Affiliation Millwood* 1991; 10: 7–27.

40. Kitzhaber JA. Prioritising health services in an era of limits: the Oregon experience. *British Medical Journal* 1993; 307: 373–377.

41. Jamison DT, Mosley WH, Measham AR, Bobadilla JL. *Disease control priorities in developing countries.* Set. New York: Oxford University Press, 1993.

42. Tengs TO, Adams ME, Pliskin JS. *et al.* Five hundred life-saving interventions and their cost-effectiveness. *Risk Analysis* 1995; 15: 369–390.

43. Tengs TO, Meyer G, Siegel JE. *et al.* Oregon's Medicaid ranking and cost-effectiveness: is there any relationship? *Medical Decision Making* 1996; 16: 99–107.

44. Bobadilla JL. *Searching for essential health services in low- and middle-income countries: a review of recent studies on health priorities.* Washington DC: Human Development Department, World Bank, 1996.

45. Bobadilla JL, Cowley P, Musgrove P. *et al.* Design, content and financing of an essential national package of health services. *Bulletin of the World Health Organization* 1994; 72: 653–662.

46. Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). *British Medical Journal* 1992; 305: 875–877.

47. Nord E, Richardson J, Macarounas KK. Social evaluation of health care versus personal evaluation of health states. Evidence on the validity of four health-state scaling instruments using Norwegian and Australian surveys. *International Journal of Technology Assessment in Health Care* 1993; 9: 463–478.

48. Nord E. The relevance of health state after treatment in prioritising between different patients. *Journal of Medical Ethics* 1993; 19: 37–42.

49. Nord E, Richardson J, Street A. *et al.* Maximizing health benefits vs egalitarianism: an Australian survey of health issues. *Social Science Medicine* 1995; 41: 1429–1437.

50. Black D, Morris JN, Smith C. *et al.* Better benefits for health: plan to implement the central recommendation of the Acheson report. *British Medical Journal* 1999; 318: 724–727.

51. Jarman B, Bosanquet N. Primary health care in London—changes since the Acheson report. *British Medical Journal* 1992; 305: 1130–1133.

52. Williams A. Intergenerational equity: an exploration of the 'fair innings' argument. *Health Economics* 1997; 6: 117–132.

53. Williams M. Rationing health care. Can a 'fair innings' ever be fair? *British Medical Journal* 1997; 314: 514.

54. Brock DW. The role of ethics in the Clinton reform proposal. *Journal of Health and Political Policy and Law* 1994; 19: 217–220.

55. Brock DW, Daniels N. Ethical foundations of the Clinton administration's proposed health care system. *Journal of the American Medical Association* 1994; 271: 1189–1196.

56. Richardson J, Nord E. The importance of perspective in the measurement of quality adjusted life years. *Medical Decision Making* 1997; 17: 33–41.

57. Gold MR, Siegel JE, Russel LB, Weinstein MC. (Eds) *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.

58. Drummond M, O'Brien BJ, Stoddart GL, Torrance G. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press, 1997.

59. Drummond MF, Bloom BS, Carrin G. *et al*. Issues in the cross-national assessment of health technology. *International Journal of Technology Assessment in Health Care* 1992; 8: 671–682.

60. Leidl RM. Some factors to consider when using the results of economic evaluation studies at the population level. *International Journal of Technology Assessment in Health Care* 1994; 10: 467–478.

61. Birch S, Gafni A. Cost-effectiveness ratios: in a league of their own. *Health Policy* 1994; 28: 133–141.

62. Drummond M, Mason J, Torrance G. Cost-effectiveness league tables: think of the fans. *Health Policy* 1995; 31: 231–238.

63. Glick H, Willke R, Polsky D. *et al*. Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage: economic evaluation of a phase III clinical trial in Europe and Australia. *International Journal of Technology Assessment in Health Care* 1998; 14: 145–160.

64. Johnston K, Gerard K, Brown J. Generalizing costs from trials: analyzing center selection bias in a breast screening trial. *International Journal of Technology Assessment in Health Care* 1998; 14: 494–504.

65. Mason J. The generalisability of pharmacoeconomic studies. *Pharmaco-Economics* 1997; 11: 503–514.

66. Schulman K, Burke J, Drummond M. *et al*. Resource costing for multinational neurologic clinical trials: methods and results. *Health Economics* 1998; 7: 629–638.

67. Willke RJ, Glick H, Polsky D. *et al*. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Economics* 1998; 7: 493.

68. Baltussen R, Ament A, Leidl R. Making cost assessments based on RCTs more useful to decision-makers. *Health Policy* 1996; 37: 163–183.

69. Commonwealth of Australia. *Guidelines for pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee:*

including economic analyses. Canberra: Department of Health and Community Services, 1995.

70. Mitchell A. Update and evaluation of Australian guidelines. Government perspective. *Medical Care* 1996; 34(Suppl): 216–225.

71. Langley PC. The November 1995 revised Australian guidelines for the economic evaluation of pharmaceuticals. *PharmacoEconomics* 1996; 9: 341–352.

72. Freund DA. Initial development of the Australian guidelines. *Medical Care* 1996; 34(Suppl): 211–215.

73. Aristides M, Mitchell A. Applying the Australian guidelines for the reimbursement of pharmaceuticals. *PharmacoEconomics* 1994; 6: 196–201.

74. Gorham P. Cost-effectiveness guidelines. The experience of Australian manufacturers. *PharmacoEconomics* 1995; 8: 369–373.

75. Canadian Coordinating Office for Health Technology Assessment (CCOHTA). *Guidelines for the economic evaluation of pharmaceuticals: Canada*. Ottawa: CCOHTA, 1994.

76. Baladi JF, Menon D, Otten N. Use of economic evaluation guidelines: 2 years' experience in Canada. *Health Economics* 1998; 7: 221–227.

77. Detsky AS. Guidelines for economic analysis of pharmaceutical products: a draft document for Ontario and Canada. *PharmacoEconomics* 1993; 3: 354–361.

78. Glennie JL, Torrance G, Baladi JF. *et al*. The revised Canadian guidelines for the economic evaluation of pharmaceuticals. *PharmacoEconomics* 1999; 15: 459–468.

79. Ontario Ministry of Health. *Ontario guidelines for economic analysis of pharmaceutical products*. Toronto: Ontario Ministry of Health, 1994.

80. Menon D, Schubert F, Torrance GW. Canada's new guidelines for the economic evaluation of pharmaceuticals. *Medical Care* 1996; 34(Suppl): 77–86.

81. Torrance GW, Blaker D, Detsky A. *et al*. Canadian guidelines for economic evaluation of pharmaceuticals. Canadian Collaborative Workshop for Pharmacoeconomics. *PharmacoEconomics* 1996; 9: 535–559.

82. Report from the Canadian Coordinating Office for Health Technology Assessment (CCOHTA). Guidelines for economic evaluation of pharmaceuticals: Canada. *International Journal of Technology Assessment in Health Care* 1995; 11: 796–797.

83. Lovatt B. The United Kingdom guidelines for the economic evaluation of medicines. *Medical Care* 1996; 34(Suppl): 179–181.

84. Clemens K, Townsend R, Luscombe F. *et al*. Methodological and conduct principles for pharmacoeconomic research. Pharmaceutical Research and Manufacturers of America. *PharmacoEconomics* 1995; 8: 169–174.

85. Ikeda S, Ikegami N, Oliver AJ. *et al*. A case for the adoption of pharmacoeconomic guidelines in Japan. *PharmacoEconomics* 1996; 10: 546–551.

86. Alban A, Gyldmark M, Pedersen AV. *et al*. The Danish approach to standards for economic evaluation methodologies. *PharmacoEconomics* 1997; 12: 627–636.

87. Graf von der Schulenburg JM. Economic evaluation of medical technologies: from theory to practice—the German perspective. *Social Science Medicine* 1997; 45: 621–633.

88. Garattini L, Grilli R, Scopelliti D. *et al.* A proposal for Italian guidelines in pharmacoeconomics: The Mario Negri Institute Centre for Health Economics. *PharmacoEconomics* 1995; 7: 1–6.

89. Rovira J. Standardizing economic appraisal of health technology in the European Community. *Social Science Medicine* 1994; 38: 1675–1678.

90. Rovira J. Standardization of the economic evaluation of health technologies. European developments. *Medical Care* 1996; 34(Suppl): 182–188.

91. Task Force on Principles for Economic Analysis of Health Care Technology. Economic analysis of health care technology. A report on principles. *Annals of Internal Medicine* 1995; 123: 61–70.

92. Bossert W. Welfarism and rationalizability in allocation problems with indivisibilities. *Mathematical Social Sciences* 1998; 35: 133–150.

93. Kassirer JP, Angell M. The journal's policy on cost-effectiveness analysis. *New England Journal of Medicine* 1994; 331: 669–670.

94. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *British Medical Journal* 1996; 313: 275–283.

95. Luce BR, Elixhauser A, Culyer AJ, Horisberger B. (Eds) *Standards for socioeconomic evaluation of health care products and services.* Berlin: Springer, 1990.

96. Rittenhouse BE. The standards debate. What have we learned? *Medical Care* 1996; 34(Suppl): 5–10.

97. Luce BR, Simpson K. Methods of cost-effectiveness analysis: areas of consensus and debate. *Clinical Therapy* 1995; 17: 109–125.

98. Freund DA, Dittus RS. Principles of pharmacoeconomic analysis of drug therapy. *PharmacoEconomics* 1992; 1: 20–31.

99. Coyle D. Statistical analysis in pharmacoeconomic studies. A review of current issues and standards. *PharmacoEconomics* 1996; 9: 506–516.

100. Schulman KA, Linas BP. Pharmacoeconomics: state of the art in 1997. *Annual Reviews in Public Health* 1997; 18: 529–548.

101. Blackmore CC, Smith WJ. Economic analyses of radiological procedures: a methodological evaluation of the medical literature. *European Journal of Radiology* 1998; 27: 123–130.

102. Blackmore CC, Magid DJ. Methodologic evaluation of the radiology cost–effectiveness literature. *Radiology* 1997; 203: 87–91.

103. Hayman J, Weeks J, Mauch P. Economic analyses in health care: an introduction to the methodology with an emphasis on radiation therapy. International Journal of Radiation in Oncology and Biological Physics 1996; 35: 827–841.

104. Talley CR. Pharmacoeconomic principles. *American Journal of Health Systems and Pharmaceutics* 1995; 52: 1871.

105. Luce BR. Cost-effectiveness analysis: obstacles to standardisation and its use in regulating pharmaceuticals. *PharmacoEconomics* 1993; 3: 1–9.

106. Lafata JE, Koch GG, Ward RE. Synthesizing evidence from multiple studies. The role of meta-analysis in pharmacoeconomics. *Medical Care* 1996; 34(Suppl): 136–145.

107. Jolicoeur LM, Jones-Grizzle AJ, Boyer JG. Guidelines for performing a pharmacoeconomic analysis. *American Journal of Hospitals and Pharmacy* 1992; 49: 1741–1747.

108. Rittenhouse BE. Is there a need for standardization of methods in economic evaluations of medicine? *Medical Care* 1996; 34(Suppl): 13–22.

109. Gafni A, Birch S. Guidelines for the adoption of new technologies: a prescription for uncontrolled growth in expenditures and how to avoid the problem. *Canadian Medical Association Journal* 1993; 148: 913–917.

110. Sacristan JA, Soto J, Galende I. Evaluation of pharmacoeconomic studies: utilization of a checklist. *Annals of Pharmacotherapy* 1993; 27: 1126–1133.

111. Lee JT, Sanchez LA. Interpretation of 'cost-effective' and soundness of economic evaluations in the pharmacy literature. *American Journal of Hospitals and Pharmacy* 1991; 48: 2622–2627.

112. Henry D. Economic analysis as an aid to subsidisation decisions: the development of Australian guidelines for pharmaceuticals. *PharmacoEconomics* 1992; 1: 54–67.

113. Johannesson M, O'Brien BJ. Economics, pharmaceuticals, and pharmacoeconomics. *Medical Decision Making* 1998; 18(Suppl): 1–3.

114. Luce BR. Working toward a common currency: is standardization of cost-effectiveness analysis possible? *Journal of Acquired Immune Deficiency Syndrome Human Retrovirology* 1995; 10: 19–22.

115. Siegel JE, Torrance GW, Russell LB. *et al.* Guidelines for pharmacoeconomic studies. Recommendations from the panel on cost effectiveness in health and medicine. Panel on Cost Effectiveness in Health and Medicine. *PharmacoEconomics* 1997; 11: 159–168.

116. Drummond MF. Guidelines for pharmacoeconomic studies. The ways forward. *PharmacoEconomics* 1994; 6: 493–497.

117. Drummond MF. A reappraisal of economic evaluation of pharmaceuticals. Science or marketing? *PharmacoEconomics* 1998; 14: 1–9.

118. Jacobs P, Bachynsky J, Baladi JF. A comparative review of pharmacoeconomic guidelines. *PharmacoEconomics* 1995; 8: 182–189.

119. Mason J, Drummond M. Reporting guidelines for economic studies. *Health Economics* 1995; 4: 85–94.

120. Drummond M, Mooney G. Economic appraisal in health care: 1. A guide to the methodology of economic appraisal. *Hospital Health Services Review* 1981; 77: 277–282.

121. Drummond M. Cost-effectiveness guidelines for reimbursement of pharmaceuticals: is economic evaluation ready for its enhanced status? *Health Economics* 1992; 1: 85–92.

122. Drummond MF, Davies L. Economic analysis alongside clinical trials. Revisiting the methodological issues. *International Journal of Technology Assessment in Health Care* 1991; 7: 561–573.

123. Drummond MF, Richardson WS, O'Brien BJ. *et al.* Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 1997; 277: 1552–1557.

124. Manning WGJ. Panel on cost-effectiveness in health and medicine recommendations: identifying costs. *Journal of Clinical Psychiatry* 1999; 60(Suppl 3): 54–56.

125. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs measurement through quality of life? A response to the recommendation of the Washington Panel. *Health Economics* 1997; 6: 253–259.

126. Cantor SB. Cost-effectiveness analysis, extended dominance, and ethics: a quantitative assessment. *Medical Decision Making* 1994; 14: 259–265.

127. Cantor SB, Ganiats TG. Incremental cost-effectiveness analysis: the optimal strategy depends on the strategy set. *Journal of Clinical Epidemiology* 1999; 52: 517–522.

128. Garber AM. *Advances in cost-effectiveness analysis of health interventions.* Working Paper 7198, NBER Working Paper Series, 1999.

129. Ament A, Baltussen R. The interpretation of results of economic evaluation: explicating the value of health. *Health Economics* 1997; 6: 625–635.

130. Weinstein MC. From cost-effectiveness ratios to resource allocation: where to draw the line? In: Sloan F. (Ed.), *Valuing health care: costs, benefits, and effectiveness of pharmaceuticals and other medical technologies.* New York: Cambridge University Press, 1996, pp. 77–97.

131. Johannesson M, Meltzer D. Some reflections on cost-effectiveness analysis. *Health Economics* 1998; 7: 1–7.

132. Laska EM, Meisner M, Siegel C. The usefulness of average cost-effective ratios. *Health Economics* 1997; 6: 504.

133. Laska EM, Meisner M, Siegel C. *et al.* Ratio-based and net benefit-based approaches to health care resource allocation: proofs of optimality and equivalence. *Health Economics* 1999; 8: 174.

134. Craven PC. Treating bone and joint infections with teicoplanin: hospitalization vs outpatient cost issues. *Hospital Formula* 1993; 28(Suppl): 41–45.

135. Tubman TR, Halliday HL, Normand C. Cost of surfactant replacement treatment for severe neonatal respiratory distress syndrome: a randomised controlled trial (published erratum appears in BMJ 1991 Jan 5; 302(6767): 27). *British Medical Journal* 1990; 301: 842–845.

136. Fendrick AM, Lee JH, LaBarge C. *et al.* Clinical and economic impact of a combination Haemophilus influenzae and Hepatitis B vaccine: estimating cost-effectiveness using decision analysis. *Archives of Pediatrics and Adolescence Medicine* 1999; 153: 126–136.

137. Goldie SJ, Kuntz KM, Weinstein MC. *et al.* The clinical effectiveness and cost-effectiveness of screening for anal squamous intraepithelial lesions in

homosexual and bisexual HIV-positive men. *Journal of the American Medical Association* 1999; 281: 1822–1829.

138. Krahn M, Guasparini R, Sherman M. *et al.* Costs and cost-effectiveness of a universal, school-based hepatitis B vaccination program. *American Journal of Public Health* 1998; 88: 1638–1644.

139. O'Brien BJ, Goeree R, Hux M. *et al.* Economic evaluation of donepezil for the treatment of Alzheimer's disease in Canada. *Journal of the American Geriatric Society* 1999; 47: 570–578.

140. Murray CJ. Rethinking DALYs. In: Murray CJ, Lopez AD. (Eds), *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020.* Cambridge, MA: Harvard University Press, 1996, pp. 1–98.

141. Davis GL, Beck JR, Farrell G. *et al.* Prolonged treatment with interferon in patients with histologically mild chronic hepatitis C: a decision analysis. *Journal of Viral Hepatitis* 1998; 5: 313– 321.

142. Murray CJ, DeJonghe E, Chum HJ. *et al.* Cost effectiveness of chemotherapy for pulmonary tuberculosis in three sub-Saharan African countries. *Lancet* 1991; 338: 1305–1308.

143. Molineaux L, Storey J, Cohen JE. *et al.* A longitudinal study of human malaria in the West African Savanna in the absence of control measures: relationships between different Plasmodium species, in particular P. falciparum and P. malariae. *American Journal of Tropical Medicine and Hygiene* 1980; 29: 725–737.

144. Struchiner CJ, Halloran ME, Spielman A. Modeling malaria vaccines. I: New uses for old ideas. *Mathematics and Biosciences* 1989; 94: 87–113.

145. Koella JC. On the use of mathematical models of malaria transmission. *Acta Tropica* 1991; 49: 1–25.

146. Joesoef MR, Remington PL, Jiptoherijanto PT. Epidemiological model and cost-effectiveness analysis of tuberculosis treatment programmes in Indonesia. *International Journal of Epidemiology* 1989; 18: 174–179.

147. Kuo HS, Chang HJ, Chou P. *et al.* A Markov chain model to assess the efficacy of screening for non-insulin dependent diabetes mellitus (NIDDM). *International Journal of Epidemiology* 1999; 28: 233–240.

148. Samsa GP, Reutter RA, Parmigiani G. *et al.* Performing cost-effectiveness analysis by integrating randomized trial data with a comprehensive decision model: application to treatment of acute ischemic stroke. *Journal of Clinical Epidemiology* 1999; 52: 259–271.

149. Van-Damme P, Tormans G, Beutels P. *et al.* Hepatitis B prevention in Europe: a preliminary economic evaluation. *Vaccine* 1995; 13(Suppl): 54–57.

150. de-Koning HJ, van-Ineveld BM, van-Oortmarssen GJ. *et al.* Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *International Journal of Cancer* 1991; 49: 531–537.

151. de-Koning HJ, van-Ineveld BM, de-Haes JC. *et al.* Advanced breast cancer and its prevention by screening. *British Journal of Cancer* 1992; 65: 950–955.

152. Jönsson B, Weinstein MC. Economic evaluation alongside multinational clinical trials: study considerations for GUSTO IIb. *International Journal of Technology Assessment in Health Care* 1997; 13: 49–58.

153. Van der Werf F, Topol EJ, Lee KL. Variations in patients management and outcomes for acute myocardial infarction in the United States and other countries: results from the GUSTO trial. *Journal of the American Medical Association* 1995; 273: 1586–1591.

2 PopMod: a longitudinal population model with two interacting disease states[1]

Jeremy A. Lauer, Klaus Röhrich, Harald Wirth, Claude Charette, Steve Gribble and Christopher J.L. Murray

Abstract

This article provides a description of the population model PopMod, which is designed to simulate the health and mortality experience of an arbitrary population subjected to two interacting disease conditions as well as all other "background" causes of death and disability. Among population models with a longitudinal dimension, PopMod is unique in modelling two interacting disease conditions; among the life-table family of population models, PopMod is unique in not assuming statistical independence of the diseases of interest, as well as in modelling age and time independently. Like other multi-state models, however, PopMod takes account of "competing risk" among diseases and causes of death.

PopMod represents a new level of complexity among both generic population models and the family of multi-state life tables. While one of its intended uses is to describe the time evolution of population health for standard demographic purposes (e.g. estimates of healthy life expectancy), another prominent aim is to provide a standard measure of effectiveness for intervention and cost-effectiveness analysis. PopMod, and a set of related standard approaches to disease modelling and cost-effectiveness analysis, will facilitate disease modelling and cost-effectiveness analysis in diverse settings and help make results more comparable.

INTRODUCTION

Measuring population health has been inseparable from modelling population health for at least three hundred years. The first accurate empirically based life table—a population model, albeit a simple one—was constructed by Edmund Halley in 1693 for the population of Breslau, Germany *(1)*. However, the 1662 life table of John Graunt, while less rigorously based on empirical mortality data, represented a reasonably good approximation of life expectancy at birth in the seventeenth century *(2)*. Indeed, because of Graunt's strong a priori assumptions about age-specific mortality, his life table could be said to represent the first population model. Recently, multi-state life tables, which explicitly model several population transitions, have become a common tool for demographers, health economists and others, and a considerable body of theory has been developed for their use and interpretation *(3–5)*. Despite the substantial complexity of existing multi-state models, a recent publication has highlighted the advantages of so-called "dynamic life tables", in which age and time would be modelled independently *(6)*.

Mathematical and computational constraints are no longer serious obstacles to solving complex modelling problems, although the empirical data required for complex models are. In particular, multi-state models present data requirements that can rapidly exceed empirical knowledge about real-world parameter values, and in many cases, the input parameters for such models are therefore subject to uncertainty. Nevertheless, even with substantial uncertainty, such models can provide robust answers to interesting questions. Indeed, the work of John Graunt demonstrates the practical value of results obtained with even purely hypothetical parameter values.

PopMod, one of the standard tools of the WHO-CHOICE programme (http://www.who.int/evidence/cea), is the first published example of a multi-state dynamic life table. Like other multi-state models, PopMod takes account of "competing risk" among diseases, causes of death and possible interventions. However, PopMod represents a new level of complexity among both generic population models and the family of multi-state life tables. Among population models with a longitudinal dimension, PopMod is unique in modelling two distinct and possibly interacting disease conditions; among the life-table family of population models, PopMod is unique in not assuming statistical independence of the diseases of interest, as well as in modelling age and time independently.

While one of PopMod's intended uses is to describe the time evolution of population health for standard demographic purposes (e.g. estimates of healthy life expectancy), another prominent aim is to provide a standard measure of effectiveness for intervention and cost-effectiveness analysis. PopMod, and a related set of standard approaches to disease

modelling and cost-effectiveness analysis used in the WHO-CHOICE programme, facilitate disease modelling and cost-effectiveness analysis in diverse settings and help make results more comparable. However, the implications of a tool such as PopMod for intervention analysis and cost-effectiveness analysis is a relatively new area with little published scholarship. Most published cost-effectiveness analysis has not taken a population approach to measuring effectiveness, and when studies have done so they have generally adopted a steady-state population metric *(7)*. Relatively little published research has noted the biases of conventional approaches when used for resource allocation *(8)*.

Despite similarities in some of the mathematical techniques *(9)*, this paper does not consider transmissible disease modelling.

Basic description of the model

PopMod simulates the evolution in time of an arbitrary population subject to births, deaths and two distinct disease conditions. The model population is segregated into male and female subpopulations, in turn segmented into age groups of one-year span. The model population is truncated at 101 years of age. The population in the first group is increased by births, and all groups are depleted by deaths. Each age group is further subdivided into four distinct states representing disease status. The four states comprise the two groups with the individual disease conditions, a group with the combined condition and a group with neither of the conditions. The states are denominated for convenience X, C, XC and S, respectively. The state entirely determines health status and disease and mortality risk for its members. For example, X could be ischaemic heart disease, C cerebrovascular disease, XC the joint condition and S the absence of X or C.

State members undergo transitions from one group to another, they are born, they get sick and recover, and they die. The four groups are collectively referred to as the total population T, births are represented as

Figure I   Network diagram of the differential equations model
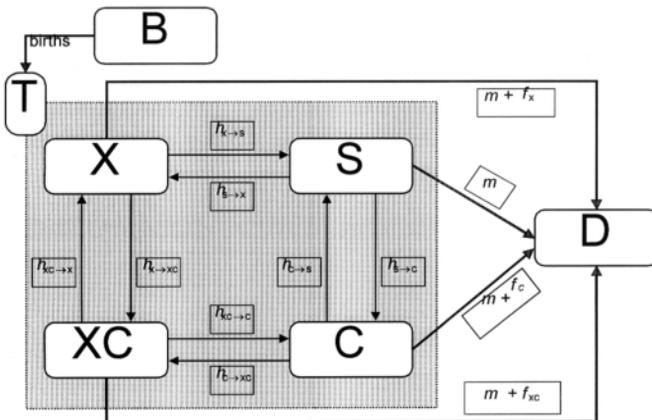
the special state B, and deaths as the special state D. A diagram for the first age group is shown in Figure 1 (notation used is explained in the section *Describing states, populations and transitions between states*). In the diagram, states are represented as boxes and flows are depicted as arrows. Basic output consists of the size of the population age-sex groups reported at yearly intervals. From this output further information is derived. Estimates of the severity of the states X, C, XC and S are required for full reporting of results, which include standard life-table measures as well as a variety of other summary measures of population health.

There now follows a more technical description of the model and its components, broken down into the following sections: describing states, populations and transitions between states; disease interactions; modelling mechanics; and output interpretation. The article concludes with a discussion of the relation of PopMod to other modelling strategies, plus a consideration of the implications, advantages and limitations of the approach.

DESCRIBING STATES, POPULATIONS AND TRANSITIONS BETWEEN STATES

DESCRIBING STATES AND POPULATIONS

In the full population model depicted in Figure 1, six age-and-sex specific states (X, C, XC, S, B and D) are distinguished. However, births B and deaths D are special states in the sense that they only feed into or absorb from other states (while the states X, C, XC and S both feed into and absorb from other states). Special states are not treated systematically in the following, which focuses on the "reduced form" of the model consisting of the states X, C, XC, and S.

States are not distinguished from their members; thus, "X" is used to mean alternatively "disease X" or "the population group with disease

Figure 2   A schematic for describing observed populations
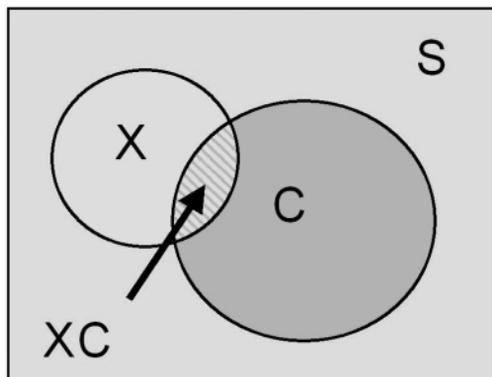
X", according to context. The second meaning is equivalent to the prevalence count for the population group.

For the differential equation system, states/groups are always denoted in the strict sense: "X" means "state X only" or "the population group with only X". However, in deriving input parameters (described more fully below in the section *Disease interactions*) from observed populations, it is convenient to describe groups in a way that allows for the possibility of "overlap". For example in Figure 2, the area "X" might be understood to mean either "the population group with X including those members with C as well" (i.e. the entire circle X) or the "the population group with only X" (i.e. the circle minus the region overlapping with circle C).

Table 1   Alternative ways to describe populations

Logical expression	Meaning	Differential equations expression
~X~C	Population group with neither X nor C	S
X~C	Population group with X but not C, i.e. with X only	X
~XC	Population group with C but not X, i.e. with C only	C
~X	Population group without X	S + C
~C	Population group without C	S + X
X	Total population group with X	X + XC
C	Total population group with C	C + XC
S	Susceptible population	S
XC	Population with both X and C	XC
T	Total population	T

Since these two valid meanings imply different uses of notation, the following conventions are adopted:

- The differential equations expressions X, C, XC and S refer only to disjoint states (or groups).
- The logical operator "~"means "not", thus "~X" is the state "not X" (or "the group without X").
- The logical expressions denoted in the left-hand column of Table 1 have the meaning and alternative description indicated in the two right-hand columns.

Prevalence rates *(p)* describe populations (i.e. prevalence counts) as a proportion of the total, for example:

$$p_X = X/T, p_C = C/T, p_{XC} = XC/T, p_S = S/T. \tag{1}$$

Here, prevalence is presented in terms of the disjoint populations X, C and XC, and the notation from the right-hand column of Table 1 is used. In the section *Disease interactions,* we discuss the case of overlapping populations.

A prevalence rate is always interpretable as a probability, but a probability is not always interpretable as a prevalence. The lower-case Greek letter pi (π) is used throughout this chapter to denote probability. Probabilities can be used to describe populations as noted in Table 2.

Table 2   Probability of finding members of population groups in
                   PopMod

Symbol	Description
ϕ_X	Probability of finding a member of T that is a member of X with random selection
ϕ_C	Probability of finding a member of T that is a member of C with random selection
ϕ_{XC}	Probability of finding a member of T that is a member of XC with random selection

Describing transitions between states

In the differential equation system, transitions (i.e. flows) between population groups are modelled as instantaneous rates, represented in Figure 1 as labelled arrows. Instantaneous rates are frequently called hazard rates, a usage generally adopted here (demographers tend to refer to instantaneous rates as "hazards" or as "forces"—e.g. force of mortality—although epidemiologists commonly use the term "rate" with the same meaning). A transition hazard is labelled here h, frequently with subscript arrows denoting the specific state transition.

In PopMod terminology, the transitions X→D, C→D and XC→D are partitioned into two parts, one of which is the cause-specific fatality hazard f due to the condition X, C or XC, and the other which is the non-specific death hazard (due to all other causes), called background mortality m:

$$h_{X \to D} = f_X + m \qquad (2a)$$

$$h_{C \to D} = f_C + m \qquad (2b)$$

$$h_{XC \to D} = f_{XC} + m \qquad (2c) \qquad\qquad (2)$$

$$h_{S \to D} = m. \qquad (2d)$$

PopMod consequently allows for up to twelve exogenous hazard parameters (Table 3).

Transition hazards

A time-varying transition hazard is denoted $h(t)$. The hazard expresses the proportion of the at-risk population (dP/P) experiencing a transition event (i.e. exiting the population) during an infinitesimal time dt:

$$h(t) = 4(1/P) \cdot dP/dt. \qquad (3)$$

"Instantaneous rate" means the transition rate obtaining during the infinitesimal interval dt, that is, during the instant in time t. If an

Table 3 Transition hazards in the population model

Hazard	Description	State transition
$h_{S\Downarrow X}$	incidence hazard	S$\Downarrow$ X
$h_{X\Downarrow S}$	remission hazard	X$\Downarrow$ S
$h_{S\Downarrow C}$	incidence hazard	S$\Downarrow$ C
$h_{C\Downarrow S}$	remission hazard	C$\Downarrow$ S
$h_{X\Downarrow D}$	case fatality hazard	X$\Downarrow$ D
$h_{C\Downarrow D}$	case fatality hazard	C$\Downarrow$ D
$h_{XC\Downarrow D}$	case fatality hazard	XC$\Downarrow$ D
$h_{T\Downarrow D}$	background mortality hazard	T$\Downarrow$ D
$h_{C\Downarrow XC}$	incidence hazard	C$\Downarrow$ XC
$h_{XC\Downarrow C}$	remission hazard	XC$\Downarrow$ C
$h_{X\Downarrow XC}$	incidence hazard	X$\Downarrow$ XC
$h_{XC\Downarrow X}$	remission hazard	XC$\Downarrow$ X

instantaneous rate does not vary, or its small fluctuations are immaterial to the analysis, PopMod parameters can be interpreted as average hazards without prejudice to the model assumptions.

Average hazards can be approximated by counting events ΔP during a period Δt and dividing by the population time at risk. If for practical purposes the instantaneous rate does not change within the time span, the approximate average hazard can be used as an estimate for the underlying instantaneous rate:

$$-(1/P) \cdot dP/dt \approx -\int dP / \int P dt \approx -\Delta P / (P \cdot \Delta t), \qquad (4)$$

where $\Delta P = \int dP$ is the cumulative number of events occurring during the interval Δt, and $\int P dt \approx P \cdot \Delta t$ is the corresponding population time at risk. Time at risk is approximated by multiplying the mid-interval population (P) by the length of the interval Δt.

For example, if ten deaths due to disease X ($\Delta P = 10$) occur in a population with approximately one million years of time at risk ($P \cdot \Delta t = 1,000,000$), an approximation of the instantaneous rate $h_{X \to D}(t)$ is given by:

$$h_{X\Downarrow D}(t) - \div P / P \cdot \div t = 10 / 1,000,000 = 0.00001 . \qquad (5)$$

Note that while eq. (3) and eq. (4) are equivalent in the limit where $\Delta t \to 0$, the approximation in eq. (4) will result in large errors when rates are high. This is discussed in the section *Proportions and hazard rates*, and an alternative formula for deducing average hazard is proposed in eq. (9).

The quantity in eq. (4) has units "deaths per year at risk", and is often called a "cause-specific mortality hazard". For the same population and

deaths, but restricting attention to the group with disease X (where, for example, $P \cdot \Delta t = 10{,}000$) the calculated hazard will be larger:

$$h_{X \to D}(t) \approx \Delta P / P \cdot \Delta t = 10 / 10{,}000 = 0.001. \tag{6}$$

The quantity in eq. (6) has the same units as that in eq. (5), but is a "case fatality hazard". Note that the same transition events (e.g. "dying of disease X") can be used to define different hazard rates depending on which population group is considered.

Proportions and hazard rates

Integration by parts of eq. (3) shows that the proportion of the population experiencing the transition in the time interval Δt (i.e. the "incident proportion") is given by:

$$\frac{\div P}{P(t_0)} \mid 14 \exp \int_{t_0}^{t_0 2 \div t} 4\, h(t)\, \mathrm{d}\,t \tag{7}$$

If the hazard is constant, that is, if $h(t) = h(t_0)h$, $\int dt = \Delta t$, and the integral collapses. The incident proportion is then written:

$$\frac{\div P}{P} \mid 14\, \mathrm{e}^{4h\,\mathord{\mid}\cdot t}. \tag{8}$$

The incident proportion can always be interpreted as the average probability that an individual in the population will experience the transition event during the interval (e.g. for mortality, this probability can be written $\pi_{P \to D} = \Delta P/P$). The qualification "average" is dropped if individuals in P are homogeneous with respect to transition risk during the interval.

Even if the hazard is not constant, eq. (8) can be rearranged to give an alternative (exact) formula for calculating the equivalent constant hazard h yielding ΔP transitions in the interval Δt:

$$h \mid 4 \ln \substack{\circledR \\ \mathrm{TI}} 4 \frac{\div P}{P} \Big\} / \div t. \tag{9}$$

However, if the true hazard is constant during the interval, the "equivalent constant hazard" equals the "average hazard" and the "instantaneous rate". The same identity applies when fluctuations in the underlying hazard are of no practical importance. PopMod requires the

assumption that hazards are constant within the unit of its standard reporting interval, defined by convention as one year.

Note that series expansion of $\exp\{- h \cdot \Delta t\}$ or $\ln\{1 - \Delta P/P\}$ shows that, for values of $h \cdot \Delta t \ll 1$ and $\Delta P/P \ll 1$, the equivalent constant hazard is well approximated by the time-normalized incident proportion, and vice versa, as in eq. (4):

$$h \approx \frac{1}{\Delta t}\frac{\Delta P}{P}. \tag{10}$$

Case-fatality hazards

Case-fatality hazards f_X, f_C, and f_{XC} are defined with respect to the specific populations X, C and XC, respectively:

$$f_X = -\frac{1}{\Delta t}\ln\left\{1 - \frac{\Delta X}{X}\right\}, \tag{11}$$

$$f_C = -\frac{1}{\Delta t}\ln\left\{1 - \frac{\Delta C}{C}\right\}, \tag{12}$$

$$f_{XC} = -\frac{1}{\Delta t}\ln\left\{1 - \frac{\Delta XC}{XC}\right\}. \tag{13}$$

Mortality hazards

Mortality hazards are defined with respect to the entire population, where cause-specific mortality hazards are conditional on cause of death:

$$m_{tot} = -\frac{1}{\Delta t}\ln\left\{1 - \frac{\Delta T}{T}\right\}, \tag{14}$$

$$m_X = -\frac{1}{\Delta t}\ln\left\{1 - \frac{\Delta T}{T}\right\}\Bigg|_{T \propto \Delta X \Downarrow D}, \tag{15}$$

$$m_C = -\frac{1}{\Delta t}\ln\left\{1 - \frac{\Delta T}{T}\right\}\Bigg|_{T \propto \Delta C \Downarrow D}. \tag{16}$$

The background mortality hazard m is defined as the instantaneous rate of deaths due to causes other than X or C.

Disease interactions

PopMod is typically used to simulate the evolution of a population subject to two disease conditions, where health status, health risk and mortality risk are conditional on disease state. Health status, health risk and mortality risk are plausibly conditional on disease state when the two primary disease conditions X and C interact. Such interactions can be analysed from various perspectives, for example, common risk factors, common treatments, common prognosis; however, the primary perspective adopted here for the pupose of analysis is that of "common prognosis", by which is meant that the two conditions mutually influence prevalence, incidence, remission and mortality risk.

A previously cited example was that of ischaemic heart disease (X) and cerebrovascular disease (C): it is well known that individuals with either heart disease or stroke history have lower health status and higher mortality risk than individuals with neither of these conditions, and that individuals with heart disease are at increased risk for stroke and vice versa.

Furthermore, individuals with history of both heart disease and stroke (XC) are known to have higher mortality risk and lower health status than either individuals with only one of the disease histories or those with neither. However, in this example as in many others, information about the joint condition (heart disease and stroke) is scarce relative to information about the two individual conditions (heart disease or stroke). The obvious reason for this is that the population group with the joint condition is smaller in size and has a lower life expectancy, reducing opportunities for data collection.

The presimulation problem

One of PopMod's guiding principles, therefore, is that while an analyst has access to information about basic parameter values for the conditions X and C (i.e. prevalence rates and incidence, remission and either case-fatality or cause-specific mortality hazards), the same is not generally true for the joint condition XC. Thus, more or less by construction, the modelling situation is one in which data for the joint condition are scarce or unavailable, and must consequently be derived from data known for the individual conditions.

An important implication is that the data available for the individual conditions (X and C) will be reported in terms of overlapping populations. Where specifically noted, therefore, the notation in the left-hand column of Table 1 (Logical expressions) is used in the following, with the particular implication that "X", for example, means "the population group with X including those members with C as well" (i.e. "X + XC" in differential equations terminology).

Once parameter values for the joint condition are determined, the minimum set of parameters required for population simulation are

known. This parameter-value problem—referred to here as the presimulation problem, since its solution must precede population simulation per se—can be divided into two principal parts: one concerning the prevalence rates defining the intial conditions (stocks) of the differential equations system, and the other the transition hazards defining its flows. These stocks and flows together make up the initial scenario of the population model. A cross-sectional approach is adopted in which deriving these two kinds of parameters values for the initial scenario are treated as separate problems.

The analytics of these derivations largely depend on which of a range of possible assumptions is made about the interactions of the two principal conditions. The simplest possible assumption is essentially an assumption of non-interaction (statistical independence). Since an understanding of the non-interacting case is an essential starting point for more complex interactions, it is discussed first.

THE INDEPENDENCE ASSUMPTION

Prevalence for the joint group

When conditions X and C are statistically independent, the joint prevalence is the product of the individual (marginal) prevalences:

$$p_{XC} = p_X \cdot p_C. \tag{17}$$

Transition hazards for the joint group

Independence implies that the hazards for the group with X or C are equal to the corresponding hazards for the group without X or C (in eq. (18) populations are denoted in differential equations (disjoint) notation from the right-hand column of Table 1):

$$h_{XC\Downarrow C} = h_{X\Downarrow S}$$

$$h_{XC\Downarrow X} = h_{C\Downarrow S}$$

$$h_{C\Downarrow XC} = h_{S\Downarrow X} \tag{18}$$

$$h_{X\Downarrow XC} = h_{S\Downarrow C}$$

Joint case fatality hazard

The probabilities $\phi_{X\infty^X\Downarrow D}$ and $\phi_{C\infty^C\Downarrow D}$ for an individual in group X or C to die of cause X or C, respectively, during an interval Δt are:

$$\phi_{X \propto X \Downarrow D} = (1 - e^{-f_X t}) = \left.\frac{-\dot X}{X}\right|_{X \propto X \Downarrow D} \quad \text{and} \quad \phi_{C \propto C \Downarrow D} = (1 - e^{-f_C t}) = \left.\frac{-\dot C}{C}\right|_{C \propto C \Downarrow D}. \quad (19)$$

So the joint probability $\phi_{XC \propto X\text{ or }C \Downarrow D}$ for someone in the group XC dying of either X or C is given by the laws of probability:

$$
\begin{aligned}
\phi_{XC \propto X\text{ or }C \Downarrow D} &= \phi_{X\propto X\Downarrow D} + \phi_{C\propto C\Downarrow D} - (\phi_{X\propto X\Downarrow D})(\phi_{C\propto C\Downarrow D}) \\
&= (1-e^{-f_X t}) + (1-e^{-f_C t}) - (1-e^{-f_X t})(1-e^{-f_C t}) \\
&= 1 - e^{-f_X t}e^{-f_C t} \\
&= 1 - e^{-(f_X + f_C)t} \\
&\equiv 1 - e^{-f_{XC}t}
\end{aligned}
\quad (20)
$$

Although individuals in the joint group XC are at risk of death from either X or C, or from other causes, the probability framework requires the assumption that they do not die of simultaneous causes (i.e. there is no cause of death "XC").

The combined case-fatality rate f_{XC} is thus:

$$f_{XC} = f_X + f_C. \quad (21)$$

This simple addition rule can be generalized to situations with more than two independent causes of death.

Background mortality hazard

The "background mortality hazard" m expresses mortality risk for population T due to any cause of death other than X and C. The "independence assumption" claims m is independent of these causes, in other words, that m acts equally on all groups (in eqs. (22–25) populations are denoted in differential equations notation from the right-hand column of Table 1):

$$m\dot T = m(\dot S + \dot X + \dot C + \dot{XC}) = m\dot S + m\dot X + m\dot C + m\dot{XC}. \quad (22)$$

The total ("all cause" or "crude") death hazard for the population is written m_{tot}. The following identity expresses the constraint that deaths in population T equal the sum of deaths in populations S, X, C and XC:

$$m_{\text{tot}}\dot T = m\dot S + (m + f_X)\dot X + (m + f_C)\dot C + (m + f_{XC})\dot{XC}. \quad (23)$$

Thus:

$$m_{tot}\ (T) = m\ ((S \cdot X \cdot C \cdot XC) \cdot f_X\ (X \cdot f_C\ (C \cdot f_{XC}\ (XC$$
$$= m\ (T \cdot f_X\ (X \cdot f_C\ (C \cdot (f_X \cdot f_C)\ (XC \qquad (24)$$
$$= m\ (T \cdot f_X\ ((X \cdot XC) \cdot f_C\ ((C \cdot XC).$$

Since by definition group X or C contributes no deaths due to cause C or X, respectively:

$$f_C\ ((C \cdot XC) = m_C\ (T, \qquad (25)$$

$$f_X\ ((X \cdot XC) = m_X\ (T, \qquad (26)$$

so:

$$m_{tot}\ (T) = m\ (T \cdot m_X\ (T \cdot m_C\ (T.$$

and:

$$m = m_{tot} - m_X - m_C. \qquad (27)$$

Likewise, this rule is generalizable to scenarios with more than three (m, X, C) independent causes of death.

Relaxing the independence assumption

As noted in the introduction, one of the primary reasons for the introduction of PopMod was to model disease interactions in a longitudinal population model. Modelling interactions requires relaxing the assumption of independence.

In the presimulation of the "stocks and flows" required for the initial scenario, three areas of interaction for the health states X and C can be distinguished. Having X (C) may make it more or less likely to:

(1) have C (X),
(2) acquire or recover from C (X),
(3) die from C (X).

Note that while interaction (1) could alternatively be considered the cumulative result of interactions (2) and (3) in the past, this is not the approach adopted here.

Interaction (1): Prevalence of the joint group

In this and subsequent sections except where noted, we revert to the notation from the left-hand column of Table 1. Table 4 shows six possible cases for calculating prevalence of the joint group depending on the type

Table 4   Options for calculating overlap probability π_{XC}

Case	π_{XC} calculated as	Comment
1	$\phi_C \cdot \phi_X$	C and X are independent
2	$\phi_{C|X} \cdot \phi_X$	C and X interact and $\phi_{C|X}$ or $\phi_{X|C}$ is known
3	$\phi_{X|C} \cdot \phi_C$	
4	$\phi_C \cdot \phi_X / [\phi_C + (1 - \phi_C) / RR_{X|C}]$	C and X are dependent and the relative risk $RR_{X|C}$ or $RR_{C|X}$ is known
5	$\phi_X \cdot \phi_C / [\phi_X + (1 - \phi_X) / RR_{C|X}]$	
6	$\phi_X \cdot \phi_C \cdot k$	X (C) either potentiates, or protects from, C (X)

of information known about the disease interaction. The probability notation π is used for prevalence, where $\pi_{X|C}$ is the probability of having disease X among those who have disease C and π_X and π_C are short forms for $\pi_{X|T}$ and $\pi_{C|T}$. Relative risk (RR) is defined here as a ratio of probabilities (risk ratio), for example, $RR_{C|X} = \pi_{C|X} / \pi_{C|\sim X}$ is the probability of having X if C is present over the probability of having X if C is not present.

Calculations for case 1 follow directly from the assumption of independence. Cases 2 and 3 follow directly from the definition of conditional probability. Cases 4 and 5 are derived as follows. Since the probability of belonging to the joint group is independent of which disease group is conditioned on, it is clear that:

$$\phi_{XC} = \phi_{X|C} \cdot \phi_C = \phi_{C|X} \cdot \phi_X. \tag{28}$$

Using the definition of conditional probability, we write:

$$\pi_X = \pi_{X|C} \cdot \pi_C + \pi_{X|\sim C} \cdot \pi_{\sim C}, \text{and}$$
$$\pi_C = \pi_{C|X} \cdot \pi_X + \pi_{C|\sim X} \cdot \pi_{\sim X}. \tag{29}$$

Now supposing $RR_{X|C}$ or $RR_{C|X}$ is known, solving either for $\pi_{X|C}$ or $\pi_{C|X}$ and substituting the result into eq. (29) and solving again for $\pi_{X|C}$ and $\pi_{C|X}$ yields:

$$\phi_{X|C} = \phi_X / [\phi_C + \phi_{\sim C} / RR_{X|C}], \text{and}$$
$$\phi_{C|X} = \phi_C / [\phi_X + \phi_{\sim X} / RR_{C|X}] \tag{30}$$

So again using the definition of conditional probability:

$$\phi_{XC} = \phi_X \cdot \phi_C / [\phi_C + \phi_{\sim C} / RR_{X|C}], \text{and}$$
$$\phi_{XC} = \phi_C \cdot \phi_X / [\phi_X + \phi_{\sim X} / RR_{C|X}] \tag{31}$$

Recalling $1 - \pi_X = \pi_{\sim X}$ and $1 - \pi_C = \pi_{\sim C}$, the required expressions in Table 4 are obtained.

The factor k in case 6 is an arbitrary multiplier that increases or reduces the prevalence of group XC compared to what would be obtained under independence, and lies between 0 and 1 if having one disease reduces the probability of having the other, and between 1 and MAX($1/\pi_C$, $1/\pi_X$) if having one disease makes it more likely to have the other. Upper bounds on k are easy to derive using the fact that $\pi_{XC} = \pi_X = \pi_C$ when X and C are obligate symbiotes.

The six cases span a range of information availability about interaction of X and C on the prevalence of the joint condition:

- Case 1 assumes independence (no interaction).
- Case 2 and 3 assume conditional prevalence is known.
- Case 4 and 5 assume relative risk is known.
- Case 6 assumes a potentiation (or protection) factor can be defined.

Interaction (2): Incidence and remission for the joint group

For incidence hazard, we write i and for remission hazard, r. Consistent with "overlapping populations", unless specifically noted, hazards are understood as "total hazards", that is, i_X includes all incidence to X regardless of whether C is also present in the population at risk. Conditional hazards are denoted $i_{X|\sim C}$ or $i_{X|C}$ to signify "incidence to X in the group without C" and "incidence to X in the group with C", respectively.

Consider total incidence i_X for the initial scenario. The product of total incidence to X and the total population without X ($\sim$X) must be equal to the sum of the products of the conditional incidences ($i_{X|\sim C}$, $i_{X|C}$) and the conditional populations ($\sim$X$\sim$C, $\sim$XC):

$$i_X \,(\sim X) = i_{X|\sim C}\,(\sim X \sim C) + i_{X|C}\,(\sim XC). \qquad (32)$$

Dividing by total population T yields:

$$i_X \frac{(\sim X)}{(T)} = i_{X|\sim C} \frac{(\sim X \sim C)}{(T)} + i_{X|C} \frac{(\sim XC)}{(T)}, \qquad (33)$$

and replacing population ratios by the corresponding prevalence rates yields:

$$i_X \, \phi_{\sim X} = i_{X|\sim C} \, \phi_{\sim X \sim C} + i_{X|C} \, \phi_{\sim XC}. \qquad (34)$$

Dividing both sides by $\pi_{\sim X}$ yields the following expression for i_X:

$$i_X \mid i_{X|\sim C} \frac{\phi_{\sim X \sim C}}{\phi_{\sim X}} 2 i_{X|C} \frac{\phi_{\sim XC}}{\phi_{\sim X}}, \tag{35}$$

where:

$$\phi_{\sim X} \mid \phi_{\sim X \sim C} 2 \phi_{\sim XC}. \tag{36}$$

It is therefore clear that total incidence to X is a weighted average of the conditional incidences, where the weights are the proportions of the population without X partitioned according to C status.

Recall that, in terms of the differential equations notation from the right-hand column of Table 1, $\pi_{\sim X} = \pi_C + \pi_S$, $\pi_{\sim X \sim C} = \pi_S$ and $\pi_{\sim XC} = \pi_C$, the values of which are determined according to one of the six cases defined above in interaction (1). Thus, when total hazard i_X is known, eq. (34) has only two unknowns ($i_{X|\sim C}$ and $i_{X|C}$). Clearly, if information on one or both conditional hazards is available, interaction (2) with respect to i_X is fully characterized for the initial scenario.

However, the guiding principle of the presimulation problem was that information on the non-overlapping populations (e.g. direct observation of the conditional hazards) is relatively scarce. When this is true, the unknown conditional hazards must remain undetermined unless one of the following three rate ratios (RR) is known or can be approximated:

$$RR(i_X)_1 \mid \frac{i_{X|C}}{i_{X|\sim C}}, RR(i_X)_2 \mid \frac{i_{X|\sim C}}{i_X}, \text{ or } RR(i_X)_3 \mid \frac{i_{X|C}}{i_X}. \tag{37}$$

A similar situation applies to the total hazards i_C, r_C, and r_X for the initial scenario, that is, eq. (34) is one of a family of equations representing the relation between the total disease hazards and the corresponding conditional hazards for subpopulations:

$$\begin{aligned}
i_X \, \phi_{\sim X} &\mid i_{X|\sim C} \, \phi_{\sim X \sim C} \, 2 \, i_{X|C} \, \phi_{\sim XC} \\
i_C \, \phi_{\sim C} &\mid i_{C|\sim X} \, \phi_{\sim X \sim C} \, 2 \, i_{C|X} \, \phi_{X \sim C} \\
r_X \, \phi_X &\mid r_{X|\sim C} \, \phi_{X \sim C} \, 2 \, r_{X|C} \, \phi_{XC} \\
r_C \, \phi_C &\mid r_{C|\sim X} \, \phi_{\sim XC} \, 2 \, r_{C|X} \, \phi_{XC}.
\end{aligned} \tag{38}$$

Note that, with respect to the initial scenario, eq. (38) forms a simultaneous system with eq. (31)—or one of the other methods of calculating π_{XC} noted in Table 4—and the system has a unique numerical solution whenever enough parameter values are known, that is, assuming the four total hazards are known, if one of the three following rate ratios (or its inverse) is known for each hazard:

$$RR(i_X)_1 = \frac{i_{X|C}}{i_{X|\sim C}},\ RR(i_X)_2 = \frac{i_{X|\sim C}}{i_X},\ \text{or } RR(i_X)_3 = \frac{i_{X|C}}{i_X},\ \text{and}$$

$$RR(i_C)_1 = \frac{i_{C|X}}{i_{C|\sim X}},\ RR(i_C)_2 = \frac{i_{C|\sim X}}{i_C},\ \text{or } RR(i_C)_3 = \frac{i_{C|X}}{i_C},\ \text{and}$$

$$RR(r_X)_1 = \frac{r_{X|C}}{r_{X|\sim C}},\ RR(r_X)_2 = \frac{r_{X|\sim C}}{r_X},\ \text{or } RR(r_X)_3 = \frac{r_{X|C}}{r_X},\ \text{and} \tag{39}$$

$$RR(r_C)_1 = \frac{r_{C|X}}{r_{C|\sim X}},\ RR(r_C)_2 = \frac{r_{C|\sim X}}{r_C},\ \text{or } RR(r_C)_3 = \frac{r_{C|X}}{r_C}.$$

Interaction (3): Mortality for the joint group

This interaction concerns causes of death. We assume that the all-cause mortality hazard m_{tot} and the total (i.e. overlapping) case-fatality hazards f_X and f_C are known. It follows that:

$$f_X\,\phi_X = f_{X|\sim C}\,\phi_{X\sim C} + f_{X|C}\,\phi_{XC},\ \text{and}$$
$$f_C\,\phi_C = f_{C|\sim X}\,\phi_{\sim XC} + f_{C|X}\,\phi_{XC}. \tag{40}$$

Following a derivation similar to that in eqs (19) – (21), one can show that, given total case-fatality hazards f_X and f_C, the case-fatality hazard for the joint condition is the sum of the conditional hazards:

$$f_{XC} = f_{X|C} + f_{C|X} \tag{41}$$

Further, since:

$$m_X\,T = f_{X|\sim C}\,(X \sim C) + f_{X|C}\,(XC),\ \text{and}$$
$$m_C\,T = f_{C|\sim X}\,(\sim XC) + f_{C|X}\,(XC), \tag{42}$$

so:

$$m_X = f_{X|\sim C}\,\phi_{X\sim C} + f_{X|C}\,\phi_{XC}, \tag{43}$$

and:

$$m_C = f_{C|\sim X}\,\phi_{\sim XC} + f_{C|X}\,\phi_{XC}. \tag{44}$$

In other words, the cause-specific mortality hazards are weighted averages of the conditional case-fatality hazards, where weights are the proportions of the total population according to disease status regarding the other condition.

It remains true that:

$$m \mid m_{\text{tot}} \; 4 \; m_X \; 4 \; m_C, \tag{45}$$

as in eq. (27).

Other interactions

Another interaction might involve relaxing the assumption of independence between background mortality hazard m and case-fatality hazards f_X and f_C. However, in cases where such dependence is suspected or known, it may be possible to "work around" it by choosing appropriate definitions for X and C. For example, to take the ischaemic heart disease (X) and stroke (C) example, suppose it is important for the research question to account for the fact that individuals with X or C are also at increased risk of mortality from other selected causes of death such as cardiac failure. While one approach might be to introduce a new box for cardiac failure, within the current structure of PopMod, the onus is effectively on the analyst to take into account such increased risk of background mortality by modifying the way state C is defined and by adjusting the corresponding incidence and case-fatality rates. For example, state C could be defined as "stroke and all other conditions (including cardiac failure) at increased risk due to heart disease". Another type of exception to the general rule of independence between background mortality and cause-specific mortality would be the existence of any common causal modifiers of m, f_X and f_C, for example, the allocation of health-care expenditure.

MODELLING MECHANICS

Initial conditions

PopMod describes population evolution conditional on initial conditions that define the state of the system at some initial time. These initial conditions consist of the population distribution in non-overlapping terms. If potentially overlapping populations (i.e. descriptions from the left hand side of Table 1) are considered, when the total prevalences p_X and p_C are known the non-overlapping population distribution can be fully determined by determining the prevalence of the joint group. Methods for this are discussed in the section *Disease interactions*.

Runge-Kutta method

The differential equation system is determined by its initial conditions and its parameters. An algebraic description of PopMod differential equation system—using notation from the right-hand side of Table 1—is:

$$dS/dt = 4(h_{S\Downarrow X} + h_{S\Downarrow C} + h_{S\Downarrow D})\cdot S + (h_{X\Downarrow S})\cdot X + (h_{C\Downarrow S})\cdot C \quad (46a)$$

$$dX/dt = 4(h_{X\Downarrow S} + h_{X\Downarrow XC} + h_{X\Downarrow D})\cdot X + (h_{S\Downarrow X})\cdot S + (h_{XC\Downarrow X})\cdot XC \quad (46b)$$

$$dC/dt = 4(h_{C\Downarrow X} + h_{C\Downarrow XC} + h_{C\Downarrow D})\cdot C + (h_{S\Downarrow C})\cdot S + (h_{XC\Downarrow C})\cdot XC \quad (46c) \quad (46)$$

$$dXC/dt = 4(h_{XC\Downarrow X} + h_{XC\Downarrow C} + h_{XC\Downarrow D})\cdot XC + (h_{X\Downarrow XC})\cdot X + (h_{C\Downarrow XC})\cdot C \quad (46d)$$

$$dD/dt = (h_{S\Downarrow D})\cdot S+(h_{X\Downarrow D})\cdot X + (h_{C\Downarrow D})\cdot C + (h_{XC\Downarrow D})\cdot XC \quad (46e)$$

Under specified conditions, which apply here, such a differential equation system has a unique solution, and the solution can be expressed in terms of the eigenvalues and eigenvectors of the 5×5 coefficient matrix (*10*).

Since finding the required eigenvalues and eigenvectors is here equivalent to solving a fifth-degree polynomial equation, specialized solution algorithms—and access to a substantial amount of processor time—will generally be required. An attractive alternative is therefore the use of numerical techniques, since they yield solutions more cheaply, and without requiring custom routines.

In PopMod, the evolution of the population in time is approximated by a 4th-order Runge-Kutta method, or, optionally, by a 5th-order Runge-Kutta method (*10*). The relevant time step is defined as a fraction of the standard reporting interval (the number of divisions of the basic reporting interval must in principle be divisible by 3, but to allow for the possibility of starting with mid-year values in the first year, the number of divisions must be divisible by 6 and the minimum number of divisions is fixed at 12). Note that an nth-order numerical method will in general provide useful results so long as the differentials are smaller than n^{th}-order in the chosen time step.

Each population age- and sex group is modelled as a separate system, and age is updated by taking end-of-year solution values for the "age = α" system as the initial values for the "age = α + 1" system in the subsequent model year.

A 4th-order Runge-Kutta method provides solutions to differential equations of the type:

$$dy_i(x)/dx = f_i(x, y_i(x)), \quad (47)$$

and is defined by the ansatz (Euler method) that:

$$y_i(x + \div x) = y_i (x) + \div x \cdot f_i (x, y_i(x)), \quad (48)$$

where:

$$y_i(x + \div x) = y_i(x) + (k_{1i} + 2k_{2i} + 2k_{3i} + k_{4i})/6 + O(\div x^5), \text{ and}$$

$$k_{1i} = \div x \cdot f_i(x, y_i)$$

$$k_{2i} = \div x \cdot f_i(x + \div x/2, y_i + k_{1i}/2) \qquad (49)$$

$$k_{3i} = \div x \cdot f_i(x + \div x/2, y_i + k_{2i}/2)$$

$$k_{4i} = \div x \cdot f_i(x + dx, y_i + k_{3i}).$$

Note that here $x = t$, y_i = S, X, C, XC and D and that the differential equations (46a–46e) are not explicitly time dependent, that is, $f_i(t, y_i(t))$ = $f_i(y_i(t))$ *(10)*.

OUTPUT INTERPRETATION

Standard PopMod output reports $P(t)$ for each population group as end-of-interval (e.g. year-end) values, corresponding to the standard life table quantity l_x. An important derived quantity also included in output is the time at risk experienced by the group during the interval ($\int P(t)\, dt$), corresponding to the life table quantity L_x (sometimes called "life-years" or "person-years").

For a constant population, population time at risk is calculated $P \cdot \Delta t$. For PopMod populations, population time at risk for the interval b–a is calculated:

$$P_{LY} \mid \int_a^b P/t \Box dt, \text{ for example } X_{LY} \mid \int_a^b X/t \Box dt . \qquad (50)$$

When the quantity resulting from eq. (50) with units "person-years" is divided by the length of the time interval with units "years", average population size for the interval ($\hat{P}$, with units "persons") is obtained:

$$\hat{P} \mid \frac{1}{b \, 4 \, a} \int_a^b P(t)\, dt . \qquad (51)$$

$\hat{P}$ thus conforms to the definition of the expected value of the function $P(t)$ on the interval b–a. Since b–a is by convention one year (or "chronon" etc.), the normalization to the interval b–a means dividing by 1. Thus, since in this case the numerical quantity is unchanged, substituting different reporting units yields two equally valid interpretations for the same output:

(1) the average population size $\hat{P}$ = E[$P(t)$] during the interval Δt, or

(2) the population time at risk P_{LY} experienced during the interval Δt.

Interpretation (*1*) also corresponds to average (count) prevalence for the population.

When transition rates are "small" (i.e. the differentials are approximately linear), average population can be interpreted as mid-interval population. Under the same assumptions, mid-year population provides a good estimate of population time at risk.

PopMod numerically evaluates P_{LY} with a standard Newton-Cotes formula for 4-point closed quadrature, sometimes also called Simpson's 3/8-rule (*11*). The quadrature formula relies on the values of $P(t)$ determined by the Runge-Kutta method at multiples of the chosen time step. Since these values involve numerical estimation error, there is no simple expression for the order of accuracy of the different output values reported in PopMod (*10*).

Discussion

Advantages of the approach

PopMod combines features of existing models (see below) with the possibility to analyse several disease states. It explicitly analyses time evolution and, even more importantly, abandons the constraint of independence of disease states.

A primary advantage of the approach adopted in PopMod is the separate modelling of age and time, and the type of bias inherent in models that do not do so has been previously pointed out (*7*). Moreover, it has been independently noted that, without this feature, life-table measures are constrained to adopt—somewhat artificially—either a "period" or a "cohort" perspective (*6*). The other chief advantage of PopMod is the ability to deal with heterogeneity of disease and mortality risk by modelling up to four disease states. No previously published generic population model has combined both these features. Note, however, that if disease conditions are independent, and population-dependent effects are not of interest, a multi-state life-table approach should probably be adopted (*12*).

A further advantage of PopMod is the introduction of a systematic analytical approach to the modelling of disease interactions. This by itself represents a relatively important advance, as modellers have until now been constrained to model only independent conditions. Furthermore, in spite of the increased informational demands made by a four-state system, the modelled functional dependency between X-related hazards conditional on C status, and vice versa, reduces the number of exogenous hazards that need to be directly observed. This is of substantial practical

importance, since, while direct observation of conditional hazards usually requires a cohort study, it will often be possible to obtain estimates of the required rate ratios from more common case-control studies (*13*).

Related models

In addition to the multi-state life table family (*3;4*), two additional families of mathematical models have some similarity to PopMod. One family comprises the class of models sometimes called incidence, prevalence and mortality (IPM) models (*14–16*). Another family (with until now one member) is that of published population models, in particular *Prevent* (*17–20*).

IPM models per se have no population or age structure; they can be conceived of as stationary population models (i.e. models of a population in equilibrium, where the numbers of births and deaths in an age group are equal). However, DisMod, probably the IPM model in most common use (*14*), has gone through several versions, and the current version allows for hazard trend analysis that relies on modelling a full population structure based on one-year age groups. Notwithstanding, IPM models analyse only a single disease condition in isolation, and, while *Prevent* was explicitly designed to analyse a full population cohort structure, it also analyses only a single disease condition.

Multi-state life tables analyse multiple disease states but published versions have invariably required the assumption of independence across diseases. In addition, multi-state life tables implicitly impose a stationary population assumption by not independently modelling population time and age.

Averaging and its implications

In all compartmental models, of which differential equations models are one type, it is assumed that health and mortality risk are conditional on disease state. In light of the seemingly infinite diversity of real phenomena, this assumption invariably results in "compression", that is, the imposition of artificial homogeneity. In many cases, compression can be considered a necessary simplifying assumption for the modelling exercise, but in other cases, heterogeneity must be explicitly modelled to avoid the phenomenon of confounding. In a differential equations system, modelling heterogeneity of disease and mortality risk amounts to introducing additional disease states. Thus, PopMod, with four disease states, respresents a substantial increase in complexity over population models modelling only two disease states (e.g. diseased and healthy). PopMod of course includes the two-disease-state model as a special case.

There is also heterogeneity other than of disease and mortality risk. In particular, although real populations change in integer steps at discrete moments in time, a differential equation system represents this process in continuous time. However, this approximation is in general acceptably good when a large number of individuals comprise the population of

interest. Moreover, an implication of representing age in a discrete number of statistical bins is modelling a birth-year cohort as though it had a single average age. If births are distributed uniformly throughout the year, the average birthday of the cohort is the mid-year point, and there is no serious objection to this procedure. However, if the cohort average birthday is not be the mid-year point, PopMod's modelled age will differ from the true average age.

It is assumed that conditional hazards are constant within a single reporting interval (e.g. one year), which will in principle be problematic for conditions with high initial case-fatality, for example heart attack (or stroke). This sort of problem can be addressed by defining condition C as "acutely fatal cases" and condition X as "long-term survivors". Similarly, for conditions of determinate duration (e.g. pregnancy), use of a constant hazard rate for "remission" will result in an exponential distribution of waiting time for transition out of the state, whereas a uniform distribution of waiting time is what would be wanted.

All compartmental population models are fundamentally simplifications of reality by means of a system of reduced dimensionality. The mathematical concept of "projection" is useful: the simplified system can be thought of as a "least-squares approximation" to the higher-order real system (*21*). The validity of input parameter values and the accuracy of the solution method determine the actual goodness of fit realized in a particular model. Nevertheless, compression applies to every modelled variable in a differential equations model. Other modelling approaches, such as microsimulation, require much less compression, so the user who wishes to avoid compression systematically should consider adopting the microsimulation approach.

Types of error in PopMod

Sources of error in PopMod can be divided into three types:

(1) Model (or "projection") error due to analysing a simplified system instead of the full one. Model error includes the characterization of scenarios for disease interaction.
(2) Numerical error due to obtaining approximate solution values with numerical techniques.
(3) Parameter error due to uncertainty about observed or derived parameter values.

The 5th-order Runge-Kutta method provides an estimate of the local truncation error inherent in the 4th-order numerical technique. Monte-Carlo analysis of distributions around transition rates can be used to examine parameter uncertainty. However, comparison with a more complex model would be necessary for quantification of model error. A way of investigating the impact of model error would be to construct progressively more realistic and complex models. A spectrum of models,

from least to most complex, can thus be imagined, where the "most complex" and necessarily imaginary model has a one-to-one relation to real system it represents. The difference between the results of two adjacent models in such a series would be an expression of model error analogous to the estimate of numerical truncation error afforded by the next-higher-order numerical method.

Although intuitively natural and mathematically valid, in most situations it would be impractical to quantify model error in this laborious way. Nevertheless, model error may, in certain data-rich cases, be estimated by "predicting" outcomes for which numerical data are available for comparison but which are not used as inputs.

Limiting assumptions

Although any state transitions are in principle possible, PopMod assumes that transitions S→XC and X→C do not occur. This is because such transitions can be thought of as the simultaneous occurrence of two transitions (for example, S→XC equals S→X plus X→XC). Note that this does not imply events S→XC and X→C cannot occur within a single reporting interval; rather, it just means the mathematics of PopMod do not represent simultaneous events. A similar feature is the absence of a modelled cause of death "XC".

However, the non-modelled transition S→XC can be imagined if someone in state S simultaneously acquires X and C as a result of, say, very high levels of common risk factors (i.e. someone who suffers a simultaneous heart attack and stroke because of high blood pressure and cholesterol). If such a "simultaneous event" results in mortality, one could potentially speak of a cause of death "XC". Similarly, the non-modelled transition X→C could occur if there were "perfect interference" between two diseases such that acquiring C caused immediate remission from X. If either of these cases is important, PopMod can miss important dynamics.

Author's contributions

JL devised the methodology, implemented the conceptual and technical development of PopMod, including coordination of co-authors' contributions, and drafted and revised the manuscript. KR and HW contributed to the development of the methodology, drafted certain sections and revised the manuscript. CC and SG contributed to the development of the methodology, revised mathematical formulae throughout, and revised the maunscript. CM provided the initial idea for the model and also contributed technical modifications throughout development of the main ideas presented in this paper. All authors approved the final manuscript.

CONFLICT OF INTEREST

ACKNOWLEDGEMENTS

REFERENCES

1. Shryock HS, Siegel JS. The life table. In: Stockwell EG, editor. *The methods and materials of demography*. Orlando, FL: Academic Press; 1976, p. 250.

2. Hacking I. *The emergence of probability*. Cambridge: Cambridge University Press; 1975.

3. Barendregt JJ, Bonneux L. *Degenerative disease in an aging population: models and conjectures*. Enschede, Netherlands: Febodruke; 1998.

4. Manton KG, Stallard E. *Chronic disease modelling*. London: Charles Griffin; 1988.

5. Schoen R. *Modelling multigroup populations*. New York and London: Plenum Press; 1987.

6. Barendregt JJ. Incidence- and prevalence-based summary measures of population health (SMPH): making the twain meet. In: Murray CJL, Salomon J, Mathers C, Lopez A, editors. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002, p. 221–231.

7. Preston SH. Health indices as a guide to health sector planning: a demographic critique. In: Gribble JN, Preston SH, editors. *The epidemiological transition, policy and planning implications for developing countries*. Washington DC: National Academy Press; 1993.

8. Murray CJL. Rethinking DALYs. In: Murray CJ, Lopez AD, editors. *The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard University Press; 1996, p. 1–98.

9. Anderson RM, May RM. *Infectious diseases of humans: dynamics and control*. Oxford: Oxford University Press; 1991.

10. Lambert JD. *Numerical methods for ordinary differential systems: the initial value problem*. Chichester: Wiley; 1991.

11. Ralston A, Rabinowitz P. *A first course in numerical analysis*. 2nd ed. Mineola, NY: Dover; 1965.

12. Barendregt JJ, van Oortmarssen GJ, van Hout BA, Van Den Bosch JM, Bonneux L. Coping with multiple morbidity in a life table. *Math Popul Stud* 1998; 7(1):29–49, 109.

13. Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia, PA: Lippincott Williams and Wilkins; 1998.

14. Kruijshaar ME, Barendregt JJ, Hoeymans N. The use of models in the estimation of disease epidemiology. *Bulletin of the World Health Organization* 2002; 80(8):622–628.

15. Murray CJ, Lopez AD. Quantifying disability: data, methods and results. *Bulletin of the World Health Organization* 1994; 72(3):481–494.

16. Barendregt JJ, Baan CA, Bonneux L. An indirect estimate of the incidence of non-insulin-dependent diabetes mellitus. *Epidemiology* 2000; 11(3):274–279.

17. Mooy JM, Gunning-Schepers LJ. Computer-assisted health impact assessment for intersectoral health policy. *Health Policy* 2001; 57(3):169–177.

18. Bronnum-Hansen H. How good is the Prevent model for estimating the health benefits of prevention? *Journal of Epidemiology and Community Health* 1999; 53(5):300–305.

19. Gunning-Schepers LJ. Models: instruments for evidence based policy. *Journal of Epidemiology and Community Health* 1999; 53(5):263.

20. Bronnum-Hansen H, Sjol A. [Prediction of ischemic heart disease mortality in Denmark 1982–1991 using the simulation model Prevent]. *Ugeskr Laeger* 1996; 158(35):4898–4904.

21. Strang G. *Linear algebra and its applications*. 3rd ed. San Diego, CA: Harcourt Brace Jovanovich; 1988.

3   Programme costs in the economic evaluation of health interventions[1]

Benjamin Johns, Rob Baltussen, Raymond Hutubessy

Abstract

Estimating the costs of health interventions is important to policy-makers for a number of reasons including the fact that the results can be used as a component in the assessment and improvement of their health system performance. Costs can, for example, be used to assess if scarce resources are being used efficiently or whether there is scope to reallocate them in a way that would lead to improvements in population health. As part of its WHO-CHOICE project, WHO has been developing a database on the overall costs of health interventions in different parts of the world as an input to discussions about priority setting.

Programme costs, defined as costs incurred at the administrative levels outside the point of delivery of health care to beneficiaries, may comprise an important component of total costs. Cost-effectiveness analysis has sometimes omitted them if the main focus has been on personal curative interventions or on the costs of making small changes within the existing administrative set-up. However, this is not appropriate for non-personal interventions where programme costs are likely to comprise a substantial proportion of total costs, or for sectoral analysis where questions of how best to reallocate all existing health resources, including administrative resources, are being considered.

This paper presents a first effort to systematically estimate programme costs for many health interventions in different regions of the world. The approach includes the quantification of resource inputs, choice of resource prices, and accounts for different levels of population coverage. By using an ingredients approach, and making tools available on the World Wide Web, analysts can adapt the programme costs reported here to their local settings. We report results for a selected number of health interventions and show that programme costs vary considerably across interventions and across regions, and that they can contribute substantially to the overall costs of interventions.

Introduction

Estimating the costs of health interventions is important to policy-makers for a number of reasons including the fact that the results can be used as a component in the assessment and improvement of the performance of their health systems. As part of its WHO-CHOICE cost-effectiveness work programme (go to www.who.int/evidence/cea for more details), WHO has undertaken an effort to assess the overall costs and effects of a wide variety of health interventions *(1)*. Single global estimates of intervention costs are not relevant to individual countries. On the other hand, very few countries are able to estimate the costs of all possible interventions in their settings. WHO-CHOICE is, therefore, assessing the costs and effects of a range of interventions for 14 epidemiologic subregions of the world. The provision of subregional estimates allows interventions to be classified into broad categories for decision-making that have broad validity across that set of countries—e.g. those that are very cost-effective, those that are cost-effective, and those that are cost-ineffective. Policy-makers can then ask if there are good reasons why very cost-effective interventions are not done in their setting, while at the same time cost-ineffective interventions are being done *(2)*. The results will be presented in a way that analysts from countries in each region will be able to judge the appropriateness of the findings for their country and adapt them to their own settings. In the future, WHO-CHOICE will provide technical assistance to selected countries interested in applying the tools of generalized cost-effectiveness analyses.

   Costs can be divided into «patient costs» and non-patient or «programme costs». Patient costs refer to all costs at the point of delivery such as outpatient visits, bed days, drugs, or laboratory tests. Programme costs include costs incurred at the administrative levels of the district, provincial or central-levels, i.e. the costs incurred at a level other than the

delivery point of an intervention to beneficiaries. The components include such items as administration, training or media campaigns *(3)*. It is not uncommon for analysts to ignore programme costs when performing CEA. For example, only one *(4)* out of nine studies examining the cost-effectiveness of tuberculosis treatment strategies clearly showed that programme costs had been incorporated *(5)*. That study estimated the average cost of different ways of directly observing tuberculosis treatment as a means of improving adherence. For the option of completely ambulatory short course chemotherapy with daily supervision, programme costs accounted for 33%, 16% and 34% of estimated total costs in Mozambique, Malawi and Tanzania in turn. These findings suggest that programme costs can be a substantial proportion of total costs, and that the proportion may well vary across settings. They also mean that using a simple rule of thumb in which programme costs are assumed to be a fixed percentage of patient costs may not always be appropriate—although probably preferable than ignoring this category of cost completely *(6;7)*.

While most CEA guidelines recommend including all relevant costs that vary between programmes, studies may ignore them because they use an "incremental" approach to costing—comparing the introduction of a new technology against an existing intervention *(8;9)*. These studies are concerned with marginal changes in costs and effects; they assume that overhead items such as programme costs will remain approximately the same for each alternative being compared, and will not affect the choice between the given alternatives *(8)*. However, this is simply not appropriate when considering non-personal health interventions, such as mass media campaigns to encourage people to exercise more, where virtually the entire intervention consists of programme costs. Nor is it appropriate in many personal health interventions, such as the tuberculosis case described above, or when analysts are interested in answering the question of how best to use existing health resources to improve population health *(10)*.

This paper presents the systematic method for estimating programme costs for health interventions across settings used for WHO-CHOICE. The method and the resulting estimates can be used for different purposes, e.g. cost-effectiveness analysis (CEA) and other types of costing exercises such as estimating the costs of scaling-up interventions as part of the activities of such bodies as the Global Fund to Fight AIDS, Tuberculosis and Malaria. The following section presents the methods for identifying, collecting and calculating programme costs, including consideration of the theoretical basis for calculating programme costs. The third section presents an application of the approach including programme cost estimates for a number of interventions. Conclusions are presented in the final section.

Methods

This section describes the methods used in calculating programme costs as part of WHO-CHOICE. The first part discusses the theoretical approach for defining relevant costs. The second and third parts document the methods used to determine the amount of resource use and their prices. The last part elucidates a means of accounting for different coverage levels of an intervention.

Conceptual Approach

Observed prices or charges do not necessarily reflect economic value. Generally, the economic definition of costs should be used in cost valuation, not the accounting (or financial) definition. This is based on the concept of "opportunity cost", i.e. the value forgone by not utilizing the same resource in its next best alternative use *(11;12)*. The concept implies that all resources consumed by an intervention should be valued, not just those constituting a budgetary line item.

   In collecting costs, several basic issues concerning the costing process arise. The following issues outline the approach used to determine costs.

1. Joint or overhead costs

Programme cost analysis to inform decisions at the sectoral level requires information on the costs of introducing each intervention singly and also in combination with other related interventions. This requires identifying all resources involved to establish and run each intervention, including the necessary overheads.

   The simplest way to identify intervention-specific overhead costs is to identify shared resources used by the different interventions and use joint costing rules or some basis of allocation related to the usage of the overhead item *(8)*. The percentage of time devoted to each individual intervention was used to allocate personnel costs and the share of equipment used. Similarly with buildings and vehicles, the proportion of intervention-specific utilization to total utilization was used *(8;13)*. This implies that the resources are divisible, or can be shared across interventions (e.g. it is feasible to use 0.2 vehicles for an individual intervention). This is appropriate since most resources can be shared across interventions and programmes, and particular types of personnel, transport, and buildings can be hired in the short term or rented out to other users. In theory, all costs related to a set of evaluated interventions could be allocated. However, WHO-CHOICE excludes two major types of 'ongoing' costs in this context. First, some of the costs of central administration are not included—those that are part of the overall planning and management of the health system that are unrelated to the development and implementation of particular interventions aimed at improving health. Second, the current level of education of health professionals is excluded; if the skills required to deliver an intervention

are available in the country under study, training costs to develop those skills are not included in the programme costs since a reallocation of health system resources does not affect these costs.

2. Capacity utilization

The extent to which capital and labor are used can critically influences unit costs *(5;8)*. Capacity utilization is defined as the proportion of the total target workload time a resource is actually used; for example, a computer used 5 hours in a 10 hour work day has a capacity utilisation of 50%. In comparing the cost-effectiveness of interventions, it is important to ensure that the observed differences are due to the intrinsic characteristics of the intervention rather than the extent to which capital and labor have been utilized in the environment in which the interventions were evaluated. WHO-CHOICE seeks to inform policy-makers on the optimum mix of interventions if health resources could be reallocated. It is not useful to perform this analysis by analyzing some interventions that are delivered inefficiently and others delivered efficiently. Therefore, for this analysis we report the cost-effectiveness estimates of interventions that are done efficiently, using 80% capacity utilization as the norm. This is consistent with recommendations made in CEA guidelines and ensures the comparability of cost-effectiveness ratios across interventions and settings *(8;9)*.

3. Ingredients approach

Rather than collecting data on total expenditures, the ingredients approach is used. The cost of any input to a production process is the product of the quantity used and the value (or price) of each unit. The ingredients approach is useful for many reasons, the most important are that it allows analysts and policy-makers to validate the assumptions used; judge whether the estimates presented can be applied to their settings; and, if necessary, change some of the parameters to replicate the analysis for their settings *(3;13;14)*.

4. Classification of costs

Costs are classified according to three characteristics: phase of implementation of the intervention, organizational level where costs are incurred, and nature of costs. This can be classified in the following three categories, with primary classifications listed first:

- *Start-up* and *Post Start-up* costs: Programmes incur different types of costs in the start-up and post-implementation phases. The definition of the start-up period is the time between the decision to implement an intervention and starting its delivery to the first beneficiary. Quantities are reported for the total time of the start-up period. If the start-up period is 18 months, the quantities used for the entire time are reported. Post start-up programme costs for the full period of

implementation of the intervention were based on an estimate of the annual cost required to run the intervention in a typical post start-up year when the programme is fully implemented.

- *Central* versus *Lower Levels* costs: Factor inputs are classified according to where in the administrative and organizational level of the health system they are used. In this analysis we collected cost data from three programme-cost levels: central, provincial and district levels, but the data can be easily adapted to the relevant administrative classification in different settings.
- *Recurrent* versus *Capital* costs: Factor inputs are further classified into recurrent and capital items. Following standard practice, capital costs are annualized over the useful life of the factor input, i.e. the 'equivalent annual costs' are calculated.

5. Discounting across time

For country-specific analysis, the local rate of return on long-term government bonds would ideally be used as the social discount rate for costs. For our purposes, to allow comparability across regions, a 3% discount rate was used as recommended by most guidelines *(8)*.

Total start-up costs of the programme were considered as a capital investment and annualized and discounted over the life of the programme. For country-specific analysis, the choice of the period over which start-up costs should be annualised would be made on a case by case basis, but to allow comparability for the sub regional analysis, 10 years was chosen as the useful life of a start-up period *(3)*. The sensitivity of the analysis to this assumption was explored in the individual intervention studies.

DATA ON QUANTITIES

In the period 2001-2, WHO-CHOICE invited regional expert teams representing countries from each of the 14 epidemiologic sub regions to gather the quantities of physical inputs (the ingredients) required for approximately 75 interventions using a standard tool (see Endnotes section, Note 1 for details of the data collection tools and procedures). Most of the ingredients were for specific interventions, but some were for generic cost components which could be used in a number of interventions —for example, the cost of training health workers on case management using different combinations of number of days, and number of participants.

The data they provided were compiled and compared to form the basis of a set of costing sheets for the different activities covered by programme costs. Next, a list of required activities and the intensity of each activity was compiled for each intervention. For example, media outreach was classified into four intensities: extensive (daily or more radio and television emissions), moderate (weekly emissions), minimal (monthly emissions or less), and printed material only (for programmes which have

some information distribution requirements). Further activities included basic administration, monitoring, evaluation, and supervision, passage of legislation, training, and law enforcement. Other activities relevant only to one or a few programmes were entered separately. Training was divided into the costs involved in setting-up and running a specific training session, and the costs of overseeing and administering a training programme. The former costs are considered to vary with the number of trainees and length of training, while the latter were considered a fixed cost needed to run any training programme, no matter how many trainees or the length of training. The use of this standardized format ensures that different programmes are valued consistently based on the activities needed. This, in turn, ensures comparability of results.

Quantities were divided into fixed and variable costs. Fixed costs include those necessary to set up and run a programme no matter how many people are covered. Some examples of fixed cost include parts of the central administration, passage of legislation, and basic monitoring activities. Some examples of items that vary by the number of people covered include people delivering a service, the amount of storage space and shipment needed, supervision, and the production of printed information materials.

The required quantities of inputs were based on the estimates by the regional expert teams. However, because there was missing data for some interventions in some regions, the quantities for the variable and fixed cost functions were standardized across regions for most interventions (this was done except in cases where difference between regions is clearly justifiable, such as random breath testing of motor vehicle drivers where significantly different traffic patterns across regions would result in very different needs for enforcement). Because of different sizes of countries within the various regions, variable costs obviously varied by region (this builds in economies of scale, where fixed costs are spread over populations of different sizes).

The regional expert teams also estimated details such as the office supplies, equipment, and office space different staff members would consume in a year. Based on these assumptions, the quantities of utilities used and maintenance costs were also estimated (further details can be found at www.who.int/evidence/cea). Within the broad categories outlined in the conceptual approach section, inputs were classified in the manner reported in Table 1.

Table 1　　　Cost categories in programme cost sheet

A. Recurrent cost	
A.1. Personnel	Personnel time allocated to each intervention is netted out from time spent by those personnel in other interventions. Personnel time used in the start-up and post start-up periods is expressed in person-months.
A.2. Materials & Supplies	Materials and supplies in terms of the quantities used for the programme. Examples are office supplies that are used by the programme.
A.3. Media operating costs	Media inputs such as radio or television time, leaflets or posters are provided in terms of their unit of measurement (e.g. minutes for radio, or quarter page ads in newspapers).
A.4. Transport operating costs	Transport is measured in terms of total kilometers traveled per mean of transport.
A.5. Equipment operating cost	In cases when equipment is rented, the number of equipment and the duration of rental (in months) are reported.
A.6. Maintenance	Maintenance costs are listed as a percentage of annual costs.
A.7. Utilities	The amounts of utility items allocated to the programme are listed here. Examples of utility items are electricity, gas, and water. The allocation of the quantities used by the programme is based on the square meter surface area used by the programme, after applying any further allocation needed if the space is shared with other programmes.
A.8. Others	
A.8.1. Rented buildings	In case buildings are rented, both the total square meter surface area of the buildings and the duration of rental (in months) are used.
A.8.2. Per diems and travel allowances	The types of personnel who are entitled for per diems and travel are listed. The types reflect the activity they are involved in, e.g. trainers, trainees, support staff in meetings, participants of meetings, supervisors visiting health facilities etc. Reported by the number of days per type of personnel.
A.8.3. Miscellaneous items	Any other category of recurrent resources used that is not provided in the list are reported here by identifying the item and the quantities used.
B. Capital Costs	
B.1. Building	Space used by the programme are reported in terms of the total square meter surface area allocated to that programme, i.e., if the space used by the programme is shared with other activities, the share of the space used for the programme under study are estimated and the value are entered here.
B.2. Transport	The number of means of transport used by the programme is listed here. If they are only partly used, the estimated share of their use are entered.
B.3. Equipment and implements	The number of office equipment, storage and distribution, maintenance, cleaning and other capital equipment are reported here. If they are only partly used, appropriate allocation is made, using the same allocation factors used for building space.
B.4. Furniture	See point B.3 above.
B.5. Other capital costs	This section is used to report any other capital resources used by the programme.

15

DATA ON PRICES

This analysis requires the unit prices used to reflect the economic cost of goods, and allow for inter-country comparison of costs of interventions. For this purpose, the world price level was chosen as the *numeraire* or price level *(11)*, and a reference currency, i.e., the International Dollar (I$), was chosen for the presentation of the results at the international level. Costs in local country currency units were converted to international dollars using purchasing power parity (PPP) exchange rates. A PPP exchange rate is the number of units of a country's currency required to buy the same amounts of goods and services in the domestic market of a reference country, in this case the United States. An international dollar is, therefore, a hypothetical currency that is used as a means of translating and comparing costs from one country to the other. Because published estimates of PPPs do not cover all 192 countries that are members of WHO, the PPP exchange rates used in this analysis were developed by WHO and are available on the WHO-CHOICE website.

Prices for traded goods

Traded goods are commodities that are available on the international market, and all countries can purchase them at an international market price. Since the international market price reflects the opportunity cost of using a good to the country, it is used as the price for traded goods, adjusted to include cost, insurance and freight (c.i.f.) for imported goods and free on board (f.o.b.) for exported goods.

   International prices were derived from price indexes compiled in WHO publications and catalogues of prices from firms and non-governmental organizations operating at an international level that excluded costs of shipment and taxes. These international values were placed in a common currency (year 2000 I$) using World Bank Gross Domestic Product (GDP) deflators, or, when GDP deflators were unavailable, Consumer Price Index deflators *(15)*.

   Generally, for small items that can be bought in bulk, the lowest internationally listed price was selected. This assumes the existence of a basic health infrastructure, enabling the purchase of items in bulk. For larger items, a middle level price was selected to represent a "typical" price. In some cases, the price range for a good was too big to justify the use of a mid-level price. For example, a given model of a four-wheel-drive vehicle can range in price from US$ 15,000 to US$ 25,000. Thus, for vehicle prices, generators, and other large cost items, the regional expert teams were asked to provide the local price of goods excluding taxes and subsidies.

   The f.o.b. (free-on-board) price of exports includes the production cost, transport costs, local marketing costs and local port charges of the exporting country *(16)*. The c.i.f. (cost-insurance-freight) price excludes import duties and subsidies (transfer payments), and includes the selling price of the producing country, freight, insurance, and unloading charges.

If a country imports the good, the costs of local transport and distribution (termed 'domestic margin') were added to the c.i.f. price in order to approximate the local opportunity cost *(16)*. Methods for calculating c.i.f./f.o.b. adjustments are discussed in the section on coverage levels.

Prices for non-traded goods

Prices of non-traded goods like labour vary across regions. The regional expert teams provided local prices for non-traded goods for reference countries in their regions. Where possible, supplementary information from other sources on country-specific prices of non-traded goods, such as the International Labour Organization (ILO) database on occupational salaries, was also used to determine a typical cost for the region as a whole.

Coverage Levels

As coverage expands into remote areas, the marginal costs of providing an intervention to each additional person will generally increase *(17–19)*. To account for the increasing marginal costs of transportation to more remote areas, the following methods were used to adjust costs for different levels of population coverage. Transportation costs consist of the cost of transporting goods to a country (c.i.f./f.o.b.) and transporting goods within a country (the domestic margin).

Adjusting prices for traded goods

The calculation of the cost of transportation was based on the only available study showing the percentage change in the price of a traded good based on the distance it travels between countries, the transportation infrastructure and the average GDP per capita of a country, and other variables relating to the availability of seaports, neighbouring trade partners, etc. *(20)*. For purposes of calculating the c.i.f./f.o.b. mark-up of goods, an infrastructure index was calculated and applied using the price elasticity coefficients reported in Limão and Venables *(21)*. Table 2 illustrates the results of this analysis for selected countries in different regions with the c.i.f./f.o.b. mark-up ranging from 1.16 (16% increase in price) in Denmark to 1.71 in Afghanistan, with a median mark-up of 1.28.

The domestic margin was calculated based on a hexagon shaped regional distribution model *(22)*. Each hexagon was assumed to cover 80 square kilometres, approximating the area served by one health centre reflecting a circular area with a radius of 5 km *(23)*. The population of each hexagon was derived from Geographical Information System (GIS) data on the population density of a country. In this model, the population density of the most crowded 80 square kilometres is assumed to be at the centre of the country, with hexagons further from the centre having progressively lower population densities. Thus, in the case of Burkina Faso, 4% of the population is assumed to live in the central hexagon, while only 2% of the population is assumed to live in the adjacent hexagon.

Table 2   Mark-up of goods to account for the cost of transport

Country	CIF/FOB Ratio	Domestic Margin			
		50% Coverage	80% Coverage	95% Coverage	100% Coverage
Afghanistan	1.71	1.73	1.73	1.74	1.74
Brunei Darussalam	1.24	1.25	1.25	1.25	1.26
Burkina Faso	1.49	1.50	1.50	1.51	1.51
China	1.30	1.30	1.31	1.31	1.31
Denmark	1.16	1.16	1.16	1.17	1.17
India	1.24	1.24	1.24	1.24	1.25
Jordan	1.31	1.31	1.32	1.32	1.33
Mexico	1.27	1.27	1.27	1.27	1.27
Nicaragua	1.41	1.41	1.42	1.42	1.43
Russian Federation	1.26	1.27	1.27	1.28	1.29
Thailand	1.29	1.29	1.30	1.31	1.31
The former Yugoslav Republic of Macedonia	1.22	1.22	1.22	1.22	1.22
United Republic of Tanzania	1.42	1.43	1.43	1.44	1.44
United States of America	1.18	1.19	1.19	1.20	1.21

Each country is also divided into provinces and districts based on the number of provinces and districts reported by WHO databases. In cases where the number of secondary or tertiary administrative units was not certain, an average was taken from the available sources. The average size of a province or district was calculated by dividing the total area of a country by the number of provinces or districts, which were then incorporated into the hexagonal grid. A traded good was assumed to travel, on average, half the distance from the central hexagon to the centre of the most peripheral province, and then to the centre of a district. The Limão and Venables price elasticity for distance was then used with this calculated distance to derive the domestic margin. Since, in this model, the central areas are more crowded than outlying areas, a programme covering 50% of the population will have a proportionately lower mark-up than a programme covering 95% of the population. However, as shown in Table 2, the domestic margin is a minor cost compared to the cost of initially transporting a good to the country.

Impact on resource utilization

When an intervention covers a larger part of the population, the resources required to run the intervention also increase. As coverage goes up, certain cost parameter values were increased as follows:

- As indicated above, the hexagon shaped regional distribution model assumes a health centre for every 80 square kilometres of space. This implies that health centres may not always run at 80% capacity, since more remote areas may have a very low population density. Since costs of training of health care professionals are independent of population density, these costs—expressed as costs per capita—will increase as coverage levels increase.
- The number of provinces covered increases as coverage expands. Under the assumptions listed above, one or two provinces may contain 50% or more of the population. Thus, as coverage expands, the number of provinces covered will increase, but each new province covered will have fewer people. Since there are fixed costs associated with running a programme at the province level, this produces diseconomies of scale.
- The distance travelled in a supervision visit increases. At the national level, this is calculated as the distance from the centre to the most remote province covered (the average distance would be half the distance from the center to the periphery; however, because supervision visits are assumed to be round trips, the full distance from the center to the periphery is used). The distance travelled for supervision visits within provinces is similarly calculated.
- Thus, the number of programme staff involved in supervision activities needs to increase both in proportion to the increased distances covered and to account for the increased number of provinces. Each province was assumed to need an equal number of supervision visits.

It is possible that salaries may be higher in very remote areas to give health personnel extra incentive to relocate to these areas. In the absence of data, this factor was not incorporated.

ORGANIZING AND USING THE DATA

The predicted quantities of resources needed were multiplied by their respective prices to calculate the total programme costs for a ten-year period of implementation. These ten-year costs are calculated in year 2000 international dollars using a standard net present value formula *(8)*.

VALIDATION

Once the data had been collected and analysed, the accuracy of the data was verified. Where possible, previous costing or CE studies which included programme costs were used as a benchmark for comparison, but very few presented programme cost estimates using the ingredients approach (e.g. *(24)*). In addition, disease and public health experts or programme managers who are familiar with the particular interventions and settings for a number of diseases reviewed the final costing figures. In the cases where the estimates did not have face validity, controls were made to ensure that there had not been mistakes with coding, and discussions were held with the regional costing experts to confirm the basis of their quantity and price estimates.

RESULTS

WHO-CHOICE has used the methods described above to produce a set of cost-effectiveness estimates, initially for 14 epidemiologic sub regions *(25)*. Table 3 reports the average annual programme cost per capita, and as a percentage of total intervention cost per capita, for selected interventions in these regions. Costs are presented in 2000 International dollars. The table shows that programme costs vary across interventions and across regions for a given intervention. For example, cost per capita of educating sex workers totals I$ 0.01 in SearB, whereas it amounts to $0.07 in AfrE. For a population of 100 million people, this would mean programme costs differ substantially—$1 million in the former and $7 million in the latter sub region. Variations are caused by differences in the number of sex workers and in the number of social workers required to train sex worker peer educators, and to differences in regional price levels of inputs. (Note also that a straight comparison of cost per capita across interventions is misleading in deciding whether an intervention is of low cost or more expensive at a population level, because there is wide variation in the target populations for each of these interventions.)

The importance of programme cost in comparison with patient cost also varies by intervention and by region. Obviously, non-personal interventions such as the introduction of random breath testing for drivers to reduce the burden of motor vehicle accidents consists entirely of programme costs. On the other hand, the provision of brief physician advice to heavy alcohol users consists largely of patient costs, with programme costs ranging from less than 1% of total costs to almost 30%

The tools used to estimate these results are available on the Internet at http://www.who.int/evidence/cea for use by local analysts. They include:

- a database of prices for traded goods,
- a database listing the reported useful life of capital goods,
- a workbook listing activities used in programme costs together with assumptions of quantities of resources used based on the data collected by WHO, and
- a costing tool CostIt© to calculate and present the final results of the costing exercise.
- a tool for uncertainty analysis MCLeague© to calculate uncertainty regions around cost-effectiveness ratios and present stochastic league tables.

All of the estimates presented could be modified by analysts to suit the particularities of their own setting. In adapting these tools, analysts have to assess if the assumptions outlined in this paper are appropriate for their own setting. The following list highlights some major considerations:

Table 3 Average annual program cost per capita for selected interventions in GBD regions* (2000 I $)

Disease / intervention	HIV/AIDS: Preventing mother-to-child transmission		HIV/AIDS: Educating sex workers						Alcohol: Random breath testing of drivers**		Alcohol: Brief physician advice to reduce heavy alcohol use†	
Coverage level	Antenatal care coverage††		50%		80%		95%		95%		50%	
GBD2000 region‡	PC‡‡	PC as % of total costs	PC	PC as % of total costs	PC	PC as % of total costs	PC	PC as % of total costs	PC	PC as % of total costs	PC	PC as % of total costs
AfrD	$0.08	8%	$0.05	70%	$0.06	63%	$0.06	61%	$0.31	100%	$ 0.011	21%
AfrE	$0.15	10%	$0.07	74%	$0.09	69%	$0.10	67%	$0.42	100%	$ 0.012	8%
AmrA	$0.19	5%	$0.06	92%	$0.09	91%	$0.10	91%	$0.29	100%	$ 0.006	0%
AmrB	$0.05	5%	$0.02	84%	$0.03	78%	$0.03	76%	$0.28	100%	$ 0.007	2%
AmrD	$0.03	4%	$0.02	62%	$0.02	55%	$0.03	52%	$0.45	100%	$ 0.005	8%
EmrB	$0.11	9%	$0.08	96%	$0.09	96%	$0.09	96%	NA	N.A.	N.A.	N.A.
EmrD	$0.08	15%	$0.06	94%	$0.06	94%	$0.06	94%	NA	N.A.	N.A.	N.A.
EurA	$0.17	9%	$0.09	97%	$0.10	96%	$0.11	96%	$0.55	100%	$ 0.025	1%
EurB	$0.14	18%	$0.05	91%	$0.05	91%	$0.05	91%	$0.51	100%	$ 0.008	2%
EurC	$0.03	9%	$0.03	83%	$0.03	83%	$0.03	83%	$0.25	100%	$ 0.003	0%
SearB	$0.07	9%	$0.01	36%	$0.02	35%	$0.02	34%	$0.19	100%	$ 0.001	2%
SearD	$0.02	4%	$0.03	69%	$0.03	65%	$0.04	64%	$0.17	100%	$ 0.004	29%
WprA	$0.19	9%	$0.10	97%	$0.11	97%	$0.12	96%	$0.95	100%	$ 0.022	4%
WprB	$0.09	15%	$0.10	89%	$0.11	89%	$0.11	89%	$0.23	100%	$ 0.006	5%

* Costs are average annual discounted programme costs of implementing an intervention during 10 years

** Only relevant at 95% coverage

† Only relevant at 50% coverage

†† Current antenatal care coverage in GBD regions; defined as percentage of target population with at least one antenatal care visit during pregnancy

‡ AFR = Africa Region; AMR = Region of the Americas; EMR = Eastern Mediterranean Region; EUR = European Region; SEAR = South East Asian Region; WPR = Western Pacific Region.
A sub regions have very low rates of adult and child mortality; B = low adult, low child; C = high adult, low child; D = high adult, high child; E = very high adult, high child mortality.

‡‡ PC = Programme costs

- Local analysts may wish to carry out the analysis using a capacity utilization rate other than 80% to better reflect their actual situation. However, one standard rate should be used for the evaluation of all interventions to ensure comparability. The CostIt© tool allows this to be done automatically.
- Local analyst may wish to use local prices rather than international prices as estimated by WHO-CHOICE. Analyst can also vary prices for non-traded goods according to the location within the country where they are incurred; for example, provincial staff may have lower salaries than staff in the capital city, or vice-versa.
- The spatial model for scaling-up can be revised to the geography of a particular country. For example, multiple points of entry for traded goods can be considered. Alternatively, local analysts may be able to gather data on the prices of goods in various parts of the country, or the costs of transportation, and thus not need to employ the model as used by WHO-CHOICE. Further, the assumption that the number of provinces expands with increasing population coverage may not accurately reflect how a country implements health interventions, and analysts should adjust their assumptions accordingly. Finally, the coverage area of health centres can be determined locally.

Conclusion

Programme costs can constitute a substantial component of costs even for personal health interventions and should not be ignored in the economic evaluation of health interventions. This paper has presented a first effort to systematically analyze programme costs in different sub regions of the world. The use of a standardized methodology ensures comparability of cost estimates across interventions and settings.

In addition, this paper has introduced "ready-to-use" tools and programme cost estimates that are available on the World Wide Web. The programme cost estimates constitute an important part of WHO-CHOICE database on costs and effects of multiple interventions in various regions in the world exploring the question of whether resources are being used to achieve the maximum possible level of population health. Analysts may wish to adapt the regional estimates to their local setting to make the results more relevant for local decision makers. This paper has shown that, in this process, special attention should be paid to issues such as capacity utilization, prices of goods, and increasing marginal costs of delivering interventions into more remote areas.

As with any innovative work, there are some limitations to the approach that has been used, which offers possibilities of further development over time. For example, in the consultation process with regional expert teams to obtain input quantities and prices, considerable efforts were made to standardize reporting approaches. Nevertheless, reported quantities still showed considerable variation beyond that

reasonably expected on the basis of regional differences, and it was necessary to return to the experts for clarification and to seek the input of external data sources and expert advice. Analysts wishing to adapt the results to their own settings should be aware that they would need to seek the advice of more than one expert in their own countries before adapting the quantities of inputs and unit prices reported here. WHO-CHOICE incorporates extensive efforts to develop methods for uncertainty analysis, to reflect uncertainty in the final cost and cost-effectiveness estimates. This is designed to help local policy makers decide the extent to which the results of the WHO-CHOICE analysis inform policy in their countries *(26;27)*.

A key element in our approach is the specification of intervention cost functions at various coverage levels. Whereas other studies have estimated costs of scaling-up health services using a linear cost function, the present study includes non-linearities *(28)*. Economies of scale have been incorporated by allowing some costs to be fixed regardless of the size of the population reached – television broadcasts are a case in point. On the other hand, diseconomies of scale have been included by using higher prices (for transport costs) and higher quantities (for training and supervision) at higher coverage levels. This is an important step for showing the impact of higher coverage on costs and outcomes. However, further work is required to add non-spatial determinants of increasing costs relating to scaling-up.

Conflict of interests

None.

Authors' contributions

BJ has day-to-day responsibility for the data management of programme costs, participated in the development of the methodology and drafted the manuscript. RB and RH participated in the development and coordination of the methodology. All authors read and approved the final manuscript.

Acknowledgements

methods used. We are grateful to Chris Murray, David Evans, and Tessa Tan Torres for general guidance throughout this project.

The open peer review comments of Dr Guy Hutton (Switzerland) and Dr Frederick Mugisha (Kenya) on an earlier version of this paper are also gratefully acknowledged.

The views expressed are those of the authors and not necessarily those of the organization they represent.

Endnote section

Note 1:

WHO-CHOICE instructed the costing experts on data gathering techniques. Each was given a standardised collecting tool and a guideline, and most attended a workshop detailing the methods to be used. The standardized data collection tool involved two Microsoft Excel spreadsheets. The first, the "general information" sheet, documents general health system parameters of a country. This sheet contains five tables, some for use in determining patient costs, some for use in determining programme costs. The second spreadsheet provided a template for recording the quantities of resource inputs for each intervention (see Table 1). A WHO-CHOICE team member made a follow-up visit to each country to determine the adequacy of the experts' techniques, answer questions, and provide further guidance. Responses were checked against those of other experts, as well as the literature, allowing outliers to be identified and the sources of any difference to be explored and corrected if necessary.

References

1. Hutubessy RCW, Baltussen RMPM, Tan Torres Edejer T, Evans DB. Generalized cost-effectiveness analysis: an aid to decision making in health. *Applied Health Economics and Health Policy* 2002; 1(2):89-95.

2. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 1997; 349(9063):1436-1442.

3. Baltussen RM, Adam T, Tan Torres T, Hutubessy RC, Acharya A, Evans DB, Murray CJ. *Generalized Cost-Effectiveness Analysis: A Guide.* http://www.who.int/evidence/cea. 2002.

4. de Jonghe E, Murray CJ, Chum HJ, Nyangulu DS, Salomao A, Styblo K. Cost-effectiveness of chemotherapy for sputum smear-positive pulmonary tuberculosis in Malawi, Mozambique and Tanzania. *The International Journal of Health Planning and Management* 1994; 9(2):151-181.

5. Adam T, Evans DB, Koopmanschap MA. Cost-effectiveness analysis: can we reduce variability in costing methods? *International Journal of Technology Assessment in Health Care* 2003;(forthcoming).

6. Floyd K, Wilkinson D, Gilks C. Comparison of cost effectiveness of directly observed treatment (DOT) and conventionally delivered treatment for tuberculosis: experience from rural South Africa. *British Medical Journal* 1997; 315(7120):1407-1411.

7. Creese A, Parker D. *Cost analysis in primary health care: A training manual for programme managers.* Geneva: World Health Organization; 1994.

8. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes.* 2nd ed. Oxford: Oxford University Press; 1997.

9. Gold MR, Siegel JE, Russel LB, Weinstein MC. *Cost-effectiveness in health and medicine.* New York: Oxford University Press; 1996.

10. Murray CJ, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Economics* 2000; 9(3):235-251.

11. Little IMD, Mirrlees JA. *Project appraisal and planning for developing countries.* London: Heinemann Educational Books; 1974.

12. Mishan EJ. *Cost-benefit analysis.* 3rd ed. London: George Allen & Unwin (Publishers) Ltd; 1982.

13. Creese A, Parker D. *Cost analysis in primary health care: a training manual for programme managers.* Geneva: World Health Organization; 1994.

14. Walker D. How to do (or not to do)... Cost and cost-effectiveness guidelines: which ones to use? *Health Policy and Planning* 2001; 16(1):113-121.

15. World Bank. *World Development Indicators 2001.* Washington, D.C: World Bank; 2001.

16. Gittinger J. *Economic analysis of agricultural projects.* John Hopkins Press; 1984.

17. Berman P, Brotowasisto, Nadjib M, Sakai S, Gani A. The costs of public primary health care services in rural Indonesia. *Bulletin of the World Health Organization* 1989; 67(6):685-694.

18. Berman P. Cost efficiency in primary health care: studies of health facilities in Indonesia. *Health Policy and Planning* 1989; 4(4):316-322.

19. Over M. The effect of scale on cost projections for a primary health care program in a developing country. *Social Science and Medicine* 1986; 22(3):351-360.

20. Limao N, Venables AJ. Infrastructure, geographical disadvantage, transport costs, and trade. *World Bank Economic Review* 2001; 15:451-479.

21. Roncek DW. Learning more from tobit coefficients: extending a comparative analysis of political protest. *American Sociological Review* 1992; 57:503-507.

22. Gusein-Zade SM. Comments on "A note on the location of medical facilities" by Z. Drezner. *Journal of Regional Sciences* 1992; 32(2):229-231.

23. Khan MM, Ali D, Ferdousy Z, Al Mamun A. A cost-minimization approach to planning the geographical distribution of health facilities. *Health Policy and Planning* 2001; 16(3):264-272.

24. Sweat M, Gregorich S, Sangiwa G, Furlonge C, Balmer D, Kamenga C et al. Cost-effectiveness of voluntary HIV-1 counselling and testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. *Lancet* 2000; 356(9224):113-121.

25. World Health Organization. *The World Health Report 2002: Reducing risks, promoting healthy life.* Geneva: World Health Organization; 2002.

26. Baltussen RM, Hutubessy RC, Evans DB, Murray CJ. Uncertainty in cost-effectiveness analysis. Probabilistic uncertainty analysis and stochastic league tables. *International Journal of Technology Assessment in Health Care* 2002; 18(1):112-119.

27. Hutubessy RC, Baltussen RM, Evans DB, Barendregt JJ, Murray CJ. Stochastic league tables: communicating cost-effectiveness results to decision-makers. *Health Economics* 2001; 10(5):473-477.

28. WHO Commission on Macroeconomics and Health. *Macroeconomics and health: investing in health for economic development. Report of the Commission on Macroeconomics and Health.* Geneva: World Health Organization; 2001.

4   ECONOMETRIC ESTIMATION OF
   COUNTRY-SPECIFIC HOSPITAL COSTS[1]

TAGHREED ADAM, DAVID B. EVANS, CHRISTOPHER J.L. MURRAY

Abstract

Information on the unit cost of inpatient and outpatient care is an essential element for costing, budgeting and economic-evaluation exercises. Many countries lack reliable estimates, however. WHO has recently undertaken an extensive effort to collect and collate data on the unit cost of hospitals and health centres from as many countries as possible; so far, data have been assembled from 49 countries, for various years during the period 1973–2000. The database covers a total of 2173 country-years of observations. Large gaps remain, however, particularly for developing countries. Although the long-term solution is that all countries perform their own costing studies, the question arises whether it is possible to predict unit costs for different countries in a standardized way for short-term use. The purpose of the work described in this paper, a modelling exercise, was to use the data collected across countries to predict unit costs in countries for which data are not yet available, with the appropriate uncertainty intervals.

The model presented here forms part of a series of models used to estimate unit costs for the WHO-CHOICE project. The methods and the results of the model, however, may be used to predict a number of different types of country-specific unit costs, depending on the purpose of the exercise. They may be used, for instance, to estimate the costs per bed-day at different capacity levels; the "hotel" component of cost per bed-day; or unit costs net of particular components such as drugs.

In addition to reporting estimates for selected countries, the paper shows that unit costs of hospitals vary within countries, sometimes by an order of magnitude. Basing cost-effectiveness studies or budgeting exercises on the results of a study of a single facility, or even a small group of facilities, is likely to be misleading.

Introduction

Information on hospital unit costs is valuable to health decision-makers and researchers for at least three purposes: budgeting (now receiving more attention with the availability of additional funds for health in poor countries through the Global Fund to Fight AIDS, Tuberculosis and Malaria); the assessment of hospital efficiency; and the assessment, by means of either cost-benefit or cost-effectiveness analysis, of the efficiency of different health interventions. Recognizing the need to make this information available on a country-specific basis, WHO has undertaken as part of the work programme WHO-CHOICE (**CHO**osing **I**nterventions that are **C**ost-**E**ffective—see www.who.int/evidence/cea), an extensive effort to collate all sources of data on unit costs from as many countries as possible *(1)*. Large gaps remain, however, particularly for developing countries. Although the long-term solution is that all countries perform their own costing studies, the question arises whether it is possible to predict unit costs for different countries in a standardized way for short-term use. The purpose of the work described in this paper is to use the data collected across countries to predict unit costs in countries for which data are not yet available (both point estimates and uncertainty intervals are reported).

Health economics has a long tradition of estimating hospital-cost functions econometrically *(2–10)*. Econometric models explain how total costs change in response to differences in service mix, inputs, input prices, and scale of operations. They allow cost and production functions to be specified with sufficient flexibility that a non-linear relationship can be demonstrated between costs and quantity of inputs: total costs can rise at a lower rate than prices *(2)*.

Previous studies have commonly used microeconomic data to analyse and estimate hospital-cost functions. This literature indicates two main approaches: behavioural cost functions and cost minimization functions *(2;3;9;11)*. Behavioural cost functions have been used to explain the variations in cost per unit of output among hospitals. They have used as determinants all variables for which a causal relationship to hospital costs is hypothesized and data are available—e.g., bed size, global indicators of hospital activity such as average length of stay and occupancy rate, dummy variables for teaching status, etc. On the other

hand, the literature on cost minimization has described the minimum cost of providing a given volume of output as a function of an exogenous vector of input prices and the volume of output. The purpose is to determine whether hospitals are cost-minimizers (profit maximizers).

When testing the hypothesis of cost-minimization, the explanatory variables typically comprise only of output quantities (e.g., number of bed days) and input prices. The remaining variables used in the behavioural cost function specification are not part of the cost minimization question but can be used to explain deviation of observed unit costs from the theoretical minimum functions—e.g., possible reasons for inefficiency (3).

To our knowledge, all previous studies have used within-country data sets; we know of none that has attempted to estimate hospital-cost functions across countries. Such studies require a large number of observations from as many countries as possible.

The model described here follows the tradition of the behavioural cost function literature because its purpose is to estimate country-specific costs per bed-day, not to test the hypothesis of cost-minimization. The analysis controls for across-country price-level differences by using unit costs adjusted for purchasing-power parity, namely in international dollars; and for differences in quantity and complexity of resource use by using macro-level indicators such as per capita GDP (12–14). The model forms part of a series of models that can be used to predict country-specific unit costs for a number of purposes. They may be used, for instance, to estimate: (i) unit costs at different capacity levels for the purposes of efficiency analysis or economic evaluation of health interventions; (ii) the "hotel" component of average cost per bed-day for budgeting exercises; or (iii) unit costs excluding components that might be funded from other sources, such as drugs. The specific objectives of this paper are to:

- explain the observed differences in hospital inpatient cost per bed-day across and within countries; and
- use the results to predict cost per bed-day for countries for which these data are not yet available.

METHODS

DATA

The search sources used to obtain the data were: Medline, Econlit, Social Science Citation Index, regional Index Medicus, Eldis (for developing-country data), Commonwealth Agricultural Bureau (CAB), and the British Library for Development Studies Databases. The range of years was set at 1960 to the present. Data covering costs and charges were included.

The search terms used were: "costs and cost analysis" and hospital costs or health centre or the abbreviations HC (health centre) or PHC

(primary health centre) or outpatient care. The language sources searched were English, French, Spanish and Arabic; no Arabic study was found. In addition, a number of studies were found in the grey literature, from such sources as electronic databases, government regulatory bodies, research institutions, and individual health economists known to the authors *(2;15–54)*. Also included were data from a number of WHO-commissioned studies on unit costs.

A standard template was used for extracting data from all sources. Database variables include: ownership; level of facility (see Annex 1 for a definition of facility types as coded in the unit cost database); number of beds; number of inpatient and outpatient specialties; cost data (cost per bed-day, outpatient visit, and admission); utilization data (bed-days, outpatient visits, admissions); types of cost included in the cost analysis (capital, drugs, ancillary, food) and whether they were based on costs or charges; capacity utilization (occupancy rate, average length of stay, bed turnover, and average number of visits per doctor per day); reference year for cost data; currency, and methods of allocation of joint costs. The database consists of unit-cost data from 49 countries for various years between 1973–2000, totalling 2173 country-years of observations. Some studies provided information on 100% of the variables described above; at the other extreme, some provided information on less than 15%. The number of observations used in this analysis was 1171 (see Annex 2 for the percentage of missing data in the model variables and Annex 3 for the list of countries).

Data cleaning comprised consistency checks and direct derivation of some of the missing variables, when possible, from other variables from the same observation (e.g., occupancy rate was calculated from number of beds and number of bed-days). STATA software was used for data analysis *(55)*.

Cost data were converted to 1998 International dollars by means of GDP deflators *(56)* and purchasing-power-parity exchange rates used for WHO's national health accounts estimates (PPP exchange rates used in this analysis are available from the WHO-CHOICE website: www.who. int/evidence/cea).

DATA IMPUTATION

Most statistical procedures rely on complete-data methods of analysis: computational programmes require that all cases contain values for all variables to be analyzed. Thus, as default, most software programs exclude from the analysis observations with missing data on any of the variables (list-wise deletion). This can give rise to two problems: compromised analytical power, and estimation bias. The latter occurs, for example, if the probability that a particular value is missing is correlated with certain determinants. For example, if the complete observation sets tend to be from observations with unit costs that are systematically higher or lower than average, the conclusions for out-of-sample estimation drawn from an analysis based on list-wise deletion will be biased upwards or downwards *(57)*.

There is a growing literature on how to deal with missing data in a way that does not require incomplete observation sets to be deleted, and several software programs have been developed for this purpose. If data are not missing in a systematic way, missing data can be imputed using the observed values for complete sets of observations as covariates for prediction purposes. Multiple imputation is an effective method for general-purpose handling of missing data in multivariate analysis; it allows subsequent analysis to take account of the level of uncertainty surrounding each imputed value, as described below *(58–61)*. The statistical model used for multiple imputation is the joint multivariate normal distribution. One of its main advantages is that it produces reliable estimates of standard errors: single imputation methods do not allow for the additional error introduced by imputation. In addition, the introduction of random error into the imputation process makes it possible to obtain largely unbiased estimates of all parameters *(58)*.

In this study, multiple imputation was performed with *Amelia,* a statistical software program designed specifically for multiple imputation of missing data *(57;59;62;63)*. First, five completed-data sets are created by imputing the unobserved data five times, using five independent draws from an imputation model. The model is constructed to approximate the true distributional relationship between the unobserved data and the available information. This reduces potential bias due to systematic difference between the observed and the unobserved data. Second, five complete-data analyses are performed by treating each completed-data set as an actual complete-data set; this permits standard complete-data procedures and software to be utilized directly. Third, the results from the five complete-data analyses are combined *(64)* to obtain the so-called repeated-imputation inference, which takes into account the uncertainty in the imputed values.

Model specifications

From the tradition of using cost functions to explain observed variations in unit costs, we estimate a long-run cost-function by means of Ordinary Least Squares regression analysis (OLS); the dependent variable is the natural log of cost per bed-day *(2;3;6–8;65)*. The primary reason for using unit cost rather than total cost as the dependent variable is to avoid the higher error terms due to non-uniform variance (heteroscedasticity) in the estimated regression. This could arise if total cost were used as the dependent variable, as the error term could be correlated with hospital size *(2;3)*. The reason for using cost per bed-day rather than cost per admission is that "bed-days" are better than "admissions" as a proxy for such hospital services as nursing, accommodation and other "hotel services" *(3),* permitting more flexibility in the use of estimated unit costs.

As the relationship between unit costs and the explanatory variables are expected to be non-linear, the Cobb-Douglas transformation was used to

approximate the normal distribution of the model variables. Natural logs were used. The Cobb-Douglas functional form can be written as follows:

Equation 1

$$Y = \alpha_0 X_1^{\alpha_1} X_2^{\alpha_2} \text{ or,}$$

Equation 2

$$\ln(Y) = \delta + \alpha_1 \ln(X_1) + \alpha_2 \ln(X_2)$$

where $\delta = \ln(\alpha_0)$. This function is non-linear in the variables Y, X_1 and X_2, but it is linear in the parameters δ, α_1, α_2, and can be readily estimated using Ordinary Least Squares *(66)*.

Log transformation has the added advantage that coefficients can be readily interpreted as elasticities *(3;66)*.

Therefore, the cost function specification of the OLS regression model may be written as:

Equation 3

$$UC_i = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i + e_i$$

Where UC_i is the natural log (ln) of cost per bed-day in 1998 I $ in the *ith* hospital; X_1 is ln of GDP per capita in 1998 I $; X_2 is ln of occupancy rate; $X_{3,4}$ are dummy variables indicating the inclusion of drug or food costs (included = 1); $X_{5,6}$ are dummy variables for hospital levels 1–2 (the comparator is level 3 hospital); $X_{7,8}$ are dummy variables indicating facility ownership (comparator is private not-for-profit hospitals); X_9 is a dummy variable controlling for USA data (USA = 1); and *e* denotes the error term.

The choice of explanatory variables is partly related to economic theory and partly determined by the purpose of the exercise, which is to estimate unit costs for countries where the data are not available. In this case, the chosen explanatory variables must be available in the out-of-sample countries. Country-specific—or in the case of large countries such as China, province-specific—GDP per capita in international dollars (I $) is used as a proxy for level of technology *(12–14)*; occupancy rate as a proxy for level of capacity utilization; and hospital level as a proxy for case mix. Unit costs are expected to be correlated positively with GDP per capita and case mix and negatively with capacity utilization.

The inclusion of the seven control variables makes it possible to estimate unit cost for different purposes to suit different types of analysis—for example, cost per bed-day in a primary-level hospital, which does not provide drugs or food; or the cost in a tertiary level hospital, with drugs and food included.

The dummy for the USA was included because all data were charges rather than costs and because there were a large number of observations

from that country. Dummies for countries other than the USA with a large number of observations, such as China and the United Kingdom, were also tested as was the use of dummy variables to capture whether the cost estimates included capital or ancillary costs. These variables were not included in the model which best fit the data. Utilization variables, such as number of bed-days or outpatient visits, and hospital indicators, such as average length of stay, were not included as explanatory variables because most out-of-sample countries do not have data on these variables, and prediction of unit costs would, therefore, be impossible.

MODEL-FIT

Regression diagnostics were used to judge the goodness-of-fit of the model. They included the tolerance test for multicollinearity, its reciprocal variance inflation factors and estimates of adjusted R square and F statistics of the regression model.

PREDICTED VALUES AND UNCERTAINTY ANALYSIS

Two types of uncertainty arise from using statistical modes: estimation uncertainty arising from not knowing β and α perfectly—an unavoidable consequence of having a finite number of observations; and fundamental uncertainty represented by the stochastic component as a result of unobservable factors that may influence the dependent variable but are not included in the explanatory variables *(62)*. To account for both types of uncertainty, statistical simulation was used to compute the quantities of interest, namely average cost per bed-day and the uncertainty around these estimates. Statistical simulation uses the logic of survey sampling to learn about any feature of the probability distribution of the quantities of interest, such as its mean or variance *(62)*.

It does so in two steps. First, simulated parameter values are obtained by drawing random values from the data set to obtain a new value of the parameter estimate. This is repeated 1000 times. Then the mean, standard deviation, and 95% confidence interval around the parameter estimates are computed. Second, simulated predicted values of $\hat{y}$ (the quantity of interest) are calculated, as follows: (1) one value is set for each explanatory variable; (2) taking the simulated coefficients from the previous step, the systematic component (g) of the statistical model is estimated, where $g = f(X,B)$; (3) the predicted value is simulated by taking a random draw from the systematic component of the statistical model; (4) these steps are repeated 1000 times to produce 1000 predicted values, thus approximating the entire probability distributiôn of y. From these simulations, the mean predicted value, standard deviation, and 95% confidence interval around the predicted values are computed. In this way, this analysis accounts for both fundamental and parameter uncertainty.

The predicted log of cost per bed day, $\ln \overline{UC}$ can then be calculated from:

Equation 4

$$\ln \overline{UC} = \overline{\alpha_0} + \overline{\alpha_1} \ln X_1 + \sum_{i=1}^{n} \overline{\alpha_i} X_i$$

where $\overline{\alpha_0}$ and $\overline{\alpha_{i..n}}$ are the estimated parameters, and $X_{i..n}$ are the independent variables. $\overline{\beta_0}$ =anti log $(\overline{\alpha_0})$ and $\overline{\beta_1} = \overline{\alpha_1}$ back-transforming Equation 4 (reduced to 1 independent log-transformed variable for simplicity) gives the power function.

Equation 5

$$\overline{UC}_{biased} = \overline{\beta_0} X^{\overline{\beta_1}}$$

where $\overline{UC}_{biased}$ denotes a biased estimate of the mean cost per bed-day due to back-transformation. This is because one of the implicit assumptions of using log-transformed models is that the least-squares regression residuals in the transformed space are normally distributed. In this case, back-transforming to estimate unit costs gives the median and not the mean. To estimate the mean it is necessary to use a bias correction technique. The smearing method described by Duan (1983) was used to correct for the back-transformation bias *(67)*. The smearing method is non-parametric, since it does not require the regression errors to have any specified distribution (e.g., normality). If the *n* residuals in log space are denoted by r_i, and b is the base of logarithm used, the smearing correction factor, $\overline{C}_{bias}$ for the logarithmic transformation is given by:

Equation 6

$$\overline{C}_{bias} = \frac{1}{n} \sum_{i=1}^{n} b^{r_i}$$

Multiplying the right side of Equation 5 by Equation 6 almost removes the bias, so that:

Equation 7

$$\overline{UC} = \overline{C}_{bias} \overline{\beta_0} X^{\overline{\beta_1}}$$

The smearing correction factor $(\overline{C}_{bias})$ for our model was 1.25.

RESULTS

Table 1 shows the variable names, description, mean and standard error, estimated after combining the results of the five datasets of the multiple imputation estimates. Table 2 presents the results of the best-fit regression model. The adjusted R square of the combined regressions is 0.80, with an F statistic of 509 (p<0.0001), indicating that the model explains a

Table 1 Descriptive statistics of the multiple imputation estimates N=1171

Variable	Description	Mean	SE
Ln cost per bed day	Natural log of cost per bed day in 1998 I $	4.98	1.63
Ln GDP per capita	Natural log of GDP per capita in 1998 I $	8.90	1.06
Ln occupancy rate	Natural log of occupancy rate	-0.41	0.61
Drug costs	Dummy variable for inclusion of drug costs. Included =1	0.96	0.18
Food costs	Dummy variable for inclusion of food costs. Included =1	0.93	0.25
Level 1 hospital	Dummy variable for level 1 hospital [1]	0.33	0.47
Level 2 hospital	Dummy variable for level 2 hospital	0.41	0.49
Public	Dummy variable for level public hospitals [2]	0.84	0.36
Private for profit	Dummy variable for level private for profit hospitals	0.08	0.27
USA	Dummy variable for USA. USA =1	0.17	0.37

(1) Dummies for levels of hospital are compared with level 3 hospitals
(2) Dummies for hospital ownership are compared with public not-for-profit hospitals

Table 2 Multiple imputation regression coefficients and SE

Dependent variable: Natural log of cost per bed-day in 1998 I $ N=1171

Adjusted R^2= 0.80	F statistic = 509		p of F statistic <0.00001	
Variable	Coefficient	SE	T	P
Ln GDP per capita	0.7624	0.0295	25.813	<0.0001
Ln occupancy rate	-0.2318	0.0474	-4.886	<0.0001
Drug costs	0.6410	0.1769	3.624	<0.0001
Food costs	0.2116	0.1394	1.518	0.152
Level 1 hospital	-0.5777	0.0742	-7.787	<0.0001
Level 2 hospital	-0.3118	0.0594	-5.253	<0.0001
Public	-0.2722	0.1172	-2.323	0.021
Private for profit	0.2444	0.1316	1.857	0.064
USA	1.7471	0.1022	17.104	0.000
Constant	-2.5036	0.3264	-7.672	0.026

large part of the variation of the cost per bed-day across countries *(68)*. The signs of the coefficients are consistent with the earlier hypotheses. For example, the GDP per capita is positively correlated with cost per bed-day, while the lower the occupancy rate the higher is the cost per bed-day. Unit costs are lower in level-one hospitals than in those of levels two and three. The coefficients for the two main explanatory variables (GDP per capita and occupancy rate) are highly significant (p<0.0001), as well as most of the control dummies, e.g., hospital level. The coefficient for food costs is not significant at the 5% level but was included in the model because it added to its explanatory power.

   The tolerance test and its reciprocal variance inflation factors (VIF) showed no evidence for multicollinearity between the model variables (tolerance ranged between 0.20 and 0.89, mean VIF 1.97; tolerance less than 0.05 and VIF more than 20 indicate the presence of multicollinearity).

The only country dummy that was included in the final model was for the USA. The most plausible explanation for the positive, highly significant coefficient for the USA dummy is that USA was the only large data set where charges were reported rather than costs. In this case, the coefficient for the USA could be interpreted as a cost-to-charge ratio, estimated as 1:1.74. In other words, costs represent 57% of the charge on average. This is consistent with published national reports on the average cost-to-charge ratio for the USA such as that published by the United States General Accounting Office (63%) *(69)*.

Figure 1 shows the three regression lines of levels one, two and three hospitals, respectively, plotted against the log of GDP per capita (the Y-axis is log of cost per bed-day). The regression lines were estimated for public hospitals, with occupancy rate of 80%, including food costs and excluding drugs. Because the original data had a lower average occupancy rate (mean 71%, SD 39%), and most observations included drug costs, it is to be expected that the regression lines will be slightly lower than the actual data points in the database. The regression lines do not pass through the USA data points situated at the upper right side of the graph because they have been calculated for the case where the US dummy was set at zero.

Figure 1   Regression lines for level one, two and three hospitals against the natural log of GDP per capita
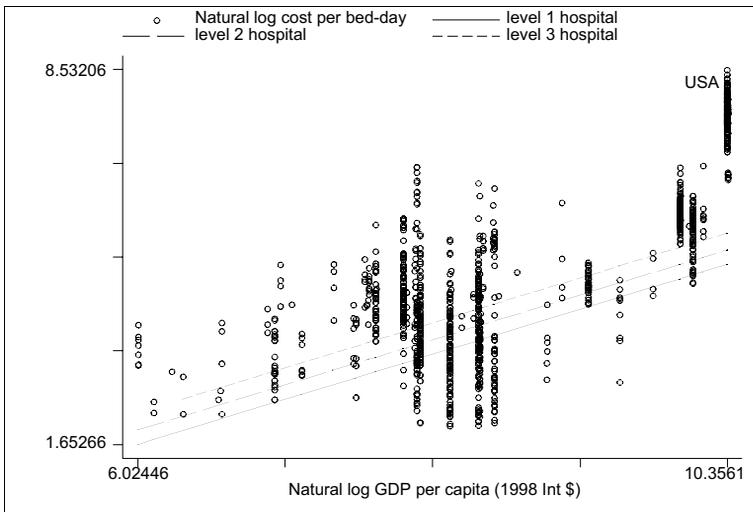
Overall, Figure 1 shows that the regression lines have a good fit with the data used to develop the model. They not only illustrate the relationship between cost per bed-day, hospital level and GDP per capita, but also show that there remains substantial variation in unit costs for any given level of GDP per capita. It would be inadvisable, therefore, to

base cost estimates on a single estimate of hospital costs in a particular setting, something that is a common feature of cost-effectiveness studies.

To use the equation reported in Table 2 to predict unit costs for a number of in and out-of-sample countries, with the appropriate uncertainty interval, requires consideration of the probability distributions of the predicted unit costs, given a specified level of the model variables. In order to derive these distributions, simulation techniques were used following the steps described in the Methods section. Table 3 presents for selected countries in different regions of the world the average simulated predicted values and 95% uncertainty intervals. The estimates are presented in 2000 I $, based on the 2000 GDP per capita in I $ and assuming that the estimated coefficients will remain constant over a short time period. They are specific to public hospitals, at an occupancy rate of 80%, excluding drug, but including food costs. Regional estimates of cost per bed day, with the same characteristics described above, are available from the WHO-CHOICE website: www.who.int\evidence\cea.

Table 3   Predicted cost per bed-day [1] in 2000 I $

| Country | GDP per capita (I$) | In or out-of-sample | Hospital level | Cost per bed day | | | |
				Mean (I $)	95% uncertainty interval low	95% uncertainty interval high	SD
Mali	581	Out	I	7.39	5.46	9.80	1.36
			II	9.64	7.07	12.73	1.74
			III	13.14	9.58	17.40	2.44
Mozambique	720	Out	I	8.70	6.45	11.43	1.58
			II	11.35	8.32	15.01	2.03
			III	15.46	11.31	20.42	2.85
Algeria	1,449	Out	I	14.82	10.95	19.45	2.65
			II	19.35	14.26	25.45	3.39
			III	26.34	19.41	34.63	4.72
Indonesia	3,167	Out	I	26.90	19.86	35.19	4.77
			II	35.12	26.05	46.12	6.11
			III	47.80	35.38	63.28	8.44
Ecuador	3,260	In	I	27.50	20.30	35.95	4.88
			II	35.90	26.63	47.17	6.25
			III	48.87	36.17	64.71	8.63
Romania	3,634	Out	I	29.88	22.05	39.11	5.30
			II	39.00	28.86	51.25	6.79
			III	53.09	39.23	70.35	9.37
Greece	6,192	Out	I	44.88	33.02	59.23	8.03
			II	58.58	43.25	76.84	10.28
			III	79.73	59.01	104.62	14.10
Russian Federation	8,035	In	I	54.75	40.17	72.43	9.86
			II	71.48	52.73	94.24	12.62
			III	97.27	71.79	127.92	17.27
Bahrain	14,159	Out	I	84.41	61.48	112.10	15.52
			II	110.19	80.27	146.22	19.85
			III	149.93	110.19	198.97	27.01
United Arab Emirates	20,330	Out	I	111.30	80.27	148.59	20.81
			II	145.28	105.96	191.58	26.62
			III	197.66	144.13	263.01	36.09
United Kingdom	24,348	In	I	127.76	91.83	170.79	24.11
			II	166.77	121.08	221.00	30.85
			III	226.87	164.90	302.91	41.76
Canada	28,087	Out	I	142.51	102.03	190.22	27.11
			II	186.02	134.55	247.23	34.69
			III	253.05	183.16	337.26	46.89

[1] Cost per bed day is estimated for public hospitals with 80% occupancy rate, excluding drug costs and including food costs

Discussion

This paper describes recent work on developing models to predict country-specific hospital unit costs, by level of hospital and ownership, for countries where these data are not available. The main purpose of this work was to feed into estimates of the costs and effects of many types of health interventions in different settings. Estimates are typically available for variables such as the number of days in hospital, or the number of outpatient visits, for certain types of interventions, but unit prices are not available for many countries. The model presented in this paper used all data on unit costs that could be collected after a thorough search to estimate costs for countries where this information does not exist. Data imputation techniques were used to impute missing data, which has the advantage of eliminating the bias introduced by list-wise deletion of observations in cases where information for some of the variables required by the model is missing.

The goodness-of-fit of the model was tested by various regression diagnostic techniques including the tolerance test for multicollinearity, adjusted R square and F statistic. All suggested a good fit of the model with the data and that GDP per capita could be used to capture different levels of technology use across countries. Although this is the first time that costs have been compared across countries, the signs of the coefficients are consistent with results from previous microeconomic studies within countries. For example, these studies have found that occupancy rate was negatively correlated with cost per bed-day while hospital level had the opposite relationship, something also found in the model presented in this paper *(70;71)*. This adds confidence to the estimated results.

In addition, the estimates produced by this model were sent to health economists and researchers in different countries to check their face validity. Experts from countries in all WHO regions, covering wide differences in GDP per capita and in technologies typically found in hospitals were consulted, including Benin, Canada, Ecuador, Egypt, Kenya, Netherlands and Thailand. They were provided with a description of the estimated unit cost (e.g., which costs were included) and were asked whether they thought they approximated the average cost per bed-day in their countries. All indicated that the results had face validity.

It is of particular note that the model incorporates a more extensive database on unit costs by hospital level and ownership than has previously been available. Increasing the range of observations will increase the validity of extrapolations of cost estimates for countries in which these data are not available. Additional sources of data are being sought for this purpose and to assist countries to develop their own studies. As this body of information grows, the predictive power of unit-cost models will continue to increase.

There are other possible uses of this model such as estimating the possible costs of scaling-up health interventions for the poor, which is receiving increasing attention with the activities of such bodies as the Global Fund to Fight AIDS, Tuberculosis and Malaria. This can be done in many ways, according to the objectives of the analysis. It may be used, for instance, to estimate:

– unit costs at different capacity levels for purposes of efficiency analysis or economic evaluation of health interventions;
– the "hotel" component of average cost per bed-day;
– unit costs, excluding specific items such as drugs or food costs.

Finally, it must be emphasized that there is wide variation in the unit costs estimated from studies within a particular country (Figure 1). These differences are sometimes of an order of magnitude, and cannot always be attributed to different methods. This implies that analysts cannot simply take the cost estimates from a single study in a country to guide their assessment of the cost-effectiveness of interventions, or the costs of scaling-up. In some cases, they could be wrong by an order of magnitude.

Conflict of interest

None.

Authors' contributions

TA was responsible for data collection, management and analysis, participated in the development of the methodology and drafted the manuscript. DE contributed to the development of the methodology, as well as data analysis and reporting. CM participated in the development and coordination of the methodology. All authors read and approved the final manuscript.

Acknowledgements

REFERENCES

1. Hutubessy RCW, Baltussen RMPM, Tan Torres Edejer T, Evans DB. Generalized cost-effectiveness analysis: an aid to decision making in health. *Applied Health Economics and Health Policy* 2002; 1(2):89-95.

2. Barnum H, Kutzin J. *Public hospitals in developing countries: resource use, cost, financing.* Baltimore: The Johns Hopkins University Press for the World Bank; 1993.

3. Breyer F. The specification of a hospital cost function. A comment on the recent literature. *Journal of Health Economics* 1987; 6(2):147-157.

4. Conrad RF, Strauss RP. A multiple-output multiple-input model of the hospital industry in North Carolina. *Applied Economics* 1983; 15:341-352.

5. Cremieux PY, Ouellette P. Omitted variable bias and hospital costs. *Journal of Health Economics* 2001; 20(2):271-282.

6. Grannemann TW, Brown RS, Pauly MV. Estimating hospital costs. A multiple-output analysis. *Journal of Health Economics* 1986; 5(2):107-127.

7. Lave JR, Lave LB. Hospital cost functions. *Annual Review of Public Health* 1984; 5:193-213.

8. Vitaliano DF. On the estimation of hospital cost functions. *Journal of Health Economics* 1987; 6(4):305-318.

9. Li T, Rosenman R. Estimating hospital costs with a generalized Leontief function. *Health Econ* 2001; 10(6):523-538.

10. Wagstaff A, Barnum H. *Hospital cost functions for developing countries.* Washington, D.C: World Bank; 1992. WPS 1044.

11. Lopez-Casasnovas G, Saez M. The impact of teaching status on average costs in Spanish hospitals. *Health Econ* 1999; 8(7):641-651.

12. Liu X, Hsiao WC. The cost escalation of social health insurance plans in China: its implication for public policy. *Social Science and Medicine* 1995; 41(8):1095-1101.

13. Newhouse JP. Medical care costs: how much welfare loss? *J Econ Perspect* 1992; 6(3):3-21.

14. Peden EA, Freeland MS. Insurance effects on US medical spending (1960-1993). *Health Econ* 1998; 7(8):671-687.

15. Garattini L, Giuliani G, Pagano E. A model for calculating costs of hospital wards: an Italian experience. *Journal of Management in Medicine* 1999; 13(2):71-82.

16. Kirigia JM, Snow RW, Fox-Rushby J, Mills A. The cost of treating paediatric malaria admissions and the potential impact of insecticide-treated mosquito nets on hospital expenditure. *Tropical Medicine & International Health* 1998; 3(2):145-150.

17. Mitchell M, Thomason J, Donaldson D, Garner P. The cost of rural health services in Papua New Guinea. *Papua and New Guinea Medical Journal* 1991; 34(4):276-284.

18. Patcharanarumol W. A *study of unit cost of out-patient and in-patient service of Khon Kaen Hospital in the fiscal year* 1996. Bangkok: Chulalongkorn University; 1997.

19. Huff-Rousselle M. *Dzongkhag costing study for Tashigang Dzongkhag.* Royal Government of Bhutan: Department of Health Services, Ministry of Social Services; 1992.

20. Rannan-Eliya R, Somanathan A. *Bangladesh facility efficiency survey.* Dhaka: Health Economics Unit, Ministry of Health and Family Welfare, Government of the People's Republic of Bangladesh and Health Policy Programme, Institute of Policy Studies of Sri Lanka; 1999. Working Paper No. 16.

21. Musau S, Kilonzo M, Newbrander W. *Development of a revised FIF user fee schedule. Report July 1996.* Boston: Management Sciences for Health. Kenya Health Care Financing Project: Contract No. 623-0245-C-00-0040-00. 2000.

22. Robertson RL. *Review of literature on costs of health services in developing countries.* Washington: Population, Health and Nutrition Department, World Bank. PHN Technical Note 85-21. 1985.

23. *Hospital services in Australia. Access and financing.* Canberra: Department of Health, Housing and Community Services; 1991. National Health Strategy Issues Paper No. 2.

24. Ojo K. *Economic and management perspectives of Windhoek state hospital complex.* Windhoek: Ministry of Health and Social Services; 1995.

25. Fox-Rushby JA, Foord F. Costs, effects and cost-effectiveness analysis of a mobile maternal health care service in West Kiang, The Gambia. *Health Policy* 1996; 35(2):123-143.

26. Brooks RG. Cost of selected health institutions in the Central Region 1972-1973. *Ghana Medical Journal* 1975; 14(3):209-214.

27. Shepard DS, Carrin G, Nyandagazi P. Household participation in financing of health care in government health centres in Rwanda. In: Lee K, Mills A, eds. *Health economics research in developing countries.* Oxford University Press; 1990. p. 140-164.

28. Flessa S. The costs of hospital services: a case study of Evangelical Lutheran Church hospitals in Tanzania. *Health Policy and Planning* 1998; 13(4):397-407.

29. Ministry of Health The Gambia, World Health Organization. *Cost analysis of the health care sector in The Gambia.* Volume 1. Ministry of Health. 1995.

30. Puglisi R, Bicknell WJ. *Functional expenditure analysis. Vol. I Final Report for Queen Elizabeth II Hospital, Maseru, Lesotho.* Boston University: Health Policy Institute. 1990.

31. Wong H. *Cost analysis of Niamey hospital.* Bethesda, MD: Abt Associates Inc. USAID Project No. 683-0254. 1989.

32. Ojo K, Foley J, Renner A, Kamara FM. *Cost analysis of health services in Sierra Leone. A case study of Connaught hospital and Waterloo Community Health Centre. Annex III.* Sierra Leone: Ministry of Health. 1995.

33. De Virgilio G, Haile M, Lemma A, Mariani D. Technical and economic efficiency of the Asella regional hospital in Ethiopia. *La Medicina Tropicale nelle Cooperazione allo Sviluppo* 1990; 6:1-7.

34. Omar AO, Komakech W, Hassan AH, Singh CH, Imoko J. Costs, resource utilisation and financing of public and private hospitals in Uganda. *East African Medical Journal* 1995; 72(9):591-598.

35. Hansen K, Chapman G, Chitsike I, Kasilo O, Mwaluko G. The costs of HIV/AIDS care at government hospitals in Zimbabwe. *Health Policy and Planning* 2000; 15(4):432-440.

36. Pepperall J, Garner P, Fox-Rushby J, Moji N, Harpham T. Hospital or health centre? A comparison of the costs and quality of urban outpatient services in Maseru, Lesotho. *The International Journal of Health Planning and Management* 1995; 10(1):59-71.

37. Carey K, Burgess JF, Jr. On measuring the hospital cost/quality trade-off. *Health Economics* 1999; 8(6):509-520.

38. Raymond SU, Lewis B, Meissner P, Norris J. *Financing and costs of health services in Belize.* Stony Brook: State University of New York. HCFLAC Research Report no. 2. 1987.

39. Lewis MA, La Forgia GM, Sulvetta MB. Measuring public hospital costs: empirical evidence from the Dominican Republic. *Social Science and Medicine* 1996; 43(2):221-234.

40. Olave M, Montano Z. *Unit cost and financial analysis for the hospital 12 de Abril in Bolivia.* Bethesda: Abt Associates Inc. Small Applied Research Report No. 11. 1993.

41. Gill L, Percy A. *Hospital costing study Glendon Hospital* – Montserrat. Organisation of Eastern Caribbean States, Health Policy & Management Unit. Report No.15. 1994.

42. Russell SS, Gwynne G, Trisolini M. *Health care financing in St Lucia and costs of Victoria Hospital.* Stony Brook: State University of New York. HCFLAC Research Report No. 5. 1988.

43. Snow J. *Papua New Guinea: health sector financing study project. Final Report – volume II hospital cost study.* Prepared for the Papua New Guinea Department of Health under contract with the Asian Development Bank, TA No. 1091-PNG. 1990.

44. Chan S. *Unit cost estimation for outpatient and inpatient departments in Nakleoung District Hospital, Cambodia.* Bangkok: Chulalongkorn University; 1997.

45. Banks DA, As-Sayaideh ASK, Shafei ARSH, Muhtash A. *Implementing hospital autonomy in Jordan: an economic cost analysis of Princess Raya Hospital.* Bethesda, MD: The Partners for Health Reformplus Project, Abt Associates Inc; 2002.

46. Department of Health. *The new NHS – 2001 reference costs.* London: Department of Health; 2001. (www.doh.gov.uk/nhsexec/refcosts.htm).

47. Department of Planning, Ministry of Health and Population, Data for Decision Making, Harvard School of Public Health, University of California, Berkeley, School of Public Health. *Cost analysis and efficiency indicators for health care: report number 1 summary output for Bani Suef General Hospital,* 1994-1994. Boston: Harvard University Press; 1997.

48. Department of Planning, Ministry of Health and Population, Data for Decision Making, Harvard School of Public Health, University of California, Berkeley, School of Public Health. *Cost analysis and efficiency indicators for health care: report number 2 summary output for Suez General Hospital,* 1993-1994. Boston: Harvard University Press; 1997.

49. Department of Planning, Ministry of Health and Population, Data for Decision Making, Harvard School of Public Health, University of California, Berkeley, School of Public Health. *Cost analysis and efficiency indicators for health care: report number 3 summary output for El Gamhuria General Hospital,* 1993-1994. Boston: Harvard University Press; 1997.

50. Department of Planning, Ministry of Health and Population, Data for Decision Making, Harvard School of Public Health, University of California, Berkeley, School of Public Health. *Cost analysis and efficiency indicators for health care: report number 4 summary output for 19 primary health care facilities in Alexandria, Bani Suef and Suez,* 1993-1994. Boston: Harvard University Press; 1997.

51. Waters H, Abdallah H, Santillan D. Application of activity-based costing (ABC) for a Peruvian NGO healthcare provider. *The International Journal of Health Planning and Management* 2001; 16(1):3-18.

52. Robertson RL, Barona B, Pabon R. Hospital cost accounting and analysis: the case of Candelaria. *Journal of Community Health* 1977; 3(1):61-79.

53. Jorbenadze A, Zoidze A, Gzirirshvili D, Gotsadze G. Health reform and hospital financing in Georgia. *Croatian Medical Journal* 1999; 40(2):221-236.

54. Mills AJ. *The cost of the district hospital. A case study from Malawi.* Washington, DC: World Bank; 1991. WPS 742.

55. StataCorp. *Stata Statistical Software: Release 7.0.* College Station, TX: Stata Corporation; 2001.

56. World Bank. *World Development Indicators 2000.* Washington, DC: World Bank; 2000.

57. King G, Honaker J, Joseph A, Scheve K. *List-wise deletion is evil: what to do about missing data in political science.* Paper presented at the Annual Meeting of the American Political Science Association, Boston; 1998.

58. Allison PD. Multiple imputation for missing data. A cautionary tale. *Sociological Methods & Research* 2000; 28(3):301-309.

59. Honaker J, Joseph A, King G, Scheve K, Singh N. *Amelia: A Program for Missing Data (Windows Version).* Cambridge, MA: Harvard University; 1999. (http://gking.harvard.edu).

60. Lu K, Tsiatis AA. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* 2001; 57(4):1191-1197.

61. Patrician PA. Multiple imputation for missing data. *Res Nurs Health* 2002; 25(1):76-84.

62. King G, Tomz M, Wittenberg J. Making the most of statistical analyses: improving interpretation and presentation. *American Journal of Political Science* 2000; 44:341-355.

63. King G, Honaker J, Joseph A, Scheve K. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review* 2000.

64. Rubin DB. *Multiple imputation for survey nonresponse*. New York: Wiley; 1986.

65. Cowing TG, Holtmann AG, Powers S. Hospital cost analysis: a survey and evaluation of recent studies. *Advances in Health Economics and Health Services Research* 1983; 4:257-303.

66. Gujarati DN. *Basic econometrics*. 3rd ed. New York: McGraw-Hill, Inc; 1995.

67. Duan N. Smearing estimate: a nonparametric retransformation method. *J Amer Statistical Assoc* 1983; 78(383):605-610.

68. King G. How not to lie with statistics: avoiding common mistakes in quantitative political science. *American Journal of Political Science* 1986; 30(3):666-687.

69. US General Accounting Office. Military health care. *Savings to CHAMPUS from using a prospective payment system*. Washington DC: United States General Accounting Office; 1990. Fact Sheet for Congressional Requesters GAO HRD-90-136FS.

70. Wouters A. The cost and efficiency of public and private health care facilities in Ogun State, Nigeria. *Health Economics* 1993; 2(1):31-42.

71. Anderson DL. A statistical cost function study of public general hospitals in Kenya. *Journal of Developing Areas* 1980; 14(2):223-235.

Annex 1   Definition of facility types as coded in the unit cost database

Facility type	Description
Primary-level hospital	Has few specialities, mainly internal medicine, obstetrics-gynecology, pediatrics, and general surgery, or only general practice; limited laboratory services are available for general but not for specialized pathological analysis; bed capacity ranges from 30 to 200 beds; often referred to as a district hospital or first-level referral.
Secondary-level hospital	Highly differentiated by function with five to ten clinical specialities; bed capacity ranging from 200-800 beds; often referred to as provincial hospital.
Tertiary-level hospital	Highly specialized staff and technical equipment, e.g., cardiology, ICU and specialized imaging units; clinical services are highly differentiated by function; may have teaching activities; bed capacity ranges from 300 to 1,500 beds; often referred to as central, regional or tertiary-level hospital.

Definition of hospital levels (adapted from Barnum and Kutzin 1993 (2))

Annex 2   Percentage of missing data in the model variables prior to data imputation

Variable name	Description	% missing
Ln GDP per capita	Natural log of GDP per capita in 1998 I $	0
Ln occupancy rate	Natural log of occupancy rate	48
Drug costs	Dummy variable for inclusion of drug costs. Included =1	3
Food costs	Dummy variable for inclusion of food costs. Included =1	19
Level 1 hospital	Dummy variable for level 1 hospital	16
Level 2 hospital	Dummy variable for level 2 hospital	16
Public	Dummy variable for level public hospitals	1
Private for profit	Dummy variable for level private for profit hospitals	1
USA	Dummy variable for USA to control for charges data. USA=1	0

Annex 3   Countries and number of unit cost observations included in
the model

Country	N	Country	N
Australia	64	Nepal	3
Bangladesh	21	Netherlands	1
Benin	1	New Zealand	4
Bolivia	1	Niger	2
Cambodia	1	Norway	6
China	367	Papua New Guinea	8
Colombia	1	Poland	4
Ecuador	70	Republic of Korea	32
Egypt	5	Russian Federation	22
Ethiopia	1	Rwanda	4
Ghana	2	Saint Lucia	1
Indonesia	5	Sri Lanka	93
Italy	2	Thailand	41
Jamaica	3	Turkey	1
Kenya	7	United Kingdom	176
Lebanon	4	United Republic of Tanzania	7
Malawi	2	United States of America	203
Mexico	2	Zimbabwe	2
Namibia	2	Total	1171

Observations range from 1973-2001, with 95% after 1990.

5   STOCHASTIC LEAGUE TABLES:
        COMMUNICATING COST-EFFECTIVENESS
        RESULTS TO DECISION-MAKERS[1]

RAYMOND C.W. HUTUBESSY, ROB M.P.M. BALTUSSEN, DAVID B.
EVANS, JAN J. BARENDREGT, CHRISTOPHER J.L. MURRAY

Summary

The presentation of the results of uncertainty analysis in cost-effectiveness analysis (CEA) in the literature has been relatively academic with little attention paid to the question of how decision-makers should interpret the information particularly when confidence intervals overlap. This question is especially relevant to sectoral CEA providing information on the costs and effects of a wide range of interventions.

This paper introduces stochastic league tables to inform decision-makers about the probability that a specific intervention would be included in the optimal mix of interventions for various levels of resource availability, taking into account the uncertainty surrounding costs and effectiveness. This information helps decision-makers decide on the relative attractiveness of different intervention mixes, and also on the implications for trading gains in efficiency for gains in other goals such as reducing health inequalities and increasing health system responsiveness.

Key words: Cost-effectiveness analysis; decision-making analysis; uncertainty analysis.

Introduction

Uncertainty in cost-effectiveness analysis (CEA) has received a lot of attention in recent years, leading to the development of a range of approaches, such as non-parametric bootstrapping (*1*), the construction of confidence planes (*2*), mathematical techniques (*3*), probabilistic sensitivity analyses using Monte Carlo simulations (*4*), and the net health benefit approach (*5*). These techniques all present study results in terms of some type of uncertainty interval. However, little or no attention is paid to the question of how decision-makers should interpret the results where uncertainty intervals overlap.

This absence of guidance to decision-makers is exacerbated in sectoral CEA based on the implicit or explicit use of cost-effectiveness league tables (*6;7*). Sectoral analysis requires that interventions are ranked on the basis of their cost-effectiveness ratios. In deterministic analysis, decision-makers are assumed to work down the list, starting with the most cost-effective, and to stop funding interventions when the resources run out. The addition of uncertainty to this analysis is more realistic, but uncertainty intervals of many of the ratios may overlap and the decision maker is left with no guidance in the literature. It is simply assumed that no decision about which intervention is more efficient can be made. Yet, decision-makers must and do make decisions about which interventions to encourage even when uncertainty is high (e.g. with overlapping confidence intervals).

We propose a new approach to presenting decision-makers with the results of CEA including uncertainty through the construction of a "stochastic league table". This informs decision-makers about the probability that a specific intervention would be included in the optimal mix of interventions for various levels or resource availability, taking into account the uncertainty surrounding its total costs and effectiveness. Each intervention should be thought of as a national programme or policy, which can only be purchased at one point (*8*). Although the argument is presented with reference to the generalized method for sectoral CEA, which we recently proposed, allowing decision-makers to assess the efficiency of the current mix of interventions as well as the relative attractiveness of changes to this mix should new resources become available (*8*), it is applicable to any form of sectoral analysis.

The analytical Framework

The construction of stochastic league tables requires four steps (a software program, MCLeague, is being developed to carry out this process). Firstly, using Monte Carlo simulations, random draws are taken from estimated distributions of total costs and effects for the interventions under study. These distributions are *a priori* defined by the analyst and may take different forms, for example normal, log-normal, and uniform

distributions (4). Table 1 presents the hypothetical costs and effect data first presented in Murray et al. (8). To reflect uncertainty, costs are here assumed to be log-normally distributed, with S.D. of 20, and effects are assumed to be normally distributed with an S.D. of 2. The covariance is assumed to be zero. The conclusions are not dependent on these assumptions. Random draws are taken from these distributions for all interventions.

The second step is to determine the optimal mix of interventions for given levels of resource availability following the procedure for choosing between mutually exclusive and independent interventions outlined in Murray et al. (8). The most efficient intervention in the set of mutually exclusive interventions is evaluated according to its average cost-effectiveness ratio (versus doing nothing), while the cost-effectiveness of others in the mutually exclusive set are evaluated incremental to the most efficient intervention.

Thirdly, this process is repeated a large number of times (here 10000) to provide 10000 estimates of the optimal mix of interventions. If P equals the number of times that an intervention is included in the optimal mix, P/10000 is the probability that the intervention is included. Hence, P is the proportion of samples from the estimated distribution for which the intervention is estimated to be optimal based on the sample average and incremental cost-effectiveness ratios. In our example, for resources equal to 50, C1 is included 4323 times, a 43% probability of being included (Table 2). P for C2 equals 1406, a probability of inclusion of 14%. In the remaining cases (43% of all random draws), costs of each possible option overrun the available resources and no intervention can be funded fully. This explains why the probabilities do not add up to 100%.

Table 1   Costs, effects and cost-effectiveness of three independent sets of mutually exclusive alternatives

Interventions	Total costs		Total effects		Cost-effectiveness[a]
	Mean	S.D.	Mean	S.D.	
A1	120	20	1	2	-
A2	140	20	5.5	2	25.4
A3	170	20	3	2	-
A4	190	20	7	2	33.3
B1	100	20	12	2	-
B2	120	20	17	2	7.1
B3	150	20	20	2	10.0
C1	50	20	22	2	2.3
C2	70	20	24.5	2	8.0
C3	120	20	29	2	11.1
C4	170	20	31	2	25.0

[a] Cost-effectiveness ratios after exclusion of dominated interventions

The fourth step involves repeating this procedure for various levels of resource availability to reveal the "resource expansion path", showing the probability that each intervention will be included at different levels of resource availability (Table 2). Decision-makers can use this information to prioritize interventions should more resources become available for health care. The probability that a more expensive alternative will be included increases with the level of resource availability. For example, the probability C2 is included increases from 14% to 47% when resources increase from 50 to 100. In our example, no intervention is included in the optimal mix with certainty – even at high levels of resource availability – because of the relative large standard deviations assumed for costs and effects.

The degree of uncertainty in costs and effects of an intervention can have a large impact on its probability of inclusion in the optimal mix. If we change the standard deviation of the cost distribution for intervention A2 from 20 to 70, its probability of inclusion at a level of resource availability of 300 increases from 5% to 22% (Table 3). This is because intervention costs now are sometimes very low thereby rendering the intervention relatively cost-effective (with resources <600, its probability of inclusion decreases because it now has to compete with the more cost-effective interventions A3 and A4 which can be afforded). The general conclusion is that the higher the uncertainty in costs and effects, the more equal the probabilities of inclusion of interventions will be, other things equal. This is true both within the same mutual exclusive set as well as between independent sets of interventions.

Table 2   Stochastic league table presenting the probability of inclusion (%) of three independent sets of mutual exclusive interventions in the optimal mix of interventions at different levels of resource availability[a]

Interventions	Resource availability							
	50	100	150	200	300	400	600	800
A1	0	0	0	0	1	2	0	0
A2	0	0	1	1	5	25	28	29
A3	0	0	0	0	0	1	4	4
A4	0	0	0	0	0	11	**67**	67
B1	0	3	14	11	2	0	0	0
B2	0	1	18	**35**	26	21	15	15
B3	0	0	4	36	**72**	**78**	**85**	85
C1	**43**	40	30	36	8	3	0	0
C2	14	**47**	25	**37**	17	10	0	0
C3	0	13	**38**	16	**45**	33	24	23
C4	0	0	8	11	31	**54**	**76**	76

[a] Numbers in bold represent interventions that would be listed in a traditional league table

In Table 2, the numbers in bold represent interventions that would be selected in a traditional league table based on the cost-effectiveness ratios calculated in Table 1. These interventions would also be chosen by the stochastic league table because of their higher probabilities of inclusion. However, the stochastic league table provides additional information to the decision maker. With resources of 200, a traditional league table would choose intervention C2 whereas our stochastic league table shows almost identical probabilities of inclusion of C1 and C2 in the optimal mix of interventions. This information provides decision-makers with more information than simply presenting the confidence intervals for all CERs. For example, it allows decision-makers to better evaluate the impact of trading off the efficiency goal against other objectives such as reducing health inequalities in their selection of interventions (9). In general, the more interventions (belonging to the same mutually exclusive set) differ regarding their probabilities of inclusion in the optimal mix, the more efficiency decision-makers give up if they choose to over-ride the results in favour of other goals in their choice of interventions – the stochastic league table in our example informs decision-makers that they are not likely to lose much in terms of efficiency if they decide to select C1 rather than C2 for equity reasons. This important information is not revealed in deterministic league tables or in the traditional approach to uncertainty in CEA.

Another advantage concerns the information provided in the expansion path, illustrated in Table 2. With resources of 200, there is little to choose between B2 and B3 but preference would be given to B2. However,

Table 3   As Table 2, with standard deviation for costs of intervention A2 increased from 20 to 70[a]

Interventions	Resource availability							
	50	100	150	200	300	400	600	800
A1	0	0	0	0	1	2	1	1
A2	7	9	14	12	22	37	28	28
A3	0	0	0	0	0	1	5	4
A4	0	0	0	0	1	9	**67**	67
B1	0	3	15	10	1	1	0	0
B2	0	1	19	**35**	28	18	15	14
B3	0	0	3	37	**71**	**81**	**85**	**86**
C1	**43**	40	30	39	11	3	0	0
C2	15	**47**	30	**33**	17	7	1	1
C3	0	13	**34**	17	**45**	30	23	24
C4	0	0	6	11	28	**60**	77	75

[a] Numbers in bold represent interventions that would be listed in a traditional league table

if the decision maker felt that additional resources would become available in the near future, and that the costs of switching from B2 to B3 might be substantial, it would be sensible for them to choose B3. Again, this type of information is not provided in the standard approach to uncertainty.

Stochastic league tables may also show that interventions that would otherwise have been ruled out by dominance in traditional league tables might well be included in some draws. In our example, intervention B1 will never be eligible for selection in a deterministic league table because it is (weakly) dominated by B2. However, taking into account uncertainty the stochastic league table (Table 2) shows that B1 has a low but non-zero probability of being included in the optimal mix. Whether decision-makers will actually select such interventions depends on the probability of inclusion compared to other mutually exclusive alternatives, and the trade-off between efficiency and other objectives of health systems.

Figure 1 depicts an alternative way of visualizing the information of Table 2. The vertical axis shows the probability of being chosen at the level of resource availability on the horizontal axis. The logic is the same as that described for the interpretation of the tables.

DISCUSSION

The stochastic league table developed in this paper is a new way of presenting uncertainty around costs and effects to decision-makers. It provides additional information beyond that offered by the traditional treatment of uncertainty in CEA, presenting the probability that each intervention is included in the optimal mix for given levels of resource

Figure 1   Probability of inclusion (%) of three independent sets of mutually exclusive interventions

availability. The most likely optimal mix will be the one that contributes the most to maximizing population health for that level of resources. Decision-makers can then decide the extent to which they should trade off gains in efficiency for gains in other goals of the health system.

Stochastic league tables are conceptually different from the recently suggested portfolio approach, borrowed from financial economics and characterizing health care resources allocation as a risky investment problem (10). This approach provides the optimal intervention mix given decision-makers' explicit preferences concerning risk and return. Our stochastic league table provides the probability of an intervention being chosen in the optimal mix, given uncertainty. Risk-neutral decision-makers would choose the most likely combination of interventions.

A drawback to our framework (and to the portfolio approach for that matter) is that distributions of costs and effects are assumed to be independent, e.g. no joint distributions are defined. Moreover, the definition of the distributions is left to the analyst, who may have very little information about the actual distribution, but whose choice is likely to have a large effects on the results. It is technically possible to include covariance between costs and outcomes in the analysis, but this requires more information about covariances than is usually available. Alternatively, where empirical data on patient' costs and effects are available, our framework could employ the technique of non-parametric bootstrapping in which samples are drawn with replacement from the original data. This approach has the advantage that is does not rely on parametric assumptions concerning the underlying distribution and that covariances between costs and effects can be easily incorporated (1). The development of stochastic league tables is an important step forward in the interpretation of uncertainty at the decision making level.

ACKNOWLEDGEMENTS

REFERENCES

1. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ* 1997; 6(4):327–340.

2. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ 1998; 7(8):723–740.

3. Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Econ* 1996; 5(4):297–305.

4. Lord J, Asante MA. Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis. *Health Econ* 1999; 8(4):323–333.

5.   Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998; 18(2 Suppl):S68–S80.

6.   Birch S, Gafni A. Cost-effectiveness ratios: in a league of their own [see comments]. *Health Policy* 1994; 28(2):133–141.

7.   Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? *Soc Sci Med* 1993; 37(1):33–40.

8.   Murray CJ, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Econ* 2000; 9(3):235–251.

9.   Murray CJ, Frenk J. A framework for assessing the performance of health systems. *Bull World Health Organ* 2000; 78(6):717–731.

10.   O'Brien BJ, Sculpher MJ. Building uncertainty into cost-effectiveness rankings: portfolio risk-return tradeoffs and implications for decision rules. *Med Care* 2000; 38(5):460–468.

6   UNCERTAINTY IN COST-EFFECTIVENESS ANALYSIS: PROBABILISTIC UNCERTAINTY ANALYSIS AND STOCHASTIC LEAGUE TABLES[1]

ROB M.P.M. BALTUSSEN, RAYMOND C.W. HUTUBESSY,
DAVID B. EVANS, AND CHRISTOPHER J.L. MURRAY

Abstract

Interest is growing in the application of standard statistical inferential techniques to the calculation of cost-effectiveness ratios (CER), but individual-level data will not be available in many cases because it is very difficult to undertake prospective controlled trials of many public health interventions. We propose the application of probabilistic uncertainty analysis using Monte Carlo simulations, in combination with nonparametric bootstrapping techniques where appropriate. This paper also discusses how decision-makers should interpret the CER of interventions where uncertainty intervals overlap. We show how the incorporation of uncertainty around costs and effects of interventions into a stochastic league table provides additional information to decision-makers for priority setting. Stochastic league tables inform decision-makers about the probability that a specific intervention would be included in the optimal mix of interventions for different resource levels, given the uncertainty surrounding the interventions.

Key words: cost-effectiveness, uncertainty analysis, priority setting

As more prospective cost-effectiveness analysis (CEA) studies are undertaken, providing stochastic data on costs and effects, interest has grown in the application of statistical techniques to the calculation of cost-effectiveness ratios (CER). Several methods have been developed, including confidence planes (*1*) mathematical techniques (*2*), and the net health benefit approach (*3*).

However, it is important to recognize that many public health interventions do not lend themselves to the collection of sampled individual-level data (by patient, health facility, region, etc.), especially in a developing-country context. For example, it is difficult to develop a feasible experimental design to identify the costs and effects of a national radio health education programme, or a policy to subsidize the use of essential pharmaceutical products. Many economic evaluations require nonstochastic parameter estimates and modeling assumptions.

Typically, uncertainty stemming from the use of such nonsampled secondary data sources in CEA has been dealt with by sensitivity analysis (*4;5*). These deterministic analyses draw inferences from point estimates of variables, but interpretation is conditional upon a range of uncertainty that is assumed for critical variables. There are three major limitations to this approach: a) the analyst has discretion as to which variables and what alternative values are included; b) interpretation is essentially arbitrary because there are no comprehensive guidelines or standards as to what degree of variation in results is acceptable evidence that the analysis is robust; and c) variation of uncertain parameters one at a time carries a risk that interactions between parameters may not be captured (*6*).

This paper examines the application of probabilistic uncertainty analysis with Monte Carlo simulations in this context.[1] This builds on work already described in the literature (*4;7–9*) and requires that analysts assume some distributional form for costs and effects from which repeated samples are drawn to determine a distribution for the CER. The definition of an uncertainty range for CER is hampered by the instability of sample estimates of CERs, causing its mean to vary (*10*). This paper applies the simple percentile method – usually employed to estimate uncertainty ranges for CER in nonparametric bootstrapping – to estimate uncertainty intervals for CERs involving probabilistic uncertainty analysis. The approach is illustrated by constructing uncertainty intervals for seven hypothetical interventions in tuberculosis control.

In addition, this paper discusses how information on uncertainty should be communicated to policy-makers. The above-mentioned techniques present study results in terms of some type of uncertainty interval. However, little or no attention is paid to the question of how decision-makers should interpret the results where uncertainty intervals overlap. The recently developed stochastic league tables inform decision-makers about the probability that a specific intervention would be included in the optimal mix of interventions for various levels or resource availability, taking into account the uncertainty surrounding costs and effectiveness

(*11*). This paper derives a stochastic league table for hypothetical interventions in tuberculosis control. We show that the incorporation of uncertainty ranges for CERs in a stochastic league table provides additional information to decision-makers for priority setting.

This methodologic work on estimating uncertainty is part of the larger World Health Organization (WHO) concept of Generalized-CEA (*12*). WHO proposes to provide policy-makers with a simple set of results that are more generalizable across settings by evaluating the costs and effectiveness of new and existing interventions, compared to the starting point of doing none of the current interventions, called the "null". This removes the constraint that the current intervention mix must be continued and eliminates differences in starting points, which traditionally makes the results of incremental analyses difficult to transfer across settings.

PROBABILISTIC UNCERTAINTY ANALYSIS BY MONTE CARLO SIMULATIONS

Probabilistic uncertainty analysis using Monte Carlo simulations has been well described elsewhere (*4;7–9*). Most applications assume a distributional form (e.g., normal, uniform, binominal) for each estimated (but nonsampled) variable. Repeated samples are then drawn from these distributions to determine an empirical distribution for some construct of the variables, such as CERs.

Table I   Costs and effects of interventions

Intervention	Costs	Effects
B1	180	200
B2	325	300
B3	600	400
A	550	500
AB1	631	600
AB2	726	650
AB3	952	700

To illustrate the procedure, consider a hypothetical example first presented in Murray et al.(*12*) related to four interventions for tuberculosis: a) passive case detection and treatment with directly observed short-course therapy (DOTS) (A); b) bacille Calmette-Guérin (BCG) vaccination at 50% coverage (B1); c) BCG vaccination at 75% coverage (B2); and d) BCG vaccination at 100% coverage (B3). In addition, three other mutually exclusive options are presented: DOTS combined with the different levels of BCG coverage, i.e. AB1, AB2, and AB3, respectively. Costs and health effects interact: the variable costs component of DOTS

Figure 1   Costs and health effects of interventions
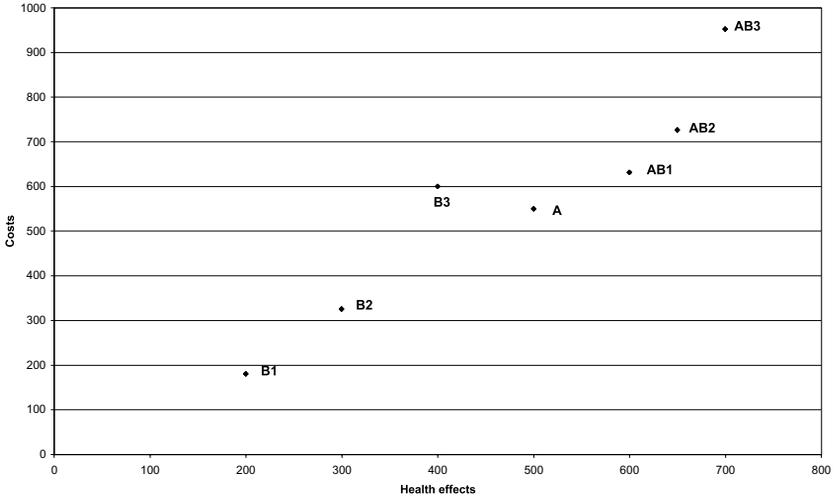
decreases when the vaccination is given, and fewer cases of tuberculosis will occur. The health benefits of BCG vaccination will be fewer in the presence of a treatment programme, because many of the deaths from tuberculosis expected in the absence of treatment will be avoided. Total costs and health effects of the interventions at the population level are presented in Table 1 and Figure 1. To reflect uncertainty, costs are assumed to be normally distributed with standard deviation of 100; health effects are assumed to be normally distributed with a standard deviation of 100. The covariance is assumed to be zero.

The procedure to generate a sample distribution for the incremental CER from expanding the intervention, for example, BCG coverage from 50% to 75% (B1$\rightarrow$B2), is as follows:

1. Take one sample of costs (C) and health effects (E) from the distribution of costs and effects from B1: C_{B1} and E_{B1}, and one sample of cost and effects from the distribution of costs and effects of B2: C_{B2} and E_{B2};
2. The sample estimate of the incremental CER is then given by C_{B2}-C_{B1} divided by E_{B2}-E_{B1}; and
3. Repeating this process a large number of times gives a vector of sample estimates that is the empirical sampling distribution of the incremental CER statistic.

We used the statistical program @RISK 4.0™ (Palisade Decision Tools) to run the analyses.

There is little stability in these CER estimates where the distributions of costs or health effects overlap – some simulations will produce negative net health effects and some positive net health effects, for example.

Figure 2   Mean incremental CERs of three interventions as a function of the number of samples
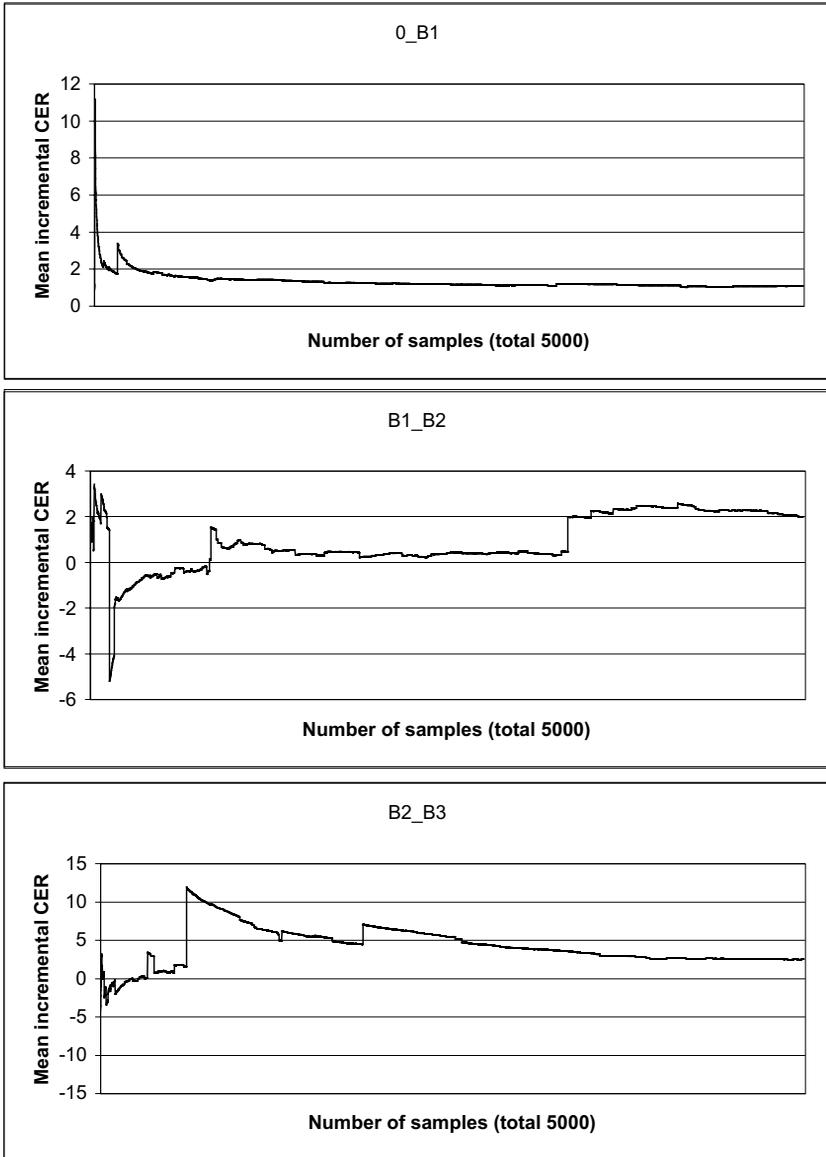
This can lead to positive or negative incremental CERs. Figure 2 shows the mean value of the sampled estimates of the incremental CERs for 0→B1, B1→B2, and B2→B3. For the assumed ranges of costs and heath effects, even after a large number of samples, some means do not stabilize. The mean CER of 0→B1 is relatively stable because the origin is fixed, and costs and health effects of intervention B1 constitute its only source of uncertainty (i.e., it is an average ratio).

The simple percentile method allows us to estimate confidence intervals in the presence of these unstable means. This approach takes the $100(\alpha/2)$ and the $100(1-(\alpha/2))$ percentile values of the bootstrap distribution as the upper and lower confidence limits for the CER (*10*). Table 2 shows the 90% confidence intervals for the seven mutually exclusive alternatives. The occasional very high values for the incremental CER of the expansion from AB1→AB2 is the reason why its confidence interval does not include its mean value.

Table 2   Sample incremental CERs for seven interventions

Intervention	Mean	Minimum	Maximum	90% confidence interval	
0→B1	1	-379	340	0	4
B1→B2	2	-912	5131	-8	8
B2→B3	3	-961	6358	-11	13
B3→A	-1	-1046	635	-5	5
A→AB1	1	-4469	1727	-5	6
AB1→AB2	21	-20545	109129	-7	7
AB2→AB3	-4	-30827	6662	-11	12

Of special interest are interventions that are weakly dominated on the basis of the point estimate of their CER but have a wide confidence interval. In such cases, some simulations might show them to be no longer dominated. In Figure 1, consider intervention B3. Because its mean is located north-west of intervention A, it appears to be strongly dominated. However, the uncertainty range of the incremental CER of B3→A ranges from -\$5 to \$5 per unit of health effect and thus includes positive values. Therefore, we cannot be sure that B3 should be excluded from the set of alternatives under consideration.

COMBINING PROBABILISTIC UNCERTAINTY ANALYSIS WITH NONPARAMETRIC BOOTSTRAP PROCEDURES

In the situation in which individual level data are available for some component of costs or effects, one feasible approach is to combine probabilistic uncertainty analysis with nonparametric bootstrapping to estimate a total "uncertainty range"[2] for CERs (*9*). The use of nonparametric bootstrapping has been advocated by many authors (*10;13–15*) and has been extensively applied to empirical data (*2;7;10;13;15–22*). Unlike probabilistic uncertainty analyses, the bootstrap approach is a nonparametric method that makes no distributional assumptions concerning the statistic in question. Instead it employs the original data in a resampling exercise in order to give an empirical estimate of the sampling distribution of that estimate.

The basic concept behind nonparametric bootstrapping is to treat the study sample as if it were the population, the premise being that it is better to draw inferences from the sample in hand rather than make potentially unrealistic assumptions about the underlying population. Using the

nonparametric bootstrap approach, successive random draws are taken *with replacement* from the study sample data As such, the fact that an observation has been selected does not preclude it from being selected again for the same resample, which leads to the construction of different bootstrap resamples. The statistic of interest and its distribution is calculated from these resamples. The number of bootstrap resamples, B, should at least be 1,000 to construct confidence intervals, in order to ensure that the tails of the empirical distributions are filled (*10*). An important advantage of the nonparametric bootstrap approach is that it is of no consequence whether the original sample is a well-behaved distribution because it forms its own probability density function.

To illustrate the combination of probabilistic uncertainty analysis and nonparametric bootstrapping, consider a CEA with costs being the product of vectors of unit prices and resource utilization. By defining a probability distribution of unit prices, and with resource utilization and effectiveness data stemming from sampled data, a total uncertainty range can be estimated by combining probabilistic uncertainty analysis with nonparametric bootstrapping. To start with, a large number (B) of samples of size n_p of sets of unit prices are obtained by random sampling from the prior distributions, and the mean price is calculated for each of the B samples. Similarly, B bootstrap samples of size n_q are taken from the resource utilization and effectiveness data, and the mean resource utilization and effectiveness is estimated for each of the B bootstrap samples. Then B replicates of the CER can be obtained by combining B bootstraps of both resource utilization and effectiveness data with the B sets of prices sampled from the prior distributions. These are then used to calculate a percentile interval.

UNCERTAINTY AT THE ALLOCATION LEVEL

Traditionally, the above results, as reported in Table 2, are placed in a single league table to inform decision-makers about the relative value of a set of (mutual exclusive) interventions. Rank ordering in the league table approach is usually made on the basis of point estimates of CE alone (*23*). However, in our example of tuberculosis control, uncertainty intervals overlap and it is not clear how decision-makers should interpret such information. This problem was also faced in many practical situations, including the study by Goodman and Mills (*24*), which incorporated the estimated uncertainty interval for their estimates of the cost-effectiveness of interventions against malaria; however, because the intervals overlapped, the authors were unwilling to suggest which ones should be given preference in the event of a shortage of resources (*24*). We believe that additional information is obtained in the data used to produce the uncertainty intervals, which could be used to guide policy-makers more than by simply saying that no decision could be made because confidence intervals overlap.

We propose the use of stochastic league tables. The approach provides the probability of inclusion of a specific intervention in the optimal mix of interventions given the uncertainty surrounding the intervention. The construction of stochastic league tables requires four steps and is described in more detail elsewhere (*11*). In a first step, CERs are calculated for the respective programmes by drawing single samples from distributions of both costs and health effects, using Monte Carlo simulations. Second, based on these samples, the optimal mix of interventions is defined, applying resource allocation decision rules as described in Murray et al.(*12*). Third, this exercise is repeated 10,000 times to obtain a distribution of rank orders of interventions, given a certain resource level. This provides information on the probability of the cost-effectiveness of interventions. Fourth, this procedure is repeated for various resource levels. This provides a so-called "budget expansion path", which shows that different interventions will be chosen at different resource levels.

Table 3 summarizes the results as probabilities (in percentages) that a particular intervention in tuberculosis would be included in the optimal set at different resource availability levels. Probabilities of inclusion of an intervention depend on the resource availability, its costs, and its relative cost-effectiveness. In our example, at a resource level of 100, intervention B1 is chosen in 19% of all cases (i.e., costs of B1 are less than 100 in 19% of the cases, and other interventions are too costly to be chosen). If resource availability increases, the probability of inclusion of other interventions also increases. Note that intervention B3, which would not be considered in a deterministic approach because of strong dominance, is now chosen in a low number of cases at certain resource levels. At the highest resource level (1200), AB3 is chosen in 53% of all cases. Decision-makers can use this information to prioritize interventions should more resources become available for health care.

Stochastic league tables present decision-makers the probability that an intervention will be included in the optimal choice and are therefore more informative than traditional league tables, which simply present uncertainty ranges (and may leave decision-makers indecisive when they overlap). They also allow decision-makers to better evaluate the impact of trading off the efficiency goal against other objectives such as reducing health inequalities in their selection of interventions. For example, the stochastic league table informs decision-makers that they are not likely to lose much in terms of efficiency if, at a resource level of 1000, they decide to select AB2 rather than AB3 for equity reasons.

DISCUSSION

This paper presents an extension and generalization of previously described methods examining uncertainty in cost-effectiveness studies. Whereas previous studies applied the concept of bootstrapping and Monte Carlo simulations in the contexts of clinical trials, here we apply

Table 3 Stochastic league table, with probabilities of inclusion of interventions (%) for different resource availability

Intervention	Resource level											
	100	200	300	400	500	600	700	800	900	1000	1100	1200
B1	19	64	57	36	15	3	0	0	0	0	0	0
B2	0	6	36	58	42	10	1	0	0	0	0	0
B3	0	0	0	1	8	12	5	2	1	1	0	0
A	0	0	0	5	27	38	22	11	6	4	3	3
AB1	0	0	0	0	7	31	43	37	27	20	16	16
AB2	0	0	0	0	0	7	29	47	48	37	32	29
AB3	0	0	0	0	0	0	0	3	17	38	48	53

them to decision models, analyzing cost-effectiveness based on any combination of primary and secondary data. Given the prevailing scarcity of sampled data on costs and effects of many public health interventions in developing countries, we propose the use of probabilistic uncertainty analysis using Monte Carlo simulations, in combination with nonparametric bootstrapping techniques where appropriate.

The reporting of some type of uncertainty range of CER in individual studies ignores the question of how policy-makers should interpret the results where uncertainty results overlap. The paper has shown that cost-effectiveness uncertainty ranges of interventions in tuberculosis control overlap and that decision-making is difficult in such a situation. The stochastic league table is a new way of presenting uncertainty around costs and effects to decision-makers. This paper shows that it provides additional information beyond that offered by the traditional treatment of uncertainty in CEA, presenting the probability that each intervention is included in the optimal mix for given levels of resource availability.

NOTES

1. There is confusion in the literature as to the definition of sensitivity analysis, on the one hand, and uncertainty analysis on the other. We argue that sensitivity analysis refers to uncertainty about social choices, such as the discount rate or the inclusion of productivity costs. Uncertainty analysis refers to variation in the distribution of costs and effects (stemming from either nonsampled or sampled data). Following that definition, we prefer to use the term *probabilistic uncertainty analysis* rather than *probabilistic sensitivity analysis* to describe the process of drawing repeated samples from nonsampled data, i.e. from some *a priori* defined distributional form of costs and/or effects. We use the term *nonparametric bootstrapping* only in relation to drawing samples from sampled data.

2. Instead of calling this a confidence interval, the term *uncertainty range* could be used since such an interval incorporates both uncertainty related to sampled and nonsampled data.

REFERENCES

1. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 1994; 32(2):150–163.

2. Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Economics* 1997; 6(3):243–252.

3. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Economics* 1999; 8(3):257–261.

4. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Economics* 1994; 3(2):95–104.

5. Briggs A, Sculpher M. Sensitivity analysis in economic evaluation: a review of published studies. *Health Economics* 1995; 4(5):355–371.

6. O'Brien BJ, Drummond MF. Statistical versus quantitative significance in the socioeconomic evaluation of medicines. *PharmacoEconomics* 1994; 5(5): 389–398.

7. Critchfield GC, Willard KE, Connelly DP. Probabilistic sensitivity analysis methods for general decision models. *Computers and Biomedical Research* 1986; 19(3):254–265.

8. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. *Medical Decision Making* 1985; 5(2):157–177.

9. Lord J, Asante MA. Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis. *Health Economics* 1999; 8(4):323–333.

10. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* 1997; 6(4):327–340.

11. Hutubessy RC, Baltussen RM, Evans DB, Barendregt JJ, Murray CJ. Stochastic league tables: communicating cost-effectiveness results to decision-makers. *Health Economics* 2001; 10(5):473–477.

12. Murray CJ, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Economics* 2000; 9(3):235–251.

13. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Statistics in Medicine* 1996; 15(13):1447–1458.

14. Manning WG, Jr, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russel LB, Weinstein MC, eds. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996. p. 247–275.

15. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999; 3(2):1–134.

16. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Economics* 1998; 7(8):723–740.

17. Campbell MK, Torgerson DJ. Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. *QJM* 1999; 92(3):177–182.

18. Hunink MG, Bult JR, de Vries J, Weinstein MC. Uncertainty in decision models analyzing cost-effectiveness: the joint distribution of incremental costs and effectiveness evaluated with a nonparametric bootstrap method. *Medical Decision Making* 1998; 18(3):337–346.

19. Mennemeyer ST, Cyr LP. A bootstrap approach to medical decision analysis. *Journal of Health Economics* 1997; 16(6):741–747.

20. Obenchain RL, Melfi CA, Crompton DW, Bueshing DP. Bootstrap analyses of cost-effectiveness in antidepressant pharmacotherapy. *PharmacoEconomics* 1997; 11:464–472.

21. Severens JL, De Boo TM, Konst EM. Uncertainty of incremental cost-effectiveness ratios. A comparison of Fieller and bootstrap confidence intervals. *International Journal of Technology Assessment in Health Care* 1999; 15(3):608–614.

22. Tambour M, Zethraeus N. Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results [comment]. *Health Economics* 1998; 7(2): 143–147.

23. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? *Social Science and Medicine* 1993; 37(1):33–40.

24. Goodman CA, Mills AJ. The evidence base on the cost-effectiveness of malaria control measures in Africa. *Health Policy and Planning* 1999; 14(4):301–312.

7 Effectiveness and costs of interventions to lower systolic blood pressure and cholesterol: a global and regional analysis on reduction of cardiovascular-disease risk[1]

Christopher J.L. Murray, Jeremy A. Lauer, Raymond C.W. Hutubessy, Louis Niessen, Niels Tomijima, Anthony Rodgers, Carlene M.M. Lawes, David B. Evans

Summary

Background: Cardiovascular disease accounts for much morbidity and mortality in developed countries and is becoming increasingly important in less developed regions. Systolic blood pressure above 115 mm Hg accounts for two-thirds of strokes and almost half of ischaemic heart disease cases, and cholesterol concentrations exceeding 3·8 mmol/L for 18% and 55%, respectively. We report estimates of the population health effects and costs of selected interventions to reduce the risks associated with high cholesterol concentrations and blood pressure in areas of the world with differing epidemiological profiles.

Methods: Effect sizes were derived from systematic reviews or meta-analyses, and the effect on health outcomes projected over time for populations with differing age, sex, and epidemiological profiles. Incidence data from estimates of burden of disease were used in a four-state longitudinal population model to calculate disability-adjusted life years (DALYs) averted and patients treated. Costs were taken from previous publications, or estimated by local experts, in 14 regions.

Findings: Non-personal health interventions, including government action to stimulate a reduction in the salt content of processed foods, are cost-effective ways to limit cardiovascular disease and could avert over 21 million DALYs per year worldwide. Combination treatment for people whose risk of a cardiovascular event over the next 10 years is above 35% is also cost effective leading to substantial additional health benefits by averting an additional 65 million DALYs per year worldwide.

Interpretation: The combination of personal and non-personal health interventions evaluated here could lower the global incidence of cardiovascular events by as much as 50%.

INTRODUCTION

Cardiovascular disease is a major contributor to the global burden of disease. It accounts for 20·3% of disability adjusted life years (DALYs) lost in more developed countries and already for 8·1% of those lost in less developed countries. The *World Health Report 2002 (1)* quantified the major contributions of tobacco use, alcohol consumption, high blood pressure, high cholesterol concentrations, low intake of fruit and vegetables, physical inactivity, and high body-mass index to the global burden of disease and of cardiovascular disease in particular *(1,2)*.

Improved data on degree of exposure and reassessments of the magnitude of hazards, have led to the recognition that high blood pressure and high cholesterol concentrations have much greater influence on population health than previously thought *(3)*. About two-thirds of strokes and almost half of cases of ischaemic heart disease can be attributed to systolic blood pressure greater than 115 mm Hg. Total cholesterol concentrations over 3·8 mmol/L account for about 18% of strokes and 55% of cases of ischaemic heart disease. The joint effects of blood pressure and cholesterol concentration would, of course, be less than additive because of the multicausality of cardiovascular disease and the joint action of these two risk factors. Regional analyses have also shown that high blood pressure and high cholesterol concentrations are major risks to health in all regions of the world, not just high-income countries.

Given the burden of disease caused by these factors, assessment of the costs and effects of the available intervention strategies to reduce these risks is important. These strategies should, however, be seen in the context of more comprehensive approaches to the control of cardiovascular disease that focus on several inter-related risks to health including blood pressure, cholesterol concentration, tobacco use, body-

mass index, physical activity, diet, and diabetes *(4,5)*. Here we take advantage of the development of standard methods and companion tools for the assessment of costs, effects, and cost effectiveness of different interventions within and across regions *(6–13)*. These methods and tools mean that results of intervention analyses can be compared more meaningfully across interventions and across locations.

Assessment of the costs and effects of the major intervention strategies for reducing the burden attributable to blood pressure and cholesterol concentrations must address two key issues. First, what are the relative roles of non-personal health services—such as mass-media messages to change diet, or legislation to lower the salt content of processed foods— and personal health services—such as the pharmacological management of cholesterol concentration and hypertension? *(4,5,14)* Second, should management of blood pressure and cholesterol concentrations be based on thresholds for each risk factor seen in isolation (such as treatment for a systolic blood pressure above 160 mm Hg), or should management be based on the absolute risk of cardiovascular disease for a given individual taking into account all his or her known determinants of risk? *(15)* We analyse the population health effects and costs of non-personal health measures, treatment of individual risk factors, and treatment based on various values of absolute risk *(16–18)*.

METHODS

INTERVENTIONS

17 non-personal and personal health-service interventions or combinations have been included in this analysis (table 1). Non-personal health interventions included health education through the mass media (focusing on blood pressure, cholesterol concentration, and body mass), and either legislation or voluntary agreements on salt content to ensure appropriate labelling and stepwise decreases in the salt content of processed foods. Personal health-service interventions included detection and treatment of people with high concentrations of cholesterol for two thresholds; treatment of individuals with high systolic blood pressure with two thresholds; treatment of individuals for both these risk factors; and treatment of individuals based on their absolute risk of a cardiovascular event in the next 10 years (the absolute risk approach *(20)*) with four thresholds. Risk values are defined by fitting mean risk-factor values to observed baseline-risk values. Estimates of the relative risk of modelled risk factors on cardiovascular events are used to predict the absolute risk of individuals with high values for risk factors. Individuals with an absolute risk of cardiovascular disease greater than the threshold all receive a β-blocker, diuretic, statin, and aspirin.

Table 1   Interventions evaluated

Intervention	Description
Non-Personal Interventions:	
N1 Salt reduction through voluntary agreements with industry	Cooperation between government and the food industry for stepwise decrease in salt content of processed foods and for labelling
N2 Population-wide reduction in salt intake legislation	Legislation to decrease salt content in processed foods and appropriate labelling
N3 Health education through mass media	Health education through broadcast and print media focusing on body-mass index and cholesterol concentrations
N4 Combined intervention of N2 and N3	Combination of N2 and N3
Personal Interventions:	
P1 and P2 Individual-based hypertension treatment and education	Treatment of SBP above 160mmHg (P1) or above 140mmHg (P2) with a standard regimen of β-blocker and diuretic
P3 and P4 Individual treatment for high cholesterol concentrations and education	Treatment with statins for total cholesterol concentrations above 6.2 mmol/L (P3) and above 5.7 mmol/L (P4)
P5 Individual treatment and health education for SBP and cholesterol concentration	Combination of P2 and P3, with treatment thresholds of 140mmHg SBP and 6.2 mmol/L for total cholesterol concentration
P6 to P9 Absolute risk approach	People with an estimated combined risk of a cardiovascular event* over the next decade above a given threshold treated for multiple risk factors (with statin, diuretic, β-blocker, and aspirin) whatever the values for individual risk actors; four different thresholds were evaluated: 35% (P6), 25% (P7), 15% (P8), and 5% (P9)

Intervention	Description
Combined Personal and Non-Personal Interventions: (C1 to C4) Addition of the absolute risk approach at the four thresholds to the combined non-personal health intervention	Combination of N4 with P6 to P9

SBP = systolic blood pressure.
* Acute myocardial infarction; angina pectoris; congestive heart failure; first-ever fatal stroke; long-term stroke survivors. The definition of a cardiovascular event differs across studies, so the results reported here may not be strictly comparable with those of similar studies.*(19)*

REGIONS

A desirable approach would be to evaluate all possible combinations of interventions in every country of the world, and for some of the larger countries, at a subnational level. No country has yet been able to do this, and many countries do not have the technical capacity to evaluate even a few of them. At the other extreme, global estimates are of little use to any individual country. WHO, through its CHOICE project, provides information on costs and health effects at a subregional level, with the different parts of the world divided by geographical proximity and epidemiology. This approach allows interventions to be put into broad categories, such as very cost effective, cost effective, and cost ineffective, revealing the extent to which strategies to reduce risks to health should differ across different settings.

In addition, CHOICE provides results in such a way that analysts from countries within a region can adapt them to their own setting if they wish. The costs, effects, and cost-effectiveness of each of the 17 interventions have been evaluated for 14 epidemiological subregions of the world (Annex 1). The results from three of them are discussed in detail—SearD (in southeast Asia with high rates of adult and child mortality); AmrB (in Latin America with low adult and child mortality); and EurA (in Europe with very low adult and child mortality)—although the costs and effectiveness estimates for all 14 subregions are given in Annex 2.

Table 2   Annual Costs, Effects, and Cost-Effectiveness of Interventions

Intervention*	AmrB			EurA			SearD		
	Costs ($x10^6)	DALYs (x10^5)	Cost/DALY ($)	Costs ($x10^6)	DALYs (x10^5)	Cost/DALY ($)	Costs ($x10^6)	DALYs (x10^5)	Cost/DALY ($)
Non-personal									
N1	82	3	244	297	7	438	199	5	373
N2	82	6	127	297	13	227	199	10	193
N3	81	6	136	202	12	165	204	14	143
N4	163	12	135	499	24	205	403	24	168
Personal									
P1	3 122	38	811	14 777	73	2 033	1 570	44	357
P2	8 806	47	1 863	37 385	82	4 536	6 399	71	904
P3	2 426	28	865	19 187	65	2 967	2 537	54	468
P4	4 397	33	1 326	27 142	69	3 933	4 835	69	705
P5	11 232	61	1 832	56 572	110	5 161	8 936	107	838
P6	1 335	51	259	11 555	87	1 324	1 650	74	224
P7	2 059	56	366	16 015	95	1 685	2 888	88	327
P8	3 352	62	542	22 226	104	2 147	5 078	106	477
P9	6 456	69	931	35 750	114	3 144	10 183	133	768
Combined									
C1	1 365	54	252	11 045	91	1 209	1 829	82	222
C2	2 056	58	352	15 474	99	1 570	2 994	95	314
C3	3 332	64	523	21 612	106	2 031	5 149	112	459
C4	6 394	71	903	35 095	116	3 023	10 173	137	745

* See Table 1 for descriptions of interventions.

Effect sizes used in the analysis are based on systematic reviews of randomised trials where possible, or meta analyses (Annexes 3, 4, 5, 6). The joint effects of interventions were assumed to be multiplicative, drawing from the evidence of large cohort studies in diverse populations (21–23).

Population health effects due to the interventions are modelled by stochastically simulating populations specific for age, sex, and subregion with the observed baseline values of cardiovascular risk and the observed distribution of risk factors (systolic blood pressure, cholesterol concentration, body-mass index, and prevalence of long-term smokers) in those regions (24,25). Interventions cover either the entire population (non-personal interventions) or subpopulations meeting specific characteristics (personal interventions), such as having a 10-year risk of a cardiovascular event, or systolic blood pressure, over a certain threshold. Population-level cardiovascular risk (incidence) is recalculated after applying the effectiveness of the intervention and the effect size of the implied change in risk-factor values to the population receiving the intervention.

To translate changes in the risk of cardiovascular disease events specific for age and sex into changes in population health quantified by DALYs, we used a standard multi-state modelling tool, PopMod (see Part Two background paper 2). In this model, health effects are estimated by tracing what would happen to each age and sex cohort of a given population over 100 years, with and without each intervention. PopMod is a four-state population model simulating the evolution of a population partitioned into four distinct health states: people who have the disorder under study, people with some other disorder, people who have both conditions, or those with none of the above (but are susceptible). Births and deaths are also included. The states can be considered either straightforwardly or as aggregates of other substates. Transition rates, such as incidence, remission, and mortality, govern movements between states.

The model is described by a system of ordinary differential equations with each population age and sex group modelled individually as a separate differential equation system (26). The model shows the time evolution of the size of the population age and sex groups through the four health states, and deaths, in yearly steps. With the appropriate health-state valuations, standard life-table measures and various summary measures of population health can be derived. Monte Carlo analysis of uncertainty in transition rates is possible. The side-effect relating to the consequences of bleeding associated with the use of aspirin was included. The entire population is subject to background mortality and morbidity, which are assumed to be independent of the cardiovascular-disease states explicitly modelled.

In some cases, mostly in more developed countries, information on intervention effects was available for only one or two settings. The

associations of blood pressure and cholesterol, however, are remarkably similar in size and shape across Asia, Europe, and North America *(22,27,28)*. We could not obtain evidence about how adherence might vary across settings, so no variation was included. Nevertheless, policy-makers must still make decisions about how to use their scarce resources. One approach would be simply to say that there is no evidence. The approach taken here is to provide the best available evidence, even if this is obtained by extrapolation from one setting to another. This approach carries additional uncertainty, especially in the case of behavioural interventions, which should be taken into account in interpretation of the results.

Costs

Costs include programme-level costs associated with running the intervention (such as administration, training, and media) and patient-level costs (such as primary-care visits, diagnostic tests, and medicines). For this analysis, potential cost-savings related to the prevention of cardiovascular-disease events were not incorporated because the major interest is in identifying the costs of improving population health by preventing these events. Costs were based on a standard ingredients-approach that has been developed by WHO to facilitate costing of interventions *(6–9)*. The units of physical inputs required were assessed and multiplied by the unit price for each input. For programme costs the quantities of the required inputs (such as labour, vehicles, office space) were identified from publications, with additional details provided by programme staff in various parts of the world. The quantity of patient-level resource inputs required for a given health intervention (e.g., hospital inpatient days, outpatient visits, medications, laboratory tests) were identified in a similar way. Reporting of costs by use of the ingredients-approach is an important part of making the results transparent to policy-makers as well as providing a way for analysts to adapt the results to their own settings.

Unit costs of programme-level and patient-level resource inputs, such as the salaries of central administrators, the capital costs of vehicles, offices, and furniture, or the cost per outpatient visit, were obtained from a review of relevant publications and supplemented by primary data from programme staff in several countries. Costs of drugs were based on the price of off-patent drugs from the vendor selling high-quality drugs at the lowest prices.

Information on the costs and effectiveness of interventions that are undertaken inefficiently has little value for decision-makers. For that reason, we assumed capacity utilisation of 80% in most settings—e.g., that health personnel are fully occupied for 80% of their time. The results identify, therefore, the set of interventions that, if done efficiently, would be cost effective in the different settings.

Costs are reported in international dollars to facilitate more meaningful comparisons across regions. An international dollar has the

same purchasing power as the US dollar has in the USA. Costs in local currency units are converted to international dollars by use of purchasing power parity (PPP) exchange rates rather than official exchange rates. A PPP exchange rate is the number of units of a country's currency required to buy the same amounts of goods and services in the domestic market as a US dollar would buy in the USA. An international dollar is, therefore, a hypothetical currency that is used as a means of translating and comparing costs taking into account differences in purchasing power. The base year is 2000. Details of the assumptions are given in Annex 3.

COST-EFFECTIVENESS

Average cost-effectiveness ratios were calculated for each intervention by combining the information on the total costs with information on the total health effects in terms of DALYs averted. All costs and effects are discounted at 3%, consistent with the Disease Control Priority Review *(29),* the first large-scale attempt to compare the cost effectiveness of interventions across diseases, and the recommendations of the US Panel on Cost-Effectiveness in Health and Medicine *(30).* By use of a standard approach, we identified the set of interventions a region should purchase to achieve the greatest health gain for different budget levels. The order in which interventions would be purchased is called an expansion path and is based on the incremental costs and benefits of each intervention compared with the last intervention purchased.

The Commission on Macroeconomics and Health recently defined interventions that have a cost effectiveness ratio of less than three times gross domestic product per head as cost effective *(31).* On this basis, we defined three broad categories: interventions that gain each year of healthy life (i.e., DALY averted) at a cost less than gross domestic product per head are defined as very cost effective; those averting each DALY at a cost between one and three times gross domestic product per head are cost effective; and the remainder are not cost effective.

The results of cost-effectiveness analysis should not be used in a formulaic way—starting with the intervention that has the lowest cost-effectiveness ratio, choosing the next most attractive intervention, and continuing until all resources have been used *(10).* There is generally too much uncertainty surrounding estimates for this approach; moreover, there are other goals of health policy in addition to improving population health. The tool is most powerful when it is used to classify interventions into broad categories such as those we used. This approach provides decision-makers with information on which interventions are low-cost ways of improving population health and which improve health at a much higher cost. This information enters the policy debate to be weighed against the effect of the interventions on other goals of health policy.

Multivariate sensitivity analysis was undertaken to assess the effect of uncertainty in the assumptions on the baseline levels of risks and effect sizes on the cost effectiveness ratios. The first step was to take several samples of hypothetical individuals from correlated distributions of four risk factors: total cholesterol concentration, systolic blood pressure, smoking, and body-mass index. Then samples were taken from distributions around the population means and SDs of the risk factors, as well as around the relative risks and effectiveness estimates from limits developed from the review of relevant publications (Annexes 3 and 4), producing upper and lower confidence limits on mean incidence. This procedure also includes the effects on costs because different numbers of people will be covered by an intervention under the different scenarios. At the same time, the price of medicines—the most important driver of costs—was allowed to vary from half to double the base estimate.

ROLE OF THE FUNDING SOURCE

The sponsor of the study had no role in the study design, collection, analysis, or interpretation of data, or the writing of the report.

RESULTS

Table 2 gives the total annualised costs, total annual health effect in terms of DALYs averted, and the average cost-effectiveness ratio for each of the 17 interventions in three subregions with differing levels of adult and child mortality and different patterns of risks to health (EurA, AmrB, and SearD; the full set of results for all 14 subregions is given in Annex 2). The health benefits of all interventions follow a roughly bell-shaped curve when plotted against age. Depending on the intervention and the region, the curve reaches its maximum at around 60 years of age, with about half

Figure I   Total intervention benefit by age (AmrB)

Figure 1: **Total intervention benefit by age (AmrB)**
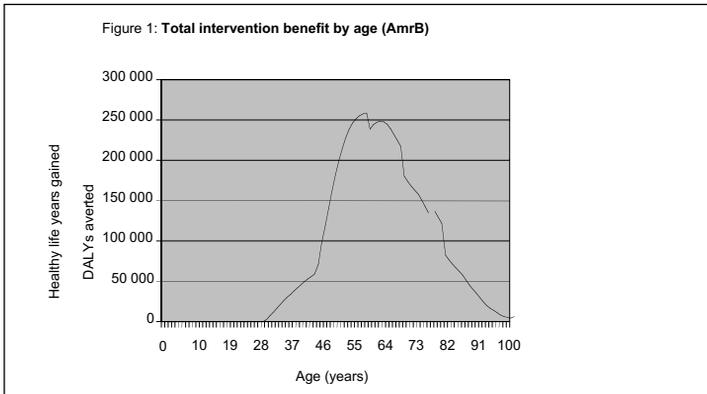
of the total intervention benefit occurring at younger ages, and about half at older ages. This relation is shown for AmrB in figure 1.

All 17 interventions in all three regions are cost effective according to the Commission on Macroeconomics and Health criterion. In all regions, the four non-personal interventions have cost-effectiveness ratios that are lower than personal health-service interventions.

When considered individually, non-personal health interventions to reduce blood pressure and cholesterol are very cost effective. Measures to decrease salt intake are potentially very cost effective, with legislation being more cost effective than voluntary agreements under the assumption that it would lead to the larger reduction in dietary salt intake. The effect of non-personal health service strategies to lower cholesterol concentrations depends on the distribution of risk factors in the region; it has a slightly lower effect on population health than legislation to lower salt intake in EurA and AmrB, but a substantially higher effect in SearD.

Perhaps surprisingly, personal health-service strategies have a much greater potential to reduce the burden of disease–even though they are slightly less cost effective than the population-wide strategies. Treatment of systolic blood pressure above 160 mm Hg is very cost effective in all regions. Statins are now available off-patent at very low cost, and their use for people with total cholesterol concentrations above 6·2 mmol/L is also very cost effective in all regions. However, a comparison of their cost-effectiveness ratios with those of the absolute-risk approach shows that treatment based on measured blood pressure or cholesterol concentrations alone would not be the preferred option on grounds of cost-effectiveness. The absolute-risk approach at a theshold of 35% is always more cost effective than treatment based on either the measured systolic blood pressure or the measured cholesterol concentration. It would avert an additional 65 million DALYs on top of the 21 million DALYs averted by the two non-personal interventions evaluated here.

As the absolute risk threshold is lowered, the health benefits increase, but so do the costs; to obtain each additional unit of health benefit becomes more and more expensive. The exact point at which policy-makers might choose to set the threshold will vary by setting and will take into account many factors in addition to cost effectiveness, but reduction of the threshold even below 15% in the three regions under consideration is very cost effective, even when consequences of bleeding associated with the additional use of aspirin are taken into account.

The cost-effectiveness ratios of the individual interventions do not tell the whole story. Figures 2, 3, and 4 plot the annual cost and DALYs averted for each of the 17 interventions in the three regions. The slope of the line connecting the origin to each point is the cost-effectiveness ratio. The steeper the slope the more expensive the intervention is per DALY averted. This figure also helps to show the incremental cost and incremental health gain of moving from one intervention strategy to another.

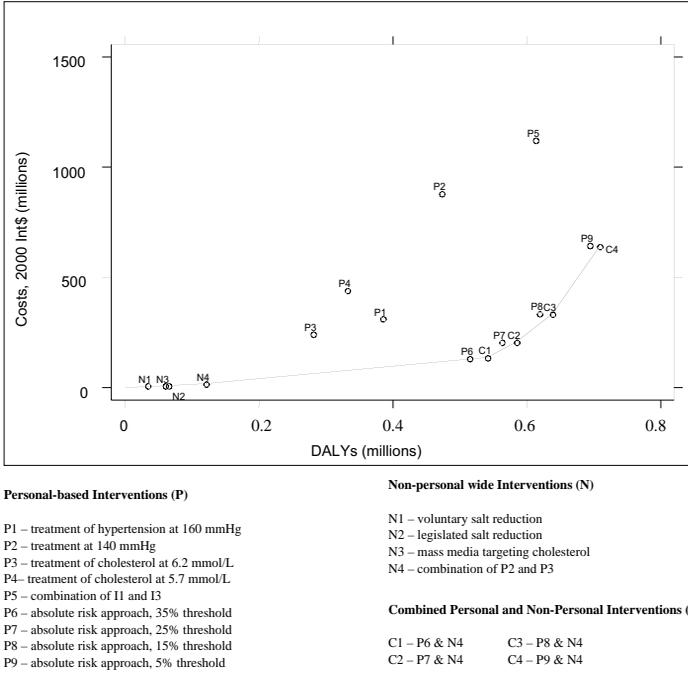**Figure 2**   Yearly costs and effectiveness of AmrB

Figure 3   Yearly costs and effectiveness of EurA
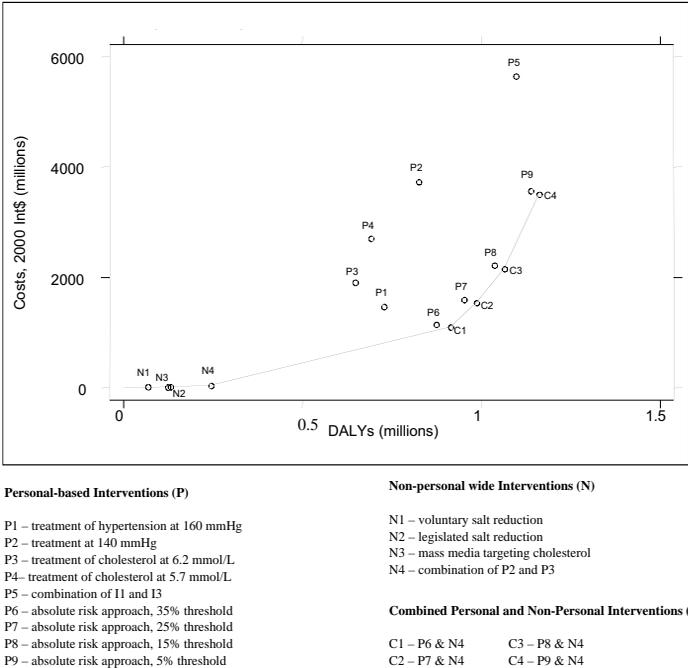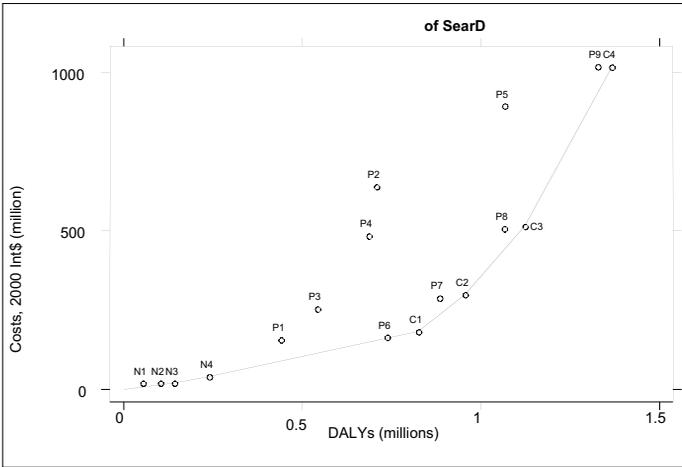
Figure 4   Yearly costs and effectiveness of SearD

Personal-based Interventions (P)

P1 – treatment of hypertension at 160 mmHg
P2 – treatment at 140 mmHg
P3 – treatment of cholesterol at 6.2 mmol/L
P4– treatment of cholesterol at 5.7 mmol/L
P5 – combination of I1 and I3
P6 – absolute risk approach, 35% threshold
P7 – absolute risk approach, 25% threshold
P8 – absolute risk approach, 15% threshold
P9 – absolute risk approach, 5% threshold

Non-personal wide Interventions (N)

N1 – voluntary salt reduction
N2 – legislated salt reduction
N3 – mass media targeting cholesterol
N4 – combination of P2 and P3

Combined Personal and Non-Personal Interventions (C)

C1 – P6 & N4   C3 – P8 & N4
C2 – P7 & N4   C4 – P9 & N4

From the perspective of how best to achieve the best population health for the available resources, the optimum overall strategy is a combination of the population-wide and individual-based interventions. The solid lines joining the most cost-effective points in figures 2, 3, and 4 show the best choice in terms of cost-effectiveness. These expansion paths join the interventions that would be selected for increasing availability of resources. The slopes between them represent the incremental cost-effectiveness ratio (the additional costs required to avert each additional DALY by moving from the lower-cost to the higher-cost intervention). The incremental costs, effects, and cost effectiveness ratios of points on the expansion path are reported in table 3. If resources are extremely scarce, the non-personal interventions will be chosen first.

Table 3   Annual incremental costs, effects, and cost-effectiveness of interventions

Intervention*	AmrB			EurA			SearD		
	Costs ($x10^6)	DALYs (x10^5)	Cost/DALY ($)	Costs ($x10^6)	DALYs (x10^5)	Cost/DALY ($)	Costs ($x10^6)	DALYs (x10^5)	Cost/DALY ($)
Non-personal									
N2	82	7	127	-	-	-	-	-	-
N3	-	-	-	202	12	165	204	14	143
N2 to N4	81	6	145	-	-	-	-	-	-
N3 to N4	-	-	-	297	12	246	199	10	205
Combined									
N4 to C1	1 201	42	286	10 546	67	1 573	1 426	58	244
C1 to C2	692	4	1 598	4 428	7	6 164	1 166	13	898
C2 to C3	1 276	5	2 392	6 138	8	7 812	2 155	17	1 289
C3 to C4	7 899	7	432	13 484	3	13 923	5 023	6	2 058

* See Table 1 for descriptions of interventions.

In all three regions, the expansion path is similar. In settings of extreme resource constraints, one of the non-personal interventions to lower salt intake, cholesterol concentration, or both would be purchased first. Decision-makers who wanted to achieve the greatest health gain for available resources would next move to a combined strategy of legislated reductions in salt content of processed foods with mass-media campaigns, and then add the absolute-risk approach to management of blood pressure and cholesterol concentration. Depending on the resources available, the absolute-risk threshold for a cardiovascular-disease event that would trigger intervention with β-blockers, diuretics, statins, and aspirin would be lowered. Although the total costs, total effects, and cost-effectiveness ratios vary across regions, the sequence of intervention strategies that would be purchased is similar.

Figures 2, 3, and 4 also show that the total effect on the burden of disease through the management of absolute risk is substantial. A population-level reduction in cardiovascular-disease events of more than 50% is possible. Even in the less developed regions of AmrB and SearD, the absolute magnitude of the changes in the burden of cardiovascular disease is impressive.

The multivariate uncertainty analysis shows that the cost-effectiveness ratios can be up to 83% higher on average or 53% lower depending on the region. This variation is illustrated for SearD in figure 5. The "clouds" or uncertainty regions show the range of possible point estimates emerging from the uncertainty analysis for any given intervention. Despite the uncertainty, the individual and combined non-personal interventions are always chosen first—the uncertainty regions for them are close to the origin and do not overlap with any of the treatment options or the combined interventions. However, certainty that population reduction of

Figure 5   Cost-effectiveness uncertainty region for SearD
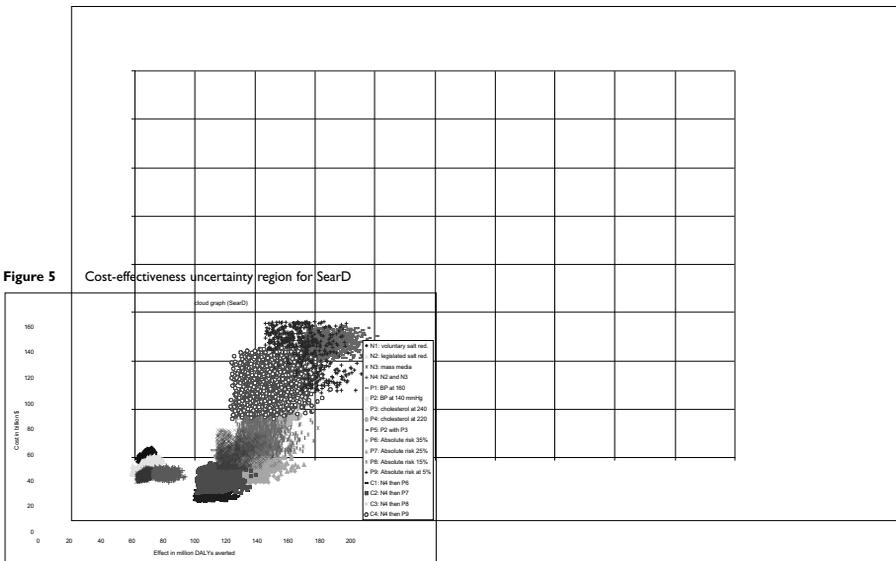
salt intake at 30% will always be more cost effective than the mass-media intervention targeting cholesterol is not possible.

The treatment interventions are highly correlated with each other as are the combination interventions. If a particular sampling for the Monte Carlo simulations shows a low effectiveness for statins, for example, it must be the same for all interventions involving statins (i.e., for all the absolute-risk options and the option to treat individuals only on the basis of measured blood cholesterol concentration). A point at the extreme left of any intervention's uncertainty region would be at the extreme left for the other uncertainty regions involving that treatment.

Therefore, the combined non-personal and personal interventions must also be more costly and more effective than the single personal option in any given situation, even if the uncertainty regions overlap—for example, the combination of the non-personal interventions with treatment at the threshold of 25% must always be more effective and more costly than the option of treatment at the 25% threshold, even though the uncertainty regions overlap. Similarly, the options to treat only blood pressure or only cholesterol concentration must always be inside the expansion path, even though the uncertainty regions overlap slightly.

The essential features of the expansion paths in the three regions do not, therefore, change with the changes in assumptions even if the slopes of the segments change somewhat. The order in which the two non-personal health interventions (reduction of salt content of processed foods and the mass-media approach to lowering cholesterol concentrations) would be introduced might not be certain, but in all cases one would be chosen as the most desirable option, then the second would be added, before the first personal intervention would be considered. In addition, the absolute-risk approach is more cost effective than treatment based on either blood pressure or cholesterol concentration alone in all cases.

DISCUSSION

In all regions, these selected non-personal and personal health interventions to lower blood pressure and cholesterol concentration are very cost effective. This finding is at odds with the perception that strategies to prevent cardiovascular disease should strictly be the concern of the very wealthy. Implied in these results is a further frameshift in thinking about priorities and public-health strategies for less developed regions. Even though the benefits documented here are large, the potential of the non-personal interventions may be even larger. The effect of these interventions observed in North Karelia, Finland, was substantially larger than that in the North American demonstration/cluster trials such as the Stanford five-city study *(32–34),* and the assumptions used here reflect both experiences. With greater understanding of the factors that influence the effectiveness of these non-personal interventions, development of strategies that increase the population health benefits even further could well be possible.

The beneficial effects of reduction in salt intake have been subject to debate, with one review suggesting that the effect on blood pressure was negligible *(35)*. That review focused on interventions involving individual dietary advice to lower salt intake rather than the option to decrease the salt content of processed food as assessed here. Indeed, the review concluded that a population-wide decrease in salt intake through decreasing salt concentrations in processed foods might achieve small reductions in blood pressure across the whole population for sustained periods, which would then have substantial health effects at the population level. This was the motivation for the intervention analysed here. In addition, there is evidence that small and repeated decreases in salt intake are not discernible on grounds of taste (e.g., less salt does not necessarily mean less taste *(36)*), so people are unlikely to resist the new foods for that reason.

The absolute-risk approach to management of blood pressure and cholesterol concentration is very cost effective in all regions, and has the potential to bring about substantial reductions in ischaemic heart disease and stroke. Many other combinations of medicines are likely to be as cost effective as those evaluated here. A meta-analysis of 354 trials involving 56 000 participants, showed the blood pressure reductions produced by the major classes of drug at standard dose are similar, independent, and additive, and that half the standard dose reduces efficacy by only 20% while more than halving side-effects *(27,37,38)*. There are also probable or proven benefits of these interventions on other important outcomes not measured here, such as dementia, renal failure, peripheral vascular disease, congestive heart failure, and the need for coronary-artery bypass grafting *(39–43)*. Furthermore, although the effects of drugs to lower blood pressure and cholesterol concentration are due largely to the reduction in the risk factor achieved *(24,25)*, there may be some additional benefits with specific agents, such as reduction in the risk of coronary disease with inhibitors of angiotensin-converting enzyme *(43)*. Implementation of risk screening can and should be tailored to the resource levels of national health systems. In high-income countries, risk assessment on the basis of age, sex, measured blood pressure, cholesterol concentration, body-mass index, diabetes, tobacco use, and clinical history of previous cardiovascular-disease events is practical. In low-resource settings, however, adequate risk screening could be based simply on age, sex, measured blood pressure, body-mass index, tobacco use, and past cardiovascular-disease events. This assessment would require no sophisticated technology or blood sampling. A "risk pill" of antihypertensive drugs, statin, and aspirin could also be packaged as a single compound, facilitating compliance.

As the absolute-risk threshold used to trigger treatment is lowered, larger proportions of the adult population would be on long-term drug treatment and the number of adverse events would increase. The consequences of this medicalisation of potentially the majority of the

adult population should be carefully considered. Issues of long-term compliance might also limit the applicability of the approach in certain populations, including younger age-groups. The potentially huge benefits and the apparent cost-effectiveness of the absolute-risk approach do seem to justify some large-scale population effectiveness studies. States or provinces in countries facing major cardiovascular-disease challenges could be enrolled to see whether the expected population benefits can be achieved in the short time-frame implied by the analysis.

In more developed countries, values of blood pressure and cholesterol concentration are well known to be worse in the poor than in the rich *(44)*. Knowledge on how to manage these risks is used more effectively by the higher income, more educated population groups. Consequently, the coverage of interventions to decrease blood pressure and cholesterol concentration is probably lower in the poor. Because of the distribution of these risks, there is a potential for both the non-personal and the absolute-risk approaches to contribute substantially to the reduction in adult health inequalities. It is a challenge for public health to develop innovative strategies to encourage the uptake of the latter in the poor and disadvantaged *(45)*. Lateral thinking may be needed. Studies show that intervention uptake can be affected by financial incentives. Perhaps lottery tickets could be given to individuals who reduce their absolute risk by a certain amount in a year.

Why is this analysis apparently suggesting a much bigger effect at lower cost for personal health-service interventions to manage blood pressure and cholesterol concentration than may have been expected? First, as part of the comparative risk analysis module of the Global Burden of Disease 2000 project, a clearer picture of the burden of these risk factors worldwide has emerged *(2)*. Second, new ways of using existing drugs such as the absolute-risk method have been developed. Third, lovastatin is now off-patent and other statins will follow soon, substantially lowering the cost of these regimens. Fourth, developments in the analysis of hazard data to deal with the effect of measurement error and regression dilution bias *(21,46–48)* have led to a near doubling of the estimated effect of reductions in blood pressure and cholesterol concentration on outcomes. These changes remind us why updating and re-evaluation of strategies that address major public-health problems will always be important.

The non-personal interventions considered here were even more cost effective than the personal interventions in the three regions despite having a lower overall effect on population health. Care is needed in interpretation of these results, because the estimates of changes resulting from the mass-media intervention were based on changes in behaviour observed in a more developed setting; however, even with a halving of the assumed effectiveness, the conclusion would not be altered. On the basis of this assumption, non-personal interventions would be the first to be introduced. Moreover, the non-personal interventions assessed here are

only a selection of those possible, and their very nature makes reliable assessment of effects challenging. But this challenge must be accepted. For example, assessment of strategies to achieve moderate but widespread changes in manufactured food (e.g., in overall fat content) would be very worthwhile, since unhealthy cholesterol concentrations and blood pressures have major dietary components to their aetiology.

Three final considerations are pertinent to the policy debate to which this paper contributes. First, the combination of medicines that prevent people at high risk of cardiovascular disease from having an event would cost just less than $14 per person per year if the cheapest medicines were purchased internationally. This is simply the costs of the medicines and does not include distribution mark-ups. The ability of poor countries to finance this intervention, and all the other possible cost-effective interventions, from their own resources is limited; some countries spend less than $10 per head on health each year. The availability of low-cost, effective ways to improve health in all settings, many of which are not affordable at current levels of health expenditure, is why WHO has argued strongly for massive injections of resources for health from richer countries that could be used to reduce the burden of disease among the poor.

Second, this paper has focused on reducing the health consequences associated with high cholesterol concentrations and high blood pressure. It shows which interventions should be given priority in development of a strategy for the control of cardiovascular disease. This is important information for policy-makers responsible for cardiovascular-disease control or health promotion. It does not, however, indicate whether control of cardiovascular disease should receive priority over reducing other risks, such as those associated with unsafe sex. Priority setting requires consideration of the costs and effects of all possible alternatives. WHO seeks to provide this information through its CHOICE project, and initial results covering several major risks to health can be found in the *World Health Report 2002 (1)*.

Third, cost-effectiveness is only one of the key inputs to final decisions about how to allocate scarce resources. Policy-makers also have other concerns, such as reducing poverty and inequalities, and questions of human rights and community acceptance also influence policy. Another key concern is how different types of interventions can be incorporated into the health infrastructure of the country, or how the infrastructure could be adapted to accommodate the desired strategies. The information presented here is one, but only one, of the critical inputs required to inform the decision-making process about efficient ways to reduce risks to health.

CONTRIBUTORS

J. Lauer and R. Hutubessy estimated population effectiveness, resource utilisation, unit prices, and programme costs. C. Murray developed the

method of simulating intervention interactions and their effect on population health. N Tomijima simulated the effect of the interventions on the incidence of cardiovascular disease and the number of people needing treatment under different scenarios and implemented the uncertainty analysis. A Rodgers and C Lawes did the meta-analysis of risk associated with values of blood pressure and cholesterol concentration, and estimated the effect of reduction in dietary salt intake on the incidence of cardiovascular disease. L Niessen and A Rodgers contributed to the development of the cardiovascular-disease model and the absolute-risk approach. D Evans guided the development of the methods and approaches to cost-effectiveness analysis and drafted the report. R Hutubessy estimated the uncertainty intervals. J Lauer developed the population effectiveness model. All authors contributed to writing the report.

The opinions in this paper are those of the authors and not necessarily those of the institutions they represent.

Conflict of interest statement

No author has conflicts of interest to declare.

Acknowledgements

References

1. World Health Organization. *The World Health Report 2002: reducing risks, promoting healthy life.* Geneva: WHO, 2002.

2. Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, Comparative Risk Assessment Collaborating Group. Selected major risk factors and global and regional burden of disease. *Lancet* 2002; 360: 1347–60.

3. Murray CJL, Lopez AD. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020, 1st edn.* Cambridge, MA: Harvard University Press, 1996.

4. Puska P. Development of public policy on the prevention and control of elevated blood cholesterol. *Cardiovasc Risk Factors* 1996; 6: 203–10.

5. World Health Organization. *Innovative care for chronic conditions: building blocks for action.* Geneva: WHO, document No. WHO/MNC/CCH/02.01: 2002.

6. Johns B, Baltussen R, Hutubessy RCW. *Programme costs in the economic evaluation of health interventions* (in press).

7. Adam T, Evans DB, Koopmanschap MA. Cost-effectiveness analysis: can we reduce variability in costing methods? *Int J Technol Assess Health Care* (in press).

8. Baltussen RM, Adam T, Tan Torres T, et al. *Generalized cost-effectiveness analysis: a guide.* Geneva: WHO, Global Programme on Evidence for Health Policy: 2002. Available at: http://www.who.int/evidence/cea

9. Adam T, Evans DB, Murray CJL. *Econometric estimation of country-specific hospital costs* (in press).

10. Murray CJ, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Econ* 2000; 9: 235–51.

11. Hutubessy RC, Baltussen RM, Evans DB, Barendregt JJ, Murray CJ. Stochastic league tables: communicating cost-effectiveness results to decision-makers. *Health Econ* 2001; 10: 473–77.

12. Hutubessy RCW, Baltussen RMPM, Tan Torres-Edejer T, Evans DB. Generalised cost-effectiveness analysis: an aid to decision making in health. *Appl Health Econ Health Policy* 2002; 1: 89–95.

13. Baltussen RM, Hutubessy RC, Evans DB, Murray CJ. Uncertainty in cost-effectiveness analysis. *Probabilistic uncertainty analysis and stochastic league tables. Int J Technol Assess Health Care* 2002; 18: 112–19.

14. Martin I. Implementation of WHO/ISH Guidelines: role and activities of WHO. *Clin Exp Hypertens* 1999; 21: 659–69.

15. Law MR, Wald NJ. Risk factor thresholds: their existence under scrutiny. *BMJ* 2002; 324: 1570–76.

16. Cooper RS, Rotimi CN, Kaufman JS, Muna WF, Mensah GA. Hypertension treatment and control in sub-Saharan Africa: the epidemiological basis for policy. *BMJ* 1998; 316: 614–17.

17. Marshall T, Rouse A. Resource implications and health benefits of primary prevention strategies for cardiovascular disease in people aged 30 to 74: mathematical modelling study. *BMJ* 2002; 325: 197–99.

18. Ebrahim S, Smith GD. Exporting failure? Coronary heart disease and stroke in developing countries. *Int J Epidemiol* 2001; 30: 201–05.

19. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991; 121: 293–98.

20. Jackson R, Barham P, Bills J, et al. Management of raised blood pressure in New Zealand: a discussion document. *BMJ* 1993; 307: 107–10.

21. Neaton JD, Wentworth D. Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease: overall findings and differences by age for 316,099 white men. *Arch Intern Med* 1992; 152: 56–64.

22. Eastern Stroke and Coronary Heart Disease Collaborative Research Group. Blood pressure, cholesterol, and stroke in eastern Asia. *Lancet* 1998; 352: 1801–07.

23. Jee SH, Suh I, Kim IS, Appel LJ. Smoking and atherosclerotic cardiovascular disease in men with low levels of serum cholesterol: the Korea Medical Insurance Corporation Study. *JAMA* 1999; 282: 2149–55.

24. Lawes C, Vander Hoorn S, Rodgers A. Blood pressure and burden of disease. In: Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, eds. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors.* Geneva: WHO (in press).

25. Lawes C, Vander Hoorn S, Law MR, MacMahon S, Rodgers A. Global disease burden attributable to cholesterol. In: Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, eds. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors.* Geneva: WHO (in press).

26. Lambert JD. *Numerical methods for ordinary differential systems: the initial value problem.* Chichester: Wiley, 1991.

27. Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *BMJ* 1994; 308: 367–72.

28. Prospective studies collaboration. Age-specfic relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; 360: 1903–13.

29. Jamison DT, Mosley WH, Measham AR, Bobadilla JL. Disease control priorities in developing countries. New York: Oxford University Press, 1993.

30. Gold MR, Siegel JE, Russel LB, Weinstein MC. *Cost-effectiveness in health and medicine.* New York: Oxford University Press, 1996.

31. WHO Commission on Macroeconomics and Health. *Macroeconomics and health: investing in health for economic development.* Report of the Commission on Macroeconomics and Health. Geneva: WHO, 2001.

32. Fortmann SP, Varady AN. Effects of a community-wide health education program on cardiovascular disease morbidity and mortality: the Stanford Five-City Project. *Am J Epidemiol* 2000; 152: 316–23.

33. Vartiainen E, Puska P, Jousilahti P, Korhonen HJ. Cardiovascular diseases and risk factors in Finland. *Prev Med* 1999; 29: S124–29.

34. Sellers DE, Crawford SL, Bullock K, McKinlay JB. Understanding the variability in the effectiveness of community heart health programs: a meta-analysis. *Soc Sci Med* 1997; 44: 1325–39.

35. Hooper L, Bartlett C, Davey SG, Ebrahim S. Systematic review of long term effects of advice to reduce dietary salt in adults. *BMJ* 2002; 325: 628.

36. Rodgers A, Neal B. Less salt does not necessarily mean less taste. *Lancet* 1999; 353: 1332.

37. Law MR, Wald NJ, Morris JK, Jordon RE. Blood pressure lowering drugs: analysis of 354 randomised trials to assess the value of low dose combination therapy. *BMJ* (in press).

38. Sacks FM, Tonkin AM, Shepherd J, et al. Effect of pravastatin on coronary disease events in subgroups defined by coronary risk factors the prospective pravastatin pooling project. *Circulation* 2000; 102: 1893–1900.

39. Richard F, Fromentin-David I, Ricolfi F, et al. The angiotensin converting enzyme gene as a susceptibility factor for dementia. *Neurology* 2001; 56: 1593–95.

40. Collins R, Armitage J, Parish S, Sleight P, Peto R. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; 360: 7–22.

41. PROGRESS Collaborative Group. Effects of blood pressure lowering with perindopril and indapamide on dementia and severe cognitive decline in patients with cerebrovascular disease. *Arch Intern Med* (in press).

42. Forette F, Seux ML, Staessen JA, et al. Prevention of dementia in randomised double-blind placebo-controlled Systolic Hypertension in Europe (Syst-Eur) trial. *Lancet* 1998; 352: 1347–51.

43. Neal B, MacMahon S, Chapman N. Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. *Lancet* 2000; 356: 1955–64.

44. Marmot M, Shipley M, Brunner E, Hemingway H. Relative contribution of early life and adult socioeconomic factors to adult morbidity in the Whitehall II study. *J Epidemiol Community Health* 2001; 55: 301–07.

45. Haynes RB, McKibbon KA, Kanani R. Systematic review of randomised trials of interventions to assist patients to follow prescriptions for medications. *Lancet* 1996; 348: 383–86.

46. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease: part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; 335: 765–74.

47. Prospective studies collaboration. Cholesterol, diastolic blood pressure, and stroke: 13 000 strokes in 450 000 people in 45 prospective cohorts. *Lancet* 1995; 346: 1647–53.

48. Asia Pacific Cohort Studies Collaboration. Blood pressure and cardiovascular diseases in the Asia Pacific region. *J Hypertens* (in press).

Annex 1   Epidemiological subregions

Region and mortality stratum*	Countries
African Region (Afr)	
D	Algeria, Angola, Benin, Burkina Faso, Cameroon, Cape Verde, Chad, Comoros, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Niger, Nigeria, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Togo
E	Botswana, Burundi, Central African Republic, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Eritrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Uganda, United Republic of Tanzania, Zambia, Zimbabwe
Region of the Americas (Amr)	
A	Canada, United States of America, Cuba
B	Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Dominica, Dominican Republic, El Salvador, Grenada, Guyana, Honduras, Jamaica, Mexico, Panama, Paraguay, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela
D	Bolivia, Ecuador, Guatemala, Haiti, Nicaragua, Peru
Eastern Mediterranean Region (Emr)	
B	Bahrain, Cyprus, Iran (Islamic Republic of), Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates
D	Afghanistan, Djibouti, Egypt, Iraq, Morocco, Pakistan, Somalia, Sudan, Yemen
European Region (Eur)	
A	Andorra, Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Sweden, Switzerland, United Kingdom
B	Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Georgia, Kyrgyzstan, Poland, Romania, Slovakia, Tajikistan, The Former Yugoslav Republic of Macedonia, Turkey, Turkmenistan, Uzbekistan, Yugoslavia
C	Belarus, Estonia, Hungary, Kazakhstan, Latvia, Lithuania, Republic of Moldova, Russian Federation, Ukraine
Southeast Asian Region (Sear)	
B	Indonesia, Sri Lanka, Thailand
D	Bangladesh, Bhutan, Democratic People's Republic of Korea, India, Maldives, Myanmar, Nepal
Western Pacific Region (Wpr)	
A	Australia, Japan, Brunei Darussalam, New Zealand, Singapore
B	Cambodia, China, Lao People's Democratic Republic, Malaysia, Mongolia, Philippines, Republic of Korea, Viet Nam
	Cook Islands, Fiji, Kiribati, Marshall Islands, Micronesia (Federated States of), Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu

* A subregions have very low rates of adult and child mortality; B have low adult and low child mortality; C have high adult and low child mortality; D have high adult and high child mortality; and E have very high adult and high child mortality.

Annex 2   Annual costs, effects and cost-effectiveness of interventions for 14 sub-regions

	AfrD			AfrE		
	Costs *(in $millions)*	**DALYs** *(100,000s)*	**Cost/DALY** *($)*	**Costs** *(in $millions)*	**DALYs** *(100,000s)*	**Cost/DALY** *($)*
Interventions*						
Non-Personal (N)						
N1: voluntary salt red.	85	1	573	118	1	919
N2: legislated salt red.	85	3	297	118	2	472
N3: mass media	124	2	779	124	2	760
N4: N2 and N3	208	4	484	241	4	602
Personal (P)						
P1: BP at 160	1464	18	825	499	16	308
P2: BP at 140 mmHg	3375	20	1678	1971	19	1011
P3: cholesterol at 240	122	3	416	262	3	964
P4: cholesterol at 220	391	5	762	619	5	1292
P5: P2 with P3	3497	21	1659	2234	20	1094
P6: Absolute risk 35%	358	16	225	198	14	138
P7: Absolute risk 25%	594	18	338	352	16	216
P8: Absolute risk 15%	1055	20	535	733	19	388
P9: Absolute risk at 5%	2318	23	1014	1947	23	850
Combined Interv. (C)						
C1: N4 then P6	516	17	306	408	15	264
C2: N4 then P7	733	18	399	543	17	316
C3: N4 then P8	1186	20	581	902	20	460
C4: N4 then P9	2412	23	1035	2085	23	893

* See Table 1 for description of interventions

Annex 2   Cont., Annual costs, effects and cost-effectiveness of interventions for 14 sub-regions

	AmrA			AmrB		
	Costs	DALYs	Cost/DALY	Costs	DALYs	Cost/DALY
*Interventions**	*(in $millions,*	*(100,000s)*	*($)*	*(in $millions)*	*(100,000s)*	*($)*
Non-Personal (N)						
N1: voluntary salt red.	92	4	229	82	3	244
N2: legislated salt red.	92	8	119	82	6	127
N3: mass media	114	11	106	81	6	136
N4: N2 and N3	207	18	115	163	12	135
Personal (P)						
P1: BP at 160	6136	50	1224	3122	38	811
P2: BP at 140 mmHg	19669	60	3294	8806	47	1863
P3: cholesterol at 240	7716	52	1475	2426	28	865
P4: cholesterol at 220	13776	60	2280	4397	33	1326
P5: P2 with P3	27384	90	3059	11232	61	1832
P6: Absolute risk 35%	9614	81	1181	1335	51	259
P7: Absolute risk 25%	13278	88	1512	2059	56	366
P8: Absolute risk 15%	18498	95	1953	3352	62	542
P9: Absolute risk at 5%	29728	103	2882	6456	69	931
Combined Interv. (C)						
C1: N4 then P6	9117	85	1077	1365	54	252
C2: N4 then P7	12783	91	1409	2056	58	352
C3: N4 then P8	17931	97	1845	3332	64	523
C4: N4 then P9	29142	105	2767	6394	71	903

* See Table 1 for description of interventions

Annex 2   Cont., Annual costs, effects and cost-effectiveness of interventions for 14 sub-regions

	AmrD			EmrB		
	Costs	DALYs	Cost/DALY	Costs	DALYs	Cost/DALY
Interventions*	*(in $millions,*	*(100,000s)*	*($)*	*(in $millions)*	*(100,000s)*	*($)*
Non-Personal (N)						
N1: voluntary salt red.	16	0.3	500	66	1	541
N2: legislated salt red.	16	1	263	66	2	273
N3: mass media	18	1	335	44	3	174
N4: N2 and N3	33	1	306	110	5	231
Personal (P)						
P1: BP at 160	417	4	1140	1151	16	712
P2: BP at 140 mmHg	1435	4	3191	3700	20	1894
P3: cholesterol at 240	292	2	1251	688	12	583
P4: cholesterol at 220	529	3	1889	1164	13	896
P5: P2 with P3	1727	6	3078	4388	23	1883
P6: Absolute risk 35%	176	4	457	499	19	264
P7: Absolute risk 25%	296	4	681	744	20	363
P8: Absolute risk 15%	541	5	1082	1140	22	512
P9: Absolute risk at 5%	1221	6	2047	2036	24	833
Combined Interv. (C)						
C1: N4 then P6	184	4	445	555	20	283
C2: N4 then P7	298	5	648	789	21	376
C3: N4 then P8	533	5	1026	1176	23	519
C4: N4 then P9	1201	6	1966	2067	25	836

* See Table 1 for description of interventions

Annex 2   Cont., Annual costs, effects and cost-effectiveness of interventions for 14 sub-regions

	EmrD			EurA		
	Costs	DALYs	Cost/DALY	Costs	DALYs	Cost/DALY
Interventions*	*(in $millions,*	*(100,000s)*	*($)*	*(in $millions)*	*(100,000s)*	*($)*
Non-Personal (N)						
N1: voluntary salt red.	76	3	246	297	7	438
N2: legislated salt red.	76	6	125	297	13	227
N3: mass media	91	6	164	202	12	165
N4: N2 and N3	167	11	149	499	24	205
Personal (P)						
P1: BP at 160	1249	38	332	14777	73	2033
P2: BP at 140 mmHg	3790	42	908	37385	82	4536
P3: cholesterol at 240	897	27	336	19187	65	2967
P4: cholesterol at 220	1516	30	511	27142	69	3933
P5: P2 with P3	4687	53	882	56572	110	5161
P6: Absolute risk 35%	598	48	126	11555	87	1324
P7: Absolute risk 25%	853	50	170	16015	95	1685
P8: Absolute risk 15%	1285	54	240	22226	104	2147
P9: Absolute risk at 5%	2294	58	399	35750	114	3144
Combined Interv. (C)						
C1: N4 then P6	706	49	144	11045	91	1209
C2: N4 then P7	952	52	184	15474	99	1570
C3: N4 then P8	1381	55	252	21612	106	2031
C4: N4 then P9	2381	59	407	35095	116	3023

* See Table 1 for description of interventions

Annex 2 Cont., Annual costs, effects and cost-effectiveness of interventions for 14 sub-regions

	EurB			EurC		
	Costs	DALYs	Cost/DALY	Costs	DALYs	Cost/DALY
	(in $millions,	*(100,000s)*	*($)*	*(in $millions)*	*(100,000s)*	*($)*
Interventions*						
Non-Personal (N)						
N1: voluntary salt red.	95	8	122	26	14	19
N2: legislated salt red.	95	15	63	26	28	10
N3: mass media	60	7	89	31	13	23
N4: N2 and N3	155	21	73	57	40	14
Personal (P)						
P1: BP at 160	3438	74	465	3910	141	276
P2: BP at 140 mmHg	6624	78	846	7233	149	487
P3: cholesterol at 240	928	28	330	3004	74	407
P4: cholesterol at 220	1841	35	524	4455	82	542
P5: P2 with P3	7552	89	852	10238	175	585
P6: Absolute risk 35%	2034	82	248	3457	166	208
P7: Absolute risk 25%	2810	86	326	4409	172	256
P8: Absolute risk 15%	3948	91	435	5759	178	324
P9: Absolute risk at 5%	6433	96	671	8482	184	461
Combined Interv. (C)						
C1: N4 then P6	1984	85	235	3249	170	191
C2: N4 then P7	2735	88	309	4198	176	239
C3: N4 then P8	3883	93	419	5557	181	307
C4: N4 then P9	6380	98	654	8321	187	444

* See Table 1 for description of interventions

Annex 2   Cont., Annual costs, effects and cost-effectiveness of interventions for 14 sub-regions

	SearB			SearD		
	Costs	DALYs	Cost/DALY	Costs	DALYs	Cost/DALY
Interventions*	*(in $millions,*	*(100,000s)*	*($)*	*(in $millions)*	*(100,000s)*	*($)*
Non-Personal (N)						
N1: voluntary salt red.	21	2	136	199	5	373
N2: legislated salt red.	21	3	70	199	10	193
N3: mass media	30	3	120	204	14	143
N4: N2 and N3	52	5	95	403	24	168
Personal (P)						
P1: BP at 160	1130	15	769	1570	44	357
P2: BP at 140 mmHg	3652	20	1834	6399	71	904
P3: cholesterol at 240	493	8	654	2537	54	468
P4: cholesterol at 220	1072	10	1066	4835	69	705
P5: P2 with P3	4145	24	1709	8936	107	838
P6: Absolute risk 35%	443	16	279	1650	74	224
P7: Absolute risk 25%	779	19	416	2888	88	327
P8: Absolute risk 15%	1541	23	684	5078	106	477
P9: Absolute risk at 5%	3708	28	1309	10183	133	768
Combined Interv. (C)						
C1: N4 then P6	422	18	238	1829	82	222
C2: N4 then P7	733	20	360	2994	95	314
C3: N4 then P8	1458	24	613	5149	112	459
C4: N4 then P9	3585	29	1229	10173	137	745

* See Table 1 for description of interventions

Annex 2   Cont., Annual costs, effects and cost-effectiveness of interventions for 14 sub-regions

	WprA			WprB		
	Costs	DALYs	Cost/DALY	Costs	DALYs	Cost/DALY
Interventions*	_(in $millions)_	_(100,000s)_	_($)_	_(in $millions)_	_(100,000s)_	_($)_
Non-Personal (N)						
N1: voluntary salt red.	183	3	536	219	12	189
N2: legislated salt red.	183	6	291	219	23	97
N3: mass media	91	3	316	1148	11	1030
N4: N2 and N3	274	9	307	1368	33	413
Personal (P)						
P1: BP at 160	4789	34	1395	6044	127	476
P2: BP at 140 mmHg	12991	38	3420	17884	154	1164
P3: cholesterol at 240	2407	11	2265	3165	36	886
P4: cholesterol at 220	5341	15	3566	6690	48	1406
P5: P2 with P3	15398	43	3602	21049	172	1223
P6: Absolute risk 35%	3972	36	1107	3166	135	235
P7: Absolute risk 25%	5569	40	1397	5179	151	344
P8: Absolute risk 15%	7978	44	1803	9337	171	545
P9: Absolute risk at 5%	13755	49	2788	19968	199	1005
Combined Interv. (C)						
C1: N4 then P6	3908	38	1034	4175	144	291
C2: N4 then P7	5476	42	1319	6072	158	383
C3: N4 then P8	7843	46	1716	10114	178	569
C4: N4 then P9	13659	51	2696	20672	204	1014

* See Table 1 for description of interventions

Annex 3   Assumptions on effectiveness and costs for non-personal interventions

Model variables	Assumptions	Sources
Basic epidemiology		
Ischaemic heart disease	Acute myocardial infarction; angina pectoris; congestive heart failure	Murray, 1996[1]
Cerebrovascular disease	First-ever fatal stroke cases; long-term stroke survivors	Murray, 1996[1]
Risk factor and epidemiology (see Annex 4)		
Blood pressure	SBPs specific for region, age, and sex; age-specific relative risks of CVD event for 1mm HG change in SBP	WHO, 2002[2]
Cholesterol concentration	Cholesterol concentrations specific for region, age, and sex; age-specific relative risks of CVD event for 1mmol/L change in total blood cholesterol	WHO, 2002[2]
Body-mass index	Body-mass index specific for region, age, and sex; age-specific relative risks of CVD event for 1 unit change in body-mass index	WHO, 2002[2]
Smoking	Prevalence of long-term smokers specific for region, age, and sex; age-specific relative risks of CVD event for unit change in prevalence of long-term smokers	WHO, 2002[2]
Programme level costs		
Intervention N1	Central administration and planning costs at 95% coverage	Johns, 2002;[3] Adam, 2002[4]
Intervention N2	Central administration, planning and enforcement costs at 95% coverage	Johns, 2002;[3] Adam, 2002[4]
Intervention N3	Central administration costs, planning, media costs and printed materials at 80% coverage	Johns, 2002;[3] Adam, 2002[4]
Effectiveness (see Annex 5)		
Intervention N1	Blood-pressure changes specific for region, age, and sex associated with a 15% reduction in total dietary salt intake	Law, 1991[5]
Intervention N2	Blood-pressure changes specific for region, age, and sex associated with a 30% reduction in total dietary salt intake	Frost, 1991[6]
Intervention N3	2% reduction in total blood cholesterol levels	Law, 1991[7]
Intervention N4	Combined effect of interventions N2 and N3	Lawes, 2002,[8] Tosteson, 1997[9]

SBP=systolic blood pressure; CVD=cardiovascular disease.

References

1. Murray CJL, Lopez AD. The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. 1ed. Cambridge, MA: Harvard University Press; 1996.

2. World Health Organization. The World Health Report 2002: Reducing risks, promoting healthy life. Geneva: World Health Organization; 2002.

3. Johns B, Baltussen R, Hutubessy RCW. Programme costs in the economic evaluation of health interventions. *Cost-effectiveness and Resource Allocation* 2003; **1**.

4. Adam T, Evans DB, Murray CJL. Econometric estimation of country-specific hospital costs. *Cost-effectiveness and Resource Allocation* 2003; **1**.

5. Law MR, Frost CD, Wald NJ. By how much does dietary salt reduction lower blood pressure? I-- Analysis of observational data among populations. *BMJ* 1991; **302**: 811-5.

6. Frost CD, Law MR, Wald NJ. By how much does dietary salt reduction lower blood pressure? II-- Analysis of observational data within populations. *BMJ* 1991; **302**: 815-8.

7. Law MR, Frost CD, Wald NJ. By how much does dietary salt reduction lower blood pressure? III-- Analysis of data from trials of salt reduction. *BMJ* 1991; **302**: 819-24.

8. Lawes C, Feigin V, Rodgers A. Estimating reductions in blood pressure following reductions in salt intake by age, sex and WHO region. Auckland: Clinical Trials Research Unit, University of Auckland; 2002.

9. Tosteson AN, Weinstein MC, Hunink MG, Mittleman MA, Williams LW, Goldman PA, Goldman L. Cost-effectiveness of populationwide educational approaches to reduce serum cholesterol levels. *Circulation* 1997; **95**: 24-30.

Annex 4   Assumptions for personal and combined interventions

Model variables	Assumption	Sources
Basic epidemiology		
Ischaemic heart disease	Acute myocardial infarction; angina pectoris; congestive heart failure	Murray, Lopez, 1996[1]
Cerebrovascular disease	First-ever fatal stroke cases; long-term stroke survivors	Murray, Lopez, 1996[1]
Upper gastrointestinal bleeding	Adverse effect of antiplatelet therapy with low-dose aspirin	Hernandez-Diaz, Rodriguez, 2002[2]
Risk factor and epidemiology		
Blood pressure	SBPs specific for region, age, and sex; global age-specific relative risks of CVD event for 1mm Hg change in SBP	WHO, 2002[3]
Cholesterol	Cholesterol concentrations specific for region, age, and sex; global age-specific in total blood cholesterol relative risks of CVD event for 1mmol/L change	WHO, 2002[3]
Body-mass index	Body-mass index specific for region, age, and sex; global age-specific relative risks of CVD event for 1 unit change in body mass index	WHO, 2002[3]
Smoking	Prevalence of long-term smokers specific for region, age, and sex; global age-specific relative risks of CVD event for unit change in prevalence of long-term smokers	WHO, 2002[3]
Intervention coverage	Coverage of antihypertensive drug treatment among respondents aware of high blood pressure Coverage of cholesterol-lowering drug treatment among respondents aware of high cholesterol	Molarius, 1998;[4] Tolonen, 1999;[5] Tolonen, 1999[6]
Patient level costs		
Intervention P1 and P2	*Drug treatment* - 50mg/day atenolol; 25mg/day hydrochlorothiazide *Provider visits* - 4 health-care provider visits/year; 1.5 outpatient visits/year for health education *Laboratory tests* - annual renal function, lipid profile, and blood sugar tests	Unit prices of health facilities based on Adam et al.[7] Unit price drugs are based on the International Drug Price indicator by Management Science for Health (www.erc.msh.org) Unit prices laboratory tests are based on Arizona Health Care
Intervention P3 and P4	*Drug treatment* - 30mg/day lovastatin *Provider visits* - 4 health-care provider visits/year; 1.5 outpatient visits/year for health education *Laboratory tests* - total cholesterol and hepatic function[1]	
Intervention P5	*Drug treatment* - 50mg/day atenolol; 25mg/day hydrochlorothiazide; and 30mg/day lovastatin *Provider visits* - 4 health-care provider visits/year; 1.5 outpatient visits/year for health education *Laboratory tests* - annual renal function, lipid profile, and blood sugar tests; total cholesterol and hepatic function[1]	
Intervention P6-P9	*Drug treatment* - 50mg/day atenolol; 25mg/day hydrochlorothiazide, 30mg/day lovastatin and 100 mg/day aspirin	

Model variables	Assumption	Sources
	Provider visits - 4 health-care provider visits/year; 1.5 outpatient visits/year for health education *Laboratory tests* - annual renal function, lipid profile, and blood sugar tests; total cholesterol and hepatic function[1]	Cost Containment System (AHCCCS) Fee Schedule (www.state.me.us/wcb/departments/omrs/1999%20MFS/radiology.htm)
Upper gastrointestinal bleeding	Secondary-level hospital stay of 2.7 days for GBD non-A subregions and tertiary-level hospital stay of 4-8 days for GBD A subregions	Hay, 1996[8]
Effectiveness		
Intervention P1 and P2	33% reduction in difference between actual SBP and 115 mm Hg	References [9-24]
Intervention P3 and P4	20% reduction in total blood cholesterol	Collins, et al, 2002[25]
Intervention P5	Combined effect of P2 and P3	
Intervention P6-P9	Combined effect of P2 and P3 with additional 20% reduction of absolute risk for antiplatelet therapy	Anon[26]
Intervention C1-C4	Effects of N4 evaluated first, then the reduced number of people at risk of a cardiovascular event are subject to the same costs and effects as with P6-9	

GBD = Global Burden of Disease   [1]Only for GBD A subregions

1. Murray CJL, Lopez AD. The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. 1ed. Cambridge, MA: Harvard University Press; 1996.

2. Hernandez-Diaz S, Rodriguez LA. Incidence of serious upper gastrointestinal bleeding/perforation in the general population: review of epidemiologic studies. *J Clin Epidemiol* 2002; 55: 157-63.

3. World Health Organization. The World Health Report 2002: Reducing risks, promoting healthy life. Geneva: World Health Organization; 2002.

4. Molarius, A., Tuomilehto, J., Kuulasmaa, K., and for the WHO MONICA Project. Quality assessment of data on hypertension control in the WHO MONICA Project. 1998. Geneva: World Health Organization. (available at www.ktl.fi/publications/monica/hyperten/hbpdrug.htm).

5. Tolonen, H., Kuulasmaa, K., and for the WHO MONICA Project. Quality assessment of data on use of aspirin in the WHO MONICA Project. 1999. Geneva: World Health Organization. (available at www.ktl.fi/publications/monica/aspirin/aspirinqa.htm).

6. Tolonen, H., Ferrario, M., Minoja, M., and for the WHO MONICA Project. Quality assessment of data on awareness and treatment of high cholesterol

in the WHO MONICA Project. 1999. Geneva: World Health Organization. (available at www.ktl.fi/publications/monica/hich/hchdrug.htm).

7. Adam T, Evans DB, Murray CJL. Econometric estimation of country-specific hospital costs. *Cost-effectiveness and Resource Allocation* 2003; 1.

8. Hay JA, Lyubashevsky E, Elashoff J, Maldonado L, Weingarten SR, Ellrodt AG. Upper gastrointestinal hemorrhage clinical--guideline determining the optimal hospital length of stay. *Am J Med* 1996; 100: 313-22.

9. Coope J, Warrender TS. Randomised trial of treatment of hypertension in elderly patients in primary care. *Br Med J (Clin Res Ed)* 1986; 293: 1145-51.

10. Amery A, Birkenhager W, Brixko P, Bulpitt C, Clement D, Deruyttere M, De Schaepdryver A, Dollery C, Fagard R et al. Mortality and morbidity results from the European Working Party on high blood pressure in the elderly trial. *Lancet* 1985; 1: 1349-54.

11. Hypertension-Stroke Cooperative Study Group. Effect of antihypertensive treatment on stroke recurrence. *JAMA* 1974; 229: 409-18.

12. MRC Working Party. Medical Research Council trial of treatment of hypertension in older adults: principal results. *BMJ* 1992; 304: 405-12.

13. Medical Research Council Working Party. MRC trial of treatment of mild hypertension: principal results. *Br Med J (Clin Res Ed)* 1985; 291: 97-104.

14. Helgeland A. Treatment of mild hypertension: a five year controlled drug trial. The Oslo study. *Am J Med* 1980; 69: 725-32.

15. Leren P, Helgeland A. Oslo hypertension study. *Drugs* 1986; 31 Suppl 1: 41-5.

16. SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). *JAMA* 1991; 265: 3255-64.

17. Anon. Effects of treatment on morbidity in hypertension. Results in patients with diastolic blood pressures averaging 115 through 129 mm Hg. *JAMA* 1967; 202: 1028-34.

18. Anon. Effects of treatment on morbidity in hypertension. II. Results in patients with diastolic blood pressure averaging 90 through 114 mm Hg. *JAMA* 1970; 213: 1143-52.

19. Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med* 2000; 342: 145-53.

20. MacMahon S, Sharpe N, Gamble G, Clague A, Mhurchu CN, Clark T, Hart H, Scott J, White H. Randomized, placebo-controlled trial of the angiotensin-converting enzyme inhibitor, ramipril, in patients with coronary or other occlusive arterial disease. PART-2 Collaborative Research Group. Prevention of Atherosclerosis with Ramipril. *J Am Coll Cardiol* 2000; 36: 438-43.

21. Anon. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6,105 individuals with previous stroke or transient ischaemic attack. *Lancet* 2001; 358: 1033-41.

22. Teo KK, Burton JR, Buller CE, Plante S, Catellier D, Tymchak W, Dzavik V, Taylor D, Yokoyama S et al. Long-term effects of cholesterol lowering and angiotensin-converting enzyme inhibition on coronary atherosclerosis: The Simvastatin/Enalapril Coronary Atherosclerosis Trial (SCAT). *Circulation* 2000; 102: 1748-54.

23. Pitt B, Byington RP, Furberg CD, Hunninghake DB, Mancini GB, Miller ME, Riley W. Effect of amlodipine on the progression of atherosclerosis and the occurrence of clinical events. PREVENT Investigators. *Circulation* 2000; 102: 1503-10.

24. Staessen JA, Fagard R, Thijs L, Celis H, Arabidze GG, Birkenhager WH, Bulpitt CJ, de Leeuw PW, Dollery CT et al. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. The Systolic Hypertension in Europe (Syst-Eur) Trial Investigators. *Lancet* 1997; 350: 757-64.

25. Collins R, Armitage J, Parish S, Sleight P, Peto R. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; 360: 7-22.

26. Anon. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002; 324: 71-86.

Annex 5   Relative risks of CVD events for unit changes in systolic blood pressure, total blood cholesterol, body-mass index, and prevalence of long-term smokers

	Relative risk in age group (years)				
	30-44	45-59	60-69	70-79	⊘80
Ischaemic heart disease					
SBP (mm Hg)	1.07	1.05	1.03	1.02	1.01
Total blood cholesterol (mmol/L)	3.65	2.08	1.55	1.42	1.42
Body mass index (kg/m^2)	1.11	1.09	1.05	1.04	1.03
Stroke					
SBP (mm Hg)	1.09	1.07	1.05	1.03	1.02
Total blood cholesterol(mmol/L)	1.48	1.35	1.25	1.17	1.09
Body mass index (kg/m^2)	1.19	1.09	1.06	1.06	1.02
Smoking and cardiovascular death					
Male	2.43	2.43	1.84	1.70	1.38
Female	2.18	2.18	2.12	1.70	1.31

Annex 6 Percentage reduction in systolic blood pressure for 15% and 30% salt reduction intake in three regions

Age-group (years)	Reduction in SBP (%)											
	AmrB				EurA				SearD			
	15% reduction		30% reduction		15% reduction		30% reduction		15% reduction		30% reduction	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
30-44	1.1%	0.9%	2.1%	1.6%	1.3%	1.1%	2.6%	1.9%	1.1%	1.0%	2.1%	1.9%
45-59	1.3%	1.1%	2.7%	2.2%	1.6%	1.3%	3.1%	2.5%	1.4%	1.3%	2.8%	2.4%
60-69	1.7%	1.4%	3.4%	2.9%	1.9%	1.6%	3.8%	3.2%	1.9%	1.7%	3.7%	3.3%
70-79	2.0%	1.7%	4.0%	3.4%	2.2%	1.8%	4.4%	3.7%	2.2%	2.0%	4.5%	3.9%
≥80	2.4%	2.0%	4.8%	3.9%	2.6%	2.1%	5.2%	4.3%	2.7%	2.4%	5.4%	4.8%

These estimates were made by applying the relation between sodium intake and blood pressure as estimated by Law et al[1-3] to the WHO age, sex and region subgroups,[4] there have been no reanalyses of the strength of this association in this paper

References

1. Law MR, Frost CD, Wald NJ. By how much does dietary salt reduction lower blood pressure? I—Analysis of observational data among populations. *BMJ* 1991; **302**:811-5.

2. Frost CD, Law MR, Wald NJ. By how much does dietary salt reduction lower blood pressure? II—Analysis of observational data among populations. *BMJ* 1991; **302**:815-8.

3. Law MR, Frost CD, Wald NJ. By how much does dietary salt reduction lower blood pressure? III—Analysis of observational data among populations. *BMJ* 1991; **302**:819-24.

4. Lawes C, Feigin V, Rodgers A. Estimating reductions in blood pressure following reductions in salt intake by age, sex and WHO region. Auckland: Clinical Trials Research Unit, University of Auckland; 2002.

8   GENERALISED COST-EFFECTIVENESS
         ANALYSIS: AN AID TO
         DECISION-MAKING IN HEALTH[1]

RAYMOND C.W. HUTUBESSY, ROB M.P.M. BALTUSSEN, TESSA
TAN-TORRES EDEJER, DAVID B. EVANS

Abstract

Health economics literature provides ample evidence for existing inefficiencies in health. Economic appraisal seeks to improve efficiency by guiding policy-makers in how scarce resources can be used to derive the greatest possible social benefit. In the past many cost-effectiveness (CE) studies have addressed sector-wide cost-effectiveness in health. However, as described in this paper, current studies suffer from a number of shortcomings, including the inability to assess the current mix of interventions, low generalisability and inconsistent methodological approaches. Most importantly, it is argued that the current incremental approach to cost-effectiveness analysis (CEA) does not provide decision-makers with sufficient guidance for sector-wide priority setting in health. Instead, a broader complementary sectoral approach is proposed via the application of a generalised CEA framework that allows examination of existing inefficiencies in health systems. The wide variations in cost-effectiveness ratios observed among interventions that are currently in use, suggest there is considerable room to improve efficiency by moving from inefficient interventions to efficient interventions that are underutilised. This information will contribute to a more informed debate on resource allocation in the long term.

Key words: cost-effectiveness analysis, priority setting in health care, resource allocation

Introduction

Cost-effectiveness analysis (CEA) provides one means by which decision-makers may assess and potentially improve the performance of health systems. This process helps ensure that resources devoted to health systems are achieving the maximum possible benefit in terms of outcomes that people value. Over the past three decades there has been an exponential growth in the number of economic appraisals performed in health. Following standard textbooks on economic evaluations, most of these CEA studies pursue an incremental approach that compares the additional costs of an intervention over current practice with additional health benefits(1–3). Such an incremental approach, however, is unable to provide policy-makers with all the necessary information relating to questions like: do the resources currently devoted to health achieve as much as they could? or, how best to use additional resources if they become available? Firstly, incremental analysis does not allow examination of whether current practice is efficient and should have been done in the first place, and secondly, it is not generalisable across settings as it is specific to the starting point (4).

This paper proposes a broader sectoral approach via the application of a generalised CEA framework, which also allows examination of existing inefficiencies in the health system. The wide variations in CE ratios observed among interventions that are currently in use suggest there is considerable room to improve efficiency by moving from inefficient interventions currently in use to efficient interventions that are underutilised (4). For developing countries in particular the reallocation of scarce financial resources is most important (5). The generalised CEA framework compares interventions to a common counterfactual or to a situation of "doing nothing". This allows both existing and new interventions to be analysed and cost-effectiveness results to be more generalisable across settings. The proposed framework focuses on the general use of cost-effectiveness information to inform health policy debates without being completely contextualised.

Here, we review evidence of existing inefficiencies in health systems both at the macro- and micro-level, indicating the need for a reallocation of health resources, and discuss past attempts at sectoral cost-effectiveness in dealing with allocative efficiency problems, including their shortcomings. In a subsequent section the WHO generalised cost-effectiveness framework will be proposed. The implementation and operationalisation of this newly introduced framework will be illustrated by presenting ongoing activities and future plans of the WHO-CHOICE initiative (CHOosing Interventions that are Cost-Effective).

Existing inefficiencies in health care systems

Both at the macro- and the micro-level there is ample evidence on existing inefficiencies in health care. On the macro-level health systems have

multiple goals, yet their defining objective is to improve health. Despite this common aim, health systems with very similar levels of health expenditure per capita can show wide variations in population health outcomes.

The World Health Report (2000) (6) published a first attempt to measure the attainment of goals by the proposed health systems of 191 countries, and considered how well countries were performing given the resources available. Evans and colleagues (6;7) showed that countries like Sri Lanka and China, which are believed to be efficient in producing health, perform less than countries at similar levels of development. Furthermore, the authors concluded that efficiency is positively correlated with health expenditure per capita, especially at low expenditure levels, and that performance sharply increases with expenditure up to about $ 80 per capita a year. These findings can in part be explained by variation in factors outside of health systems, such as the education level of the population. However, a further part can be explained by the fact that some systems devote resources to expensive interventions with small effects on population health, while at the same time low cost interventions which would result in relatively large health improvements are not fully implemented or even ignored.

At the micro-level Tengs (8) and Murray and colleagues (9) argued that health both in the United States and sub-Saharan Africa could be greatly improved by reallocating available resources from interventions that are not cost effective to those that are more cost-effective but not fully implemented. For the case of the United States, it was estimated that a set of 185 currently publicly-funded interventions costs about US$ 214.4 billion, for an estimated saving of 592 000 years of life. Reallocating those funds to the most cost-effective interventions could save an additional 638 000 life-years if all potential beneficiaries were reached (8).

SECTORAL CEA

One approach that has been developed to facilitate policy-makers in decisions to reallocate resources is the construction of a 'league table' that rank-orders interventions by their cost-effectiveness ratios (cost per Quality Adjusted Life Years (QALY)). Many published league tables have been criticised for including only a few interventions (10–12), or only including interventions within one disease area. For example, recently Pinkerton and colleagues (13) constructed league tables to compare interventions to prevent sexual transmission of HIV. Only rarely has the "league table" approach been applied in an explicit broader sectoral perspective, in which CE studies are compared on a wide range of health interventions in a single research effort. Exceptions are the work of Oregon Health Services (14), the Harvard Life Saving Project (15) and World Bank Health Sector Priorities Review (HSPR) (16). What these studies have in common is their aim to allocate health care resources

across many interventions and population groups to generate the highest possible overall level of population health in a single exercise. Each study will be described in more detail hereafter.

World Bank Health Sector Priorities Review Project

The most comprehensive sectoral CEA example on a global level is the World Bank HSPR. In 1987, as recognition surrounding the importance of the HIV epidemic mounted, many groups called upon the health sector of the World Bank to make HIV control their number one priority in health. This provoked a debate on substantive priorities for action in the health sector. The World Bank initiated the HSPR to address this problem. A list of more than twenty important conditions or clusters of conditions was drawn up. The main results of the HSPR are estimates of the long-term average cost-effectiveness of a set of interventions.

Overall, the study showed that categorical assessments such as "primary health care is cost-effective and hospital care is cost-ineffective" are too simplistic: each intervention needs to be evaluated, and one cannot guess cost-effectiveness on the basis of an intervention being curative or preventive or delivered at a given level of the health system. But one of the key findings was that many of the interventions currently undertaken are very expensive ways of improving health, while many of the low cost ways of improving health are not fully funded. This implies there is considerable room to improve allocative efficiency, even if technical efficiency is also low. The World Development Report 1993 (*17*) introduced a global league table of priority health interventions, cardinally ranked by health gain per dollar spent in order to improve efficiency of public health expenditure. Based on this global league table the World Bank proposed a minimum package of basic public and curative health interventions.

Oregon Health Plan

The Oregon Health Plan (OHP) has been widely heralded as an important innovation in American medical care policy. Oregon's pioneering model of prioritising funding for health care through systematically ranking services has drawn an extraordinary amount of national and international attention. The rationing of services rested on an elaborate technical analysis, one that merged cost-effectiveness analysis and medical outcomes research with public participation in policy making decisions.

A Health Services Commission was organised to compile clinical information from physicians, treatment costs and benefit data, and community values from the public. This Commission reduced over 10 000 services to a prioritised list that initially ranked 709 condition and treatment pairs.

The net effect has been to exclude a limited number of services such as medical management of back pain, but to expand coverage of Medicaid

to more people without increasing the budget. The Oregon Health Plan has sparked significant controversy in the US concerning the role of the state in controlling the set of available services in the health sector.

The Harvard Life-Saving Project

A project at the Harvard Center for Risk Analysis was undertaken to review the published literature on the cost-effectiveness of interventions that reduce mortality (*15;18*). It was based on published papers, with minor amendments for differences in methods, and does not include non-fatal health outcomes. As with the HSPR, the study shows a substantial range of cost-effectiveness ratios across interventions that are currently undertaken in the USA. The Harvard Life-Saving Project estimated that this type of reallocation for primary prevention interventions in the USA would save an additional 600 000 years of life annually for the same level of investment. Tengs (*18*) has subsequently shown that reallocating resources from those that are cost-ineffective to those that are cost-effective in the US could save a very considerable number of life years.

REQUIREMENTS FOR SECTORAL CEA

The sectoral CEA studies presented in the previous section have demonstrated major inefficiencies in the current allocation of resources, implying that countries could make significant gains in population health by shifting resources from high-cost, low-effect interventions currently in use, to low-cost, high-effect interventions that are not used, or underutilised. However, it is not always clear how to interpret the results from current CE studies with the aim of sectoral analysis. Some of the difficulties in using current CE studies for sectoral analysis are presented below. These problems (or requirements) should be evaluated for any CE study to be useful to the allocation of resources across a broad range of interventions:

1. Current CE studies are typically based on the incremental or "intervention-mix- constrained" CEA approach, which is appropriate in settings where policy-makers are constrained not only by the availability of resources, but also by the current level of care for the condition under discussion. However, in the long-term where policy is *not* constrained by the current mix of interventions, incremental analysis does not provide best guidance to policy-makers. It ignores the question of whether current interventions themselves are cost-effective. Yet, there is considerable evidence that some interventions currently undertaken are not cost-effective.

2. This form of incremental analysis has limited use for decision-makers in settings other than the one in which a study is undertaken. The starting points for an incremental analysis vary across settings

(according to the current state of infrastructure and the current mix of interventions), while the additional health effects achieved from a given increase in resource use is dependent on what is currently done. This makes it very difficult to generalise CEA results.

3. As has been pointed out in CE literature the comparison of CE results becomes problematic when studies are based on varying costing methods and if economic evaluations are undertaken at different points in time (*1;3;19*). For the sake of sectoral analysis, standardised methods must be used consistently across individual CEA studies to ensure external validity (*20;21*).

4. The World Bank (*17*) estimated that a minimum package of basic public and curative health interventions, each of which was considered to be cost-effective in its own right, would cost US$ 12. Yet this package was unaffordable in many of the poorest countries where health expenditure per capita was as low as US$ 2 (*22;23*). The usefulness of such a general statement might be questioned, and a regional or national league table might be more appropriate. As a minimum, CE studies should identify the full resource implications of implementing interventions identified to be cost-effective; a practice that is slowly beginning to occur in the literature (*22;24–29*). Take the case of malaria: at low levels of health expenditure in a country with a high burden of the disease, case management and prophylaxis for pregnant women would be very cost-effective and affordable. Only with more resources available might impregnated mosquito nets also be implemented (*30*).

5. Current CE studies typically do not consider synergistic effects between interventions. In reality, costs and/or effects of intervention A may influence the costs and/or effects of intervention B because of the relationship between them. Intervention A could be a preventive intervention for tuberculosis (TB) (e.g. BCG vaccination), while intervention B is a treatment for TB (e.g. directly observed short course therapy (DOTS)). BCG vaccination reduces the remaining TB cases which results in fewer patients requiring DOTS and therefore costs for this treatment. Likewise, the health benefits of BCG in the presence of a treatment programme are less because many of the deaths from tuberculosis expected in the absence of treatment will be avoided (*4*).

6. Changing strategies from cost-ineffective to cost-effective interventions will incur transaction costs that are typically not taken into account in current CE studies. That is, it is assumed that what the health system is currently doing or trying to do with its existing infrastructure can be easily redirected. For example, in their Health Resource Allocation Model (HRAM) the authors point out that the presence of existing

capital investments such as staff, buildings and other infrastructure play a major role in budget allocation processes (*31*). Another example is that the cost and effectiveness of delivering antimalarials closer to households will depend critically on whether a network of village workers currently exists (*27*) or on the current and past environmental management of malaria control (*32*). The evidence on transaction costs in the health care sector is scarce. Examples can be found in health care reform initiatives in the United States (*33*), United Kingdom (*34*) and New Zealand (*35*).

7. Finally, current CE studies typically only handle uncertainty around cost-effectiveness ratios (CER) at the individual study level or do not take uncertainty into account at all. When uncertainty ranges around CERs of different interventions overlap, the question is how decision-makers should interpret this information when allocating resources across a large number of interventions. For example, the World Development Report (WDR) 1993 (*17*) only reported point estimates of the CERs. The league tables proposed in these sectoral studies do not provide information about uncertainty to a decision-maker who is risk averse. In particular, this may be troublesome when a fixed budget applies as there may also be considerable uncertainty about the actual costs of a programme.

GENERALISED CEA FRAMEWORK AND WHO-CHOICE

As discussed above the shortcomings of current CE studies for sectoral priority setting in health care are closely related to the use of league tables in general. Many commentators have cautioned against the unthinking use of league tables because of non-comparability of methods, inappropriate comparators and non-generalisability of results (*10;12;36*). Most of the issues and shortcomings raised are addressed within the newly developed WHO generalised CEA approach (*4*). The proposed framework provides policy-makers with a simple set of results that are generalisable across settings. It does this by evaluating costs and effectiveness of new and existing interventions compared to the starting point of doing none of the current interventions. Importantly, the use of such a common reference removes the constraint that the current intervention mix must be continued, and eliminates differences in starting points which makes the results of incremental analysis difficult to transfer across settings. Only one constraint remains; the budget, which allows simple decision rules to be developed based on the calculated cost-effectiveness ratios. It should be recognised that there is still a need within this approach to elicit incremental CERs between interventions, i.e. generalised CEA builds on and incorporates incremental analysis.

Current CE studies and therefore previous sectoral analyses have been restricted to assessing the efficiency of adding a single new intervention

to the existing set, or replacing one existing intervention with an alternative. The generalised approach is of considerable policy importance. Because the analysis is not constrained by what is already being done, policy-makers now have a tool to revisit and possibly revise past choices made, and they will have a rational basis if they decide to reallocate resources from less to more cost-effective interventions. However, as with current (sectoral) CE attempts it will remain a challenge within the generalised CE approach as to how to deal with additional costs of changing strategies (i.e. transition costs). Furthermore, the use of a common methodology enhances comparability between disease areas and transferability of findings across countries. Bearing in mind that obtaining context-specific cost-effectiveness information is intensive, time consuming and costly, the issue of generalisability of information is important, in particular for low- and middle-income countries.

The interactions between interventions, in terms of both costs and effectiveness, are a major focus within the generalised CEA approach. As explained earlier in the tuberculosis example, interventions that are likely to be delivered together in a way that reduces the unit costs are analysed singly or together, and likewise on the effectiveness side: interventions in which the effectiveness is likely to be altered if delivered with another intervention are also analysed singly and together. This approach approximates more closely the practical situation faced by policy-makers.

WHO-CHOICE introduces stochastic league tables to inform decision-makers about the probability that a specific intervention would be included in the optimal mix of interventions for various levels of resource availability, taking into account the uncertainty around cost and effectiveness of different interventions simultaneously (*37*). This would overcome the shortcomings outlined earlier on uncertainty of the existing sectoral league tables. This information helps decision-makers decide on the relative attractiveness of different mixes of interventions given the resources available. Moreover, stochastic league tables inform policy-makers about the total budget impact of an intervention. More recently similar attempts to incorporate affordability thresholds and uncertainty around CE results have been proposed by other authors (Sendi et al 2002 (*38*); Fenwick et al 2001 (*39*); Laska et al 2002 (*40*)).

The WHO-CHOICE project will provide league tables of the cost-effectiveness of interventions (expressed in terms of cost per healthy life year or disability adjusted life year (DALY) averted) for a group of 17 world sub-regions that have been chosen to ensure maximum amount of comparability between countries in terms of health systems and epidemiological profiles. The league table will cover a range of preventive, curative and rehabilitative interventions clustered with various target populations and disease areas such as "children under five" (e.g. diarrhoeal diseases, food fortification and vaccination programmes); women aged 15–44 (e.g. antenatal care and perinatal care);

adolescents and adults (e.g. cancers, stroke, diabetes, mental disorders, HIV, TB) and diseases affecting all ages (e.g. malaria, blindness).

DISCUSSION AND CONCLUSIONS

For sector-wide priority setting, cost-effectiveness information should be collected in a way that will allow policy-makers to address the policy questions raised earlier in the introduction of this paper: do the resources currently devoted to health achieve as much as they could? and, how best to use additional resources if they become available? It has been shown that current CE studies and therefore sectoral analysis have their limitations, e.g. they do not allow assessment of the current mix of interventions, they are setting-specific, and based on incremental CE information with inconsistent methodologies and typically inappropriate comparators. Generalised CEA used by WHO-CHOICE permits both questions raised on technical and allocative efficiency at sectoral level to be answered and deals with them simultaneously.

In reality, many factors may alter the actual cost-effectiveness of a given intervention programme during implementation. These include the availability of the intervention, mix and quality of inputs, local prices, implementation capacity, underlying organisational structures and incentives, and the supporting institutional framework (41;42). All these obstacles imply that even on the sole criterion of cost-effectiveness, analysis of a health system's potential for getting more health from what it spends needs to begin with the current capacities, activities and outcomes, and consider what steps can be taken from that starting point to add, modify or eliminate services. This is likely to have profound implications for investment if little can be changed simply by redirecting the existing staff, facilities and equipment (31). Since the generalised CEA approach focuses on the general assessment of the costs and health benefits of different interventions in the absence of various highly variable local decision constraints the only remaining constraint using a generalised league table for priority setting is the availability of resources. It will give policy-makers indications of how to plan and organise their health system from a long-term perspective.

But nevertheless, information other than cost-effectiveness league tables is also important, such as: evidence about major causes of ill-health and death; responsiveness of the system to people's non-health needs; and inequalities in health outcomes, responsiveness, and the way in which households contribute financially to the system (43). The debates on the use of CE information from the Oregon experience clearly showed that political, ethical, or social issues can easily take precedence over economic criteria (44;45). To choose the appropriate mix of interventions, cost-effectiveness information is only one of a set of criteria that a health system may be asked to respect. It ought to protect people from financial risk to be consistent with the goal of fair financial

contribution; strive for both horizontal and vertical equity; and, it should spend public funds in favour of the poor (*46*). In addition, what makes setting priorities among interventions particularly difficult is that these different criteria are not always compatible. In particular, efficiency and equity can easily conflict as the costs of treating a given health problem differ among individuals, or because the severity of a disease bears little relation to the effectiveness of interventions against it or to their costs. The application of generalised CEA is one way to ensure that sound evidence on cost and effects is used in the sector-wide policy making process.

REFERENCES

1. Drummond M, McGuire A. *Economic evaluation in health care. Merging theory with practice*. 1st ed. Oxford: Oxford University Press; 2001.

2. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 2nd ed. Oxford: Oxford Univeristy Press; 1997.

3. Gold MR, Siegel JE, Russel LB, Weinstein MC. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996.

4. Murray CJ, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Econ* 2000; 9(3):235–251.

5. Hutubessy RC, Bendib LM, Evans DB. Critical issues in the economic evaluation of interventions against communicable diseases. *Acta Trop* 2001; 78(3): 191–206.

6. WHO. *World Health Report 2000: Health systems: Improving performance*. Geneva: World Health Organization; 2000.

7. Evans DB, Tandon A, Murray CJ, Lauer JA. Comparative efficiency of national health systems: cross national econometric analysis. *BMJ* 2001; 323(7308): 307–310.

8. Tengs TO. *Dying Too Soon: How Cost-Effectiveness Analysis Can Save Lives*. Dallas, US: National Center for Policy Analysis; 1997. NCPA Policy Report.

9. Murray CJL, DeJonghe E, Chum HJ, Nyangulu DS, Salomao A, Styblo K. Cost effectiveness of chemotherapy for pulmonary tuberculosis in three sub-Saharan African countries. *Lancet* 1991; 338(8778):1305–1308.

10. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? *Soc Sci Med* 1993; 37(1):33–40.

11. Birch S, Gafni A. Cost-effectiveness ratios: in a league of their own [see comments]. *Health Policy* 1994; 28(2):133–141.

12. Mason J, Drummond M, Torrance G. Some guidelines on the use of cost effectiveness league tables [see comments]. *BMJ* 1993; 306(6877):570–572.

13. Pinkerton SD, Johnson-Masotti AP, Holtgrave DR, Farnham PG. Using cost-effectiveness league tables to compare interventions to prevent sexual transmission of HIV. *AIDS* 2001; 15(7):917–928.

14. Blumstein JF. The Oregon experiment: the role of cost-benefit analysis in the allocation of Medicaid funds. *Soc Sci Med* 1997; 45(4):545–554.

15. Tengs TO, Adams ME, Pliskin JS, Safran DG, Siegel JE, Weinstein MC et al. Five-hundred life-saving interventions and their cost-effectiveness [see comments]. *Risk Anal* 1995; 15(3):369–390.

16. Jamison DT, Mosley WH, Measham AR, Bobadilla JL. *Disease Control Priorities in Developing Countries*. New York: Oxford University Press; 1993.

17. World Bank. *World development report 1993: Investing in health*. New York: Oxford University Press. 1993.

18. Tengs TO. Enormous variation in the cost-effectiveness of prevention: implications for public policy. *Current Issues in Public Health* 1996; 2:13–17.

19. Jefferson T, Demicheli V. Quality of economic evaluations in health care. *British Medical Journal* 2002; 324(7333):313–314.

20. Walker D, Fox-Rushby J. Economic evaluation of parasitic diseases: a critique of the internal and external validity of published studies. [Review] [44 refs]. *Tropical Medicine & International Health* 2000; 5(4):237–249.

21. Walker D. Cost and cost-effectiveness guidelines: which ones to use? *Health Policy Plan* 2001; 16(1):113–121.

22. Marseille E, Kahn JG, Mmiro F, Guay L, Musoke P, Fowler MG et al. Cost effectiveness of single-dose nevirapine regimen for mothers and babies to decrease vertical HIV-1 transmission in sub-Saharan Africa [see comments]. *Lancet* 1999; 354(9181):803–809.

23. World Bank. *Better Health in Africa*. Washington, DC: World Bank; 1994.

24. Broomberg J, Soderlund N, Mills A. Economic analysis at the global level: a resource requirement model for HIV prevention in developing countries. *Health Policy* 1996; 38(1):45–65.

25. Farmer P, Leandre F, Mukherjee JS, Claude M, Nevil P, Smith-Fawzi MC et al. Community-based approaches to HIV treatment in resource-poor settings. *Lancet* 2001; 358(9279):404–409.

26. Forsythe SS. The affordability of antiretroviral therapy in developing countries: what policymakers need to know. *AIDS* 1998; 12 Suppl 2:S11–S18.

27. Goodman CA, Coleman PG, Mills AJ. Cost-effectiveness of malaria control in sub-Saharan Africa. *Lancet* 1999; 354(9176):378–385.

28. Newell ML, Dabis F, Tolley K, Whynes D. Cost-effectiveness and cost-benefit in the prevention of mother-to-child transmission of HIV in developing countries. Ghent International Working Group on Mother-to-Child Transmission of HIV [editorial; comment]. *AIDS* 1998; 12(13):1571–1580.

29. Soucat A, Levy-Bruhl D, De Bethune, X, Gbedonou P, Lamarque JP, Bangoura O et al. Affordability, cost-effectiveness and efficiency of primary health care: the Bamako Initiative experience in Benin and Guinea. *Int J Health Plann Manage* 1997; 12 Suppl 1:S81–108.

30. WHO. *The World Health Report 1999: Making a Difference*. Geneva, Switzerland: World Health Organization. 1999.

31. Murray CJL, Kreuser J, Whang W. Cost-effectiveness analysis and policy choices: investing in health systems. *Bull World Health Organ* 1994; 72(4):663–674.

32. Utzinger J, Tozan Y, Singer BH. Efficacy and cost-effectiveness of environmental management for malaria control. *Trop Med Int Health* 2001; 6(9):677–687.

33. Sekhri NK. Managed care: the US experience. *Bull World Health Organ* 2000; 78(6):830–844.

34. Ferguson B, Keen J. Transaction costs, externalities and information technology in health care. *Health Econ* 1996; 5(1):25–36.

35. Ashton T. Contracting for health services in New Zealand: a transaction cost analysis. *Soc Sci Med* 1998; 46(3):357–367.

36. Gerard K, Mooney G. QALY league tables: handle with care [see comments]. *Health Econ* 1993; 2(1):59–64.

37. Hutubessy RC, Baltussen RM, Evans DB, Barendregt JJ, Murray CJ. Stochastic league tables: communicating cost-effectiveness results to decision-makers. *Health Econ* 2001; 10(5):473–477.

38. Sendi P, Gafni A, Birch S. Opportunity costs and uncertainty in the economic evaluation of health care interventions. *Health Economics* 2002; 11(1):23–31.

39. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics* 2001; 10(8):779–787.

40. Laska EM, Meisner M, Siegel C, Wanderling J. Statistical determination of cost-effectiveness frontier based on net health benefits. *Health Economics* 2002; 11(3):249–264.

41. Hammer JS. *Economic analysis for health projects*. USA: The World Bank; 1996. Policy Research Working Paper No 611.

42. Peabody JW, Omar Rahman M, Gertler PJ, Luck J, Carter GM, Farley DO. *Policy and health; implications for development in Asia*. Cambridge: Cambridge University Press; 1999.

43. Murray CJ, Frenk J. A framework for assessing the performance of health systems. *Bull World Health Organ* 2000; 78(6):717–731.

44. Callahan D. Ethics and priority setting in Oregon. *Health Aff (Millwood)* 1991; 10(2):78–87.

45. Daniels N. Is the Oregon rationing plan fair? *JAMA* 1991; 265(17):2232–2235.

46. Musgrove P. Public spending on health care: how are different criteria related? *Health Policy* 1999; 47(3):207–223.

9   ETHICAL ISSUES IN THE USE OF
     COST-EFFECTIVENESS ANALYSIS FOR THE
     PRIORITIZATION OF HEALTH CARE
     RESOURCES[*]

DAN W. BROCK

Resources to improve health are and always have been scarce, in the sense that health must compete with other desirable social goals like education and personal security for resources.[2] It is not possible to provide all the resources to health, including health care and health care research, that might provide some positive health benefits without great and unacceptable sacrifices in other important social goods. This should go without saying, and in other areas of social expenditures resource scarcity is not denied, but in health care many people mistakenly persist in denying this fact. It follows from resource scarcity that some form of health care rationing is unavoidable, where by rationing I mean some means of allocating health care resources that denies to some persons some potentially beneficial health care. That rationing may take many forms. In most countries with a national health system it is done through some form of global budgeting for health care. In the United States much rationing is by ability to pay, but in both public programs like the Oregon Medicaid program and in many private managed care plans more systematic efforts to prioritize health care resources have been carried out.

   To many health policy analysts it is an unquestioned, and so generally undefended, assumption that in the face of limited health care resources, those resources should be allocated so as to maximize the health benefits they produce, measured by either the aggregate health status or disease burden of a population. Cost effectiveness analysis (CEA) that compares the aggregate health benefits secured from a given resource expenditure devoted to alternative health interventions is the standard analytic tool for determining how to maximize the health benefits from limited resources. Natural, even self-evident, as this maximization standard may

appear to many health policy analysts and economists, it assumes a utilitarian or consequentialist moral standard, and more specifically standard of distributive justice, and the utilitarian account of distributive justice is widely and I believe correctly taken to be utilitarianism's most problematic feature.

Cost effectiveness analysis comparing alternative health interventions in the quality-adjusted life years (QALYs) produced from a given level of resources constitutes a quantitative method for prioritizing different interventions to improve health. There are many unresolved technical and methodological issues in QALYs and CEA, none of which will be my concern here. My concern will be instead with the ethical issues in the construction and use of CEAs for the prioritization of health care resources. The specific issues that I shall briefly discuss below all constitute potential ethical criticisms of CEA as a normative standard, specifically criticisms concerning justice or equity, and so one might hope concerns for justice or equity could be integrated into these quantitative methodologies. There are at least two reasons, however, for caution, at least in the near term, about the possibility of integrating some of these ethical concerns into cost effectiveness models and analyses. First, although a great deal of work in economics and health policy has gone into the development and validation of measures of health status and the burdens of disease, as well as of cost effectiveness methodologies, much less work has been done on how to integrate concerns of ethics and equity into cost effectiveness measures, although I shall mention one means of doing so later. The theoretical and methodological work necessary to do so remains largely undone. Second, each of the issues of ethics and equity that I take up below remain controversial. Since no clear consensus exists about how each should be treated, there is in turn no consensus about what qualifications or constraints they might justify placing on the cost effectiveness goal of maximizing health.

This second difficulty is not likely to be solely a near term limitation, awaiting further work on the ethical issues that I will identify. Instead, most of these issues represent deep divisions in normative ethical theory and in the ethical beliefs of ordinary people; I believe they are likely a permanent fact of ethical life. As I understand and shall present these ethical issues, in most cases there is not a single plausible answer to them. Even from within the standpoint of a particular ethical theory or ethical view, these issues' complexity means that different answers may be appropriate for a particular issue in the different contexts in which CEAs are used. Thus, what is necessary at this point is work developing more clearly and precisely the nature of the issues at stake, the alternative plausible positions on them together with the arguments for and against those positions. Until much more of this work is done, we will not know how deep the conflicts go and the degree to which any can be resolved.

Norman Daniels and James Sabin have recently argued that because ethical theories and theories of justice are indeterminate and/or in conflict

on some of these issues, we must turn to fair procedures to arrive at practical solutions to them for health policy *(1)*. As practical policy matters that need resolution now they are no doubt correct, and a single quantitative measure or model of equity and justice for health care resource prioritization is certainly not possible now, if it will ever be. But that is not to deny that much important work remains to be done on the substantive issues of equity in health care, and that work should inform the deliberations of those taking part in the fair procedures that we will need to reach practical resolutions and compromises on these issues in real time. What then are some of the main issues of equity raised by cost effectiveness approaches to resource allocation of health care?

FIRST ISSUE: How should states of health and disability be evaluated?

Any CEA in health care requires some summary measure of the health benefits of interventions designed to improve the health status and reduce the burden of disease of a given population. Early summary measures of the health status of populations and of the benefits of health interventions often assessed only a single variable, such as life expectancy or infant mortality. The usefulness of life expectancy or infant mortality rates is clearly very limited, however, since they give us information about only one of the aims of health interventions, extending life or preventing premature loss of life, and they provide only limited information about that aim. They give us no information about another, at least as important, aim of health interventions, to improve or protect the quality of life by treating or preventing suffering and disability.

Multi-attribute measures like the Sickness Impact Profile *(2)* and the SF 36 *(3)* provide measures of different aspects of overall health related quality of life (HRQL) on which a particular population can be mapped, and an intervention assessed for its impact on these different components of health, or HRQL. Since these measures do not assign different relative value or importance to the different aspects or attributes of HRQL, they do not provide a single overall summary measure of HRQL. Thus, if one of two populations or health interventions scores higher in some respect(s) but lower in others, no conclusion can be drawn about whether the overall HRQL of one population, or from one intervention, is better than the other. Much quantitative based resource prioritization requires a methodology that combines in a single measure the two broad kinds of benefits produced by health interventions—extension of length of life and improvements in the quality of life *(4)*.

Typical summary measures of the benefits over time of health interventions that combine and assign relative value to these two kinds of benefits include QALYs and Disability-Adjusted Life Years (DALYs). Measures like QALYs and DALYs require a measure of the health status

of individuals and in turn populations at different points in time, such as the Health Utilities Index (HUI) *(5)* and the Quality of Well-Being Scale (QWB) *(6),* so as to be able to measure the health benefits in terms of changes in HRQL and length of life produced by different health interventions. The construction of any measure like the HUI or QWB requires a two step process: first, different states of disability or conditions limiting HRQL are described; second, different relative values or utilities are assigned to those different conditions.

The determination of a person's or group's different health related conditions in terms of the various areas of function on the HUI or QWB both before and after a particular health intervention is an empirical question, which should be answered by appeal to relevant data regarding the burden of a particular disease and the reduction in that burden that a particular health intervention can be expected to produce. Needless to say, often the relevant data are highly imperfect, but that is a problem to be addressed largely by generating better data, not by ethical analysis.

The second step of assigning different relative values or utilities to the different areas and levels of function described by a measure like the HUI is typically done by soliciting people's preferences for life with the various functional limitations. This raises the fundamental question of whose preferences should be used to determine the relative value of life with different limitations in function and how they should be obtained. The developers of the DALY used the preferences of expert health professionals, in part for the practical reason that they are more knowledgeable about the nature of different health states, but the degree to which various conditions reduce overall HRQL is not a matter to be settled by professional expertise. Moreover, health professionals may have systematic biases that skew their value judgments about quality of life from those of ordinary persons. Other measures like the HUI and QWB use the value judgments of a random group of ordinary citizens to evaluate different states of disability or limitations in function.

A central issue concerning whose evaluations of different states of disability or functional limitation should be used arises from the typical responses of individuals to becoming disabled: adaptation, that is improving one's functional performance through learning and skills development, coping, that is altering one's expectations for performance so as to reduce the self-perceived gap between them and one's actual performance; and adjustment, that is altering one's life plans to give greater importance to activities in which performance is not diminished by disability *(7).* The result is that the disabled who have gone through these processes often report less distress and limitation of opportunity and a higher quality of life with their disability than the non disabled in evaluating the same condition. If the evaluations of disability states by the non disabled are used for ranking different states of health and disability, then disabilities will be ranked as more serious health needs, but these rankings are open to the charge that they are distorted by the ignorance

of the evaluators of what it is like to live with the conditions in question. Moreover, those valuations will assign less value to extending the lives of persons with disabilities. If the evaluations of the disabled themselves are used, however, the rankings are open to the charge that they reflect a different distortion by unjustifiably underestimating the burden of the disability because of the process of adaptation, coping, and adjustment that the disabled person has undergone. Moreover, they will assign less value to prevention or rehabilitation for disability because of the results of this process. The problem here is to determine an appropriate evaluative standpoint for ranking the importance of different disabilities which avoids these potential distortions *(8)*.

Since the preferences for different states of disability or HRQL used to determine their relative values should be informed preferences, it is natural to think that the preferences of those who actually experience the disabilities should be used. Because they should have a more informed understanding of what it is actually like to live with the particular disability in question, we can hope to avoid uninformed evaluations. But this is to miss the deeper nature of the problem caused by adaptation, coping, and adjustment to disabilities.

Fundamental to understanding the difficulty posed by adaptation, coping, and adjustment to disabilities for preference evaluation of HRQL with various disabilities is that neither the nondisabled nor the disabled need have made any mistake in their different evaluations of quality of life with that disability. They arrive at different evaluations of the quality of life with that disability because they use different evaluative standpoints as a result of the disabled person's adaptation, coping, and adjustment. Disabled persons who have undergone this process can look back and see that before they became disabled they too would have evaluated the quality of life with that disability as nondisabled people now do. But this provides no basis for concluding that their pre-disability evaluation of the quality of life with that disability was mistaken, and so in turn no basis for discounting or discarding it because mistaken. The problem that I call the perspectives problem is that the nondisabled and the disabled evaluate the quality of life with the disability from two different evaluative perspectives, neither of which is mistaken. It might seem tempting to use the non-disabled's preferences for assessing the importance of prevention or rehabilitation programs, but the disabled's preferences for assessing the importance of life-sustaining treatments for the disabled, but this ignores the necessity of a single unified perspective in order to compare the relative benefits from, and prioritize, the full range of different health interventions.

Moreover, what weight to give to the results of coping with one's condition may depend on the causes of that condition, for example disease or injury that are no one's fault as opposed to unjust social conditions. Most measures of HRQL include some measure of subjective satisfaction or distress, a factor that is importantly influenced by people's

expectations. In a society which has long practiced systematic discrimination against women, for example, women may not be dissatisfied with their unjustly disadvantaged state, including the health differences that result from that discrimination. The fact that victims are sufficiently oppressed that they accept an injustice as natural and cope with it by reducing their expectations and adjusting their life plans should not make its effects less serious, as measures of HRQL with a subjective satisfaction or distress component would imply.

When measures like the HUI or QWB are applied across different economic, ethnic, cultural, and social groups, the meaningful states of health and disability and their importance in different groups may vary greatly; for example, in a setting in which most work is manual labor limitations in physical functioning will have greater importance than it does in a setting in which most individuals are engaged in non physical, knowledge-based occupations, where certain cognitive disabilities are of greater importance. Different evaluations of health conditions and disabilities as seem to be necessary for groups with significantly different relative needs for different functional abilities, but then cross-group comparisons of health and disability, and of the relative value of health interventions, in those different groups will not be possible. The health program benefits will have been measured on two different and apparently incommensurable valuational scales. These differences will be magnified when summary measures of population health are employed for international comparisons across very disparate countries.

Some of this variability of perspective may be avoided by a focus on the evaluation of disability instead of handicap, as these are traditionally distinguished, such as in the 1980 International Classification of Impairments, Disabilities and Handicaps (ICIDH) *(9)*. The ICIDH understands disabilities as "any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner or within the range considered normal for a human being," whereas handicap is "a disadvantage for a given individual, resulting from an impairment or disability, that limits or prevents the fulfillment of a role that is normal (depending on age, sex, and social and cultural factors) for that individual." There will be greater variability between individuals, groups, and cultures in the relative importance of handicaps than of disabilities since handicaps take account of differences in individuals' roles and social conditions that disabilities do not. But it is problematic whether these differences should be ignored in prioritizing health resources for individuals, groups, and societies, that is, whether disabilities or handicaps are the correct focus for evaluation.

Second Issue: Do all QALYs count equally?

QALYs standardly assume that an additional year of life has the same value regardless of the age of the person who receives it, assuming that

the different life years are of comparable quality. A year of life extension for an infant, a forty-year-old, and an eighty-year-old all have the same value in QALYs produced, and in turn in a cost effectiveness analysis using QALYs, assuming no difference in the quality of the year of life extension. This is compatible, of course, with using age-based quality adjustments for interventions affecting groups of different age patients to reflect differences in the average quality of life of those different groups; for example, if average quality of life in a group of patients of average age 85 is less than that of patients of average age 25, a year of life extension for the 25 year old would have greater value in QALYs than would a year of life extension for the 85 year old.

In the World Bank Study, *World Development Report 1993; Investing in Health, (10),* the alternative DALY measure was developed to measure the burden of disease in reducing life expectancy and quality of life. Probably the most important ethical difference between QALYs and DALYs is that DALYs assign different value to a year of life extension of the same quality, depending on the age at which an individual receives it; specifically, life extension for individuals during their adult productive work years is assigned greater value than a similar period of life extension for infants and young children or the elderly. The principal justification offered for this feature of DALYs was the different social roles that individuals typically occupy at different ages and the typical emotional, physical, and financial dependence of the very young and the elderly on individuals in their productive work years *(11).*

This justification of age-based differences in the value of life extension implicitly adopts an ethically problematic social perspective on the value of health care interventions that extend life, or maintain or restore function, that is, an evaluation of the benefits *to others* of extending an individual's life, or maintaining or restoring his or her function, in addition to the benefit to that individual of doing so. This social perspective is in conflict with the usual focus in clinical decision making and treatment only on the benefits to the individuals who receive the health care interventions in question. Typical practice in health policy and public health contexts is more ambiguous on this point, since benefits to others besides the direct recipient of the intervention are sometimes given substantial weight in the evaluation and justification of health programs; for example, treatment programs for substance abuse are argued to merit high priority because of their benefits in reductions in lost work days and in harmful effects on the substance abusers' family members. This social perspective is ethically problematic because it gives weight to differences between individuals in their social and economic value to others; in so doing, it discriminates against persons with fewer dependencies and social ties, which arguably is not ethically relevant in health care resource allocation. The social perspective justifying the DALY measure is therefore ethically problematic, in a way the alternative QALY measure is not, if the value of health benefits for individuals should focus on the

value to the individuals treated of the health benefits, not on the social value for others of treating those persons. The ethical difficulty here is briefly explored further in the section below on what costs and benefits should count in a CEA.

Giving different value to life extension at different ages, however, might be justified ethically if done for different reasons. For example, Norman Daniels has argued that because everyone can expect to pass through the different stages of the life span, giving different value to a year of life extension at different stages in the life span need not unjustly discriminate against individuals in the way giving different weight to life extension for members of different racial, ethnic, or gender groups would unjustly discriminate *(12)*. Each individual can expect to pass through all the life stages in which life extension is given different value, but is a member of only one race, ethnic group, and gender. Thus, all persons are treated the same at comparable stages of their lives regarding the value of extending their lives, and so the use of DALYs would not constitute unjust age discrimination comparable to gender, ethnic or racial discrimination.

Moreover, individuals, and in turn their society, might choose to give lesser weight to a year of life extension beyond the normal life span than to a year of life extension before one has reached the normal life span based on a conception of what equality of opportunity requires, or on what Alan Williams calls the "fair innings argument" *(13)*. People's plans of life and central long term projects will typically be constructed to fit within the normal life span, and so the completion of these central projects will typically require reaching, but not living beyond, the normal life span *(12;14)*.

Third Issue: What costs and benefits should count in cost effectiveness analyses of health programs?

It is widely agreed that cost effectiveness analyses in health should reflect the direct health benefits for individuals of their medical treatment, such as improving renal function or reducing joint swelling, and of public health programs, such as reducing the incidence of infectious diseases through vaccination programs. The direct costs of medical treatment and public health programs, such as the costs of health care professionals' time and of medical equipment and supplies, should also be reflected. But medical and public health interventions typically also have indirect non health benefits and costs. For example, some disease and illness principally affects adults during their working years, thereby incurring significant economic costs in lost work days associated with the disease or illness, whereas other disease and illness principally affects either young children, such as some infectious diseases, or the elderly, such as Alzheimer's' dementia, who in each case are not typically employed and

so do not incur lost wages or lost work time from illness. Should an indirect economic burden of disease of this sort be given weight in a cost effectiveness analysis used to prioritize between different health interventions?

From an economic perspective, as well as from a broad utilitarian moral perspective, indirect non health benefits and costs are real benefits and costs of disease and of efforts to treat or prevent it, even if not direct health benefits and direct treatment costs; they should be reflected in the overall cost effectiveness accounting of how to use scarce health resources so as to produce the maximum aggregate benefit. A possible moral argument for ignoring these indirect non health costs and benefits in health resource prioritization is grounded in a conception of the moral equality of persons. Giving priority to the treatment of one group of patients over another because treating the first group would produce indirect non health benefits for others (for example, other family members who were dependent on these patients) or would reduce indirect economic costs to others (for example, the employers of these patients who incur less lost work time) could be argued to fail to treat each group of patients with the equal moral concern and respect that all people deserve; in particular, doing so would fail to give equal moral concern and weight to each person's health care needs. Instead, giving lower priority to the second group of patients simply because they are not a means to the indirect non health benefits or cost savings produced by treating the first group of patients gives the second group of patients and their health care needs lower priority simply because they are not a means to these indirect non health benefits or cost savings to others. It would violate the Kantian moral injunction against treating people solely as means for the benefit of others.

In public policy we often use a notion of "separate spheres," which in this case could be used to argue that the purpose of health care and of public health is health and the reduction of disease, and so only these goals and effects should guide health care and public health programs (15;16). There are obvious practical grounds for the separate spheres view from the difficulty of fully determining and calculating indirect benefits and costs. But the Kantian moral argument could serve as a principled moral basis for ignoring indirect benefits and costs in a cost effectiveness analysis to be used to prioritize health resources and interventions that serve different individuals or groups.

FOURTH ISSUE: Should discount rates be applied to health care benefits?

It is both standard and recommended practice in cost effectiveness analyses, within health care and elsewhere, to assume a time preference by applying a discount rate to both the benefits and costs of different programs under evaluation, although the reasons for doing so and the proper rate of

discount are controversial *(17)*. It is important to separate clearly the ethical issue about whether health benefits should be discounted from other economic considerations for discounting, as well as to be clear why the issue is important for health policy. It is not ethically controversial that a discount rate should be applied to economic costs and economic benefits; a dollar received today is worth more than a dollar received 10 years from now because we have its use for those ten years, and there is a similar economic advantage in delaying the incurring of economic costs. The ethical issue is whether a discount rate should be applied directly to changes in life extension and well-being or health. Is an improvement in well-being, such as a specific period of life extension, a reduction in suffering, or an improvement in function, extending, say, for one year of substantially less value if it occurs twenty years from now than if it occurs next year?

Future benefits are appropriately discounted when they are more uncertain than proximate benefits. Proximate benefits, such as restoration of an individual's function, also are of more value than distant benefits if they make possible a longer period of, and thus larger, benefit by occurring sooner. But neither of these considerations require the use of a discount rate—they will be taken account of in the measurement of expected benefits of alternative interventions. The ethical question is whether an improvement in an individual's well-being is of lesser value if it occurs in the distant future than if it occurs in the immediate future, simply and only because it occurs later in time. This is a controversial issue in the literature on social discounting and my own view is that no adequate ethical justification has been offered for applying a discount rate directly to changes in health and well-being, though I cannot pursue the justifications offered by proponents of discounting here. The avoidance of paradoxes that arise when no discount rate is applied or when different discount rates are applied to costs and benefits, has influenced many economists to support use of the same discount rate for costs and benefits *(18)*, but I believe these are properly dealt with not through discounting, but rather through directly addressing the ethical issues they raise, usually about equity between different generations.

The policy importance of this issue is relatively straightforward in the prioritization of health care interventions. Many health care and public health programs take significantly different lengths of time to produce their benefits. Applying a discount rate to those benefits leads to an unwarranted priority to programs producing benefits more rapidly. It results in a program that produces benefits in health and well-being say twenty years into the future being given lower priority than an alternative health care program that produces substantially less overall improvement in health and well-being, but produces that improvement much sooner. Many public health and preventive interventions, for example, vaccination programs and changes in unhealthy behavior, reap their health benefits years into the future. If those benefits are unjustifiably discounted, they will be given lower priority than alternative programs

that produce fewer aggregate benefits. The result is a health policy that produces fewer overall health benefits over time than could have been produced with the same resources.

FIFTH ISSUE: What life expectancies should be used for calculating the benefits of life saving interventions?

In calculating QALYs it is standard practice to take account of differences in the average ages and in turn life expectancies of patients served by different health care programs; for example, a treatment for a life-threatening childhood disease would produce more QALYs than a comparable treatment for a life-threatening disease affecting primarily the elderly. Similarly, accurate estimates of the expected QALYs from different interventions would adjust for differences in the average life expectancies of patients caused by diseases other than those treated by the interventions; for example, an intervention that improved the quality of life of patients with cystic fibrosis, who have a much lower than average life expectancy as a result of their disease, would produce fewer QALYs than an intervention with a comparable improvement in lifetime quality of life for patients with average life expectancies undiminished by disease. This latter case raises difficult issues about discrimination against people with disabilities that I take up later. But there are other differences in the life expectancies of different groups that an accurate estimate of QALYs produced by health interventions serving those groups would seemingly have to reflect; for example, there are significant differences in the life expectancies between different genders, racial and ethnic groups, and socio-economic groups within most countries. Internationally, the differences in life expectancies between different countries are often much larger. Should these differences affect calculations of the QALYs gained by health care and public health interventions that extend life or improve quality of life? An accurate estimate of the additional life years actually produced by those interventions should not ignore differences in life expectancies that the health care interventions will not affect, but the result will be that it is less valuable to save the life of a poor person in an underdeveloped country than a rich person in a developed country.

The differences in life expectancies between different racial, ethnic, and socio-economic groups within a single country, as well as the very large differences between life expectancies in economically developed and poor countries, are often principally the result of unjust conditions and deprivations suffered by those with lower life expectancies. It would seem only to compound those injustices to give less value to interventions that save lives or improve quality of life for groups with lower life expectancies caused by the unjust conditions and deprivations from which they suffer. Differences in life expectancies between the genders,

on the other hand, are believed to rest in significant part on biological differences, not on unjust social conditions. Whether the biologically based component of gender differences in life expectancies should be reflected in measures like QALYs or DALYs is more controversial. For example, on the one hand, the lower life expectancy of men does not result from any independent injustice, but, on the other hand, it is explicit public policy and required by law in the United States to ignore this gender-based difference in most calculations of pension benefits and annuity costs so as to avoid gender discrimination. The developers of the DALY explicitly chose to use a single uniform measure of life expectancy (except for the biological component of the gender difference), specifically that observed in Japan which has the highest national life expectancy, to measure gains from life saving interventions. They justified their choice in explicitly ethical terms as conforming to a principle of "treating like events as like," although the reasoning was not pursued in any detail *(11)*. How this issue is treated can have a substantial impact on the priorities that result from the cost effectiveness analysis, especially at the international level where country differences tend often to be greater than group differences within specific countries.

Each of the preceding five ethical issues can be considered issues in the *construction* of a cost-effectiveness analysis in health care. The other issues I want to briefly note can be considered issues in the *use* of cost effectiveness analysis in health resource prioritization. They are each issues of distributive justice or equity raised by the fact that a cost effectiveness analysis is insensitive to the distribution of health benefits and of the costs of producing them. Yet people's beliefs about equity and justice directly affect the relative priority they assign to different health interventions. One standard response to this point is that a CEA can only be an aid to policy making in general, and health resource prioritization in particular, and that policy makers must take account of considerations of equity in final policy decisions and choices. But as with the ethical issues in the construction of CEAs, much work remains to be done to clarify and assess alternative positions on these issues of equity so the policy choices on them can at least be better informed, even if they remain controversial. Here, there is only space to state four of the main equity issues in the use of CEAs and some of the principal ethical considerations supporting different positions on them *(19)*. After doing that, I shall mention an alternative quantitative methodology that, unlike CEA, incorporates considerations of equity within the quantitative analysis.

Sixth Issue: What priority should be given to the sickest or worst off? *(20)*

It is a commonplace that most theories of distributive justice require some special concern for those who are worst off or most disadvantaged; for

example, it is often said that the justice of a society can be measured by how it treats its least well off members. In the context of health care allocation and the prioritization of health interventions, the worst off with regard to need for the good being distributed might reasonably be thought to be the sickest patients. In many cases, the sickest will be given priority by a CEA comparing treating them as opposed to less sick patients; the sickest have greater possible improvements in HRQL because they begin from a lower HRQL, and so, for example, in comparing fully effective treatments those for the sickest will produce the greater benefits. But in other cases giving priority to the sickest will require a sacrifice in aggregate health benefits. An abstract example makes the point most concisely. Suppose Group A patients have a very serious disease that leaves them with a health utility level of .25 as measured by the HUI, and this would be raised only to .45 with the best available treatment because no treatment is very effective for their disease; for example, patients with severe chronic obstructive pulmonary disease or with severe chronic schizophrenia that is largely resistant to standard pharmacological treatments. A similar number of Group B patients have a health utility level of .60 because they have a considerably less serious disease, but since treatment for their disease is more effective, although no more costly, it would raise their health utility level to .90; for example, patients with asthma, or with milder forms of pulmonary disease or schizophrenia that both leave them less disabled without treatment and are more responsive to treatment. Should we give priority to treating Group B because doing so would produce a 50% greater aggregate health benefit at the same cost, as the CEA standard implies, or to treating Group A who are the sickest? In some empirical studies, both ordinary people and health professionals prefer to sacrifice some aggregate health benefits in order to treat the sickest patients, although the degree of sacrifice they are prepared to make is variable and not statistically reliable *(21)*.

One difficulty raised by this issue is determining what weight to give to this particular aspect of equity—concern for the worst off. Virtually no one would prefer to treat the sickest, no matter how costly their treatment and how small the benefit to them of doing so, and no matter how beneficial and inexpensive treatment for the less sick might be. However, there seems no objective, principled basis for determining how much priority to give the sickest, that is, how much aggregate health benefits should be sacrificed in order to treat or give priority to the sickest. Instead, the most one can say is that most people and many theories of distributive justice have a concern both for maximizing overall benefits with scarce health resources and for helping the worst off or sickest, but there is a large range of indeterminacy regarding the proper trade off between these two concerns when they are in conflict.

One issue in understanding this concern for the worst off important for health care priorities is whether it should focus on who is worst off at a

point in time or instead over an extended period of time, such as a lifetime. When choosing between patients to receive a scarce resource, such as in organ transplantation, it is often plausible to focus on lifetime well being, since otherwise we may give priority to the patient who is worst off at the time the distributive choice is made, but whose lifetime level of well being is far higher than the other patient. Frances Kamm has defended a notion of need in this context according to which the neediest patient is the patient whose life will have gone worst if he or she does not get the scarce resource, such as an organ transplant *(15)*. However, some justifications for giving priority to the worst off may support focusing on the sickest here and now.

What are the ethical justifications for giving priority to the worst off? I can mention only two possibilities here. One is that we must give priority to the worst off in order to avoid increasing the already unjustified disadvantage or inequality they suffer relative to those better off. But it is worth noting that a concern for the worst off is not always the same as a concern to produce equality in outcomes. In the example above of Groups A and B, equality could be achieved by what Derek Parfit has called "leveling down", that is by bringing B's health utility level down to that of A's instead raising A's level up to that of B *(22)*. If equity here is equivalent to equality in outcomes, then if it were not possible to raise A's level above .40 with treatment, equity would seem to support not treating Group B and letting their condition deteriorate until it reached the lower level of Group A. The fact that no one would defend doing this suggests that this aspect of our notion of equity or justice is best captured by the idea of giving priority to improving the condition of the worst off, rather than by a simple concern for equality in outcomes. A different justification for giving priority to treating the sickest, offered by some participants in Nord's research, is that it would be subjectively more important to the sickest to obtain treatment, even if the health benefits they receive from treatment are less than those that would go to the less sick; this justification might support focusing on who is worst off at the point in time at which the decision about who to treat is made, not whose lifetime well-being will be lowest *(21)*.

One further issue concerning the priority to the worst off should be mentioned. In the context of health resource prioritization in health policy it seems natural to understand the worst off as the sickest. But this may not always be correct. At the most fundamental ethical level in our general theories of equity and distributive justice, our concern should be for those who are overall or all things considered worst off, and they will not always be the sickest. It could be argued that giving priority to the worst off in health resource prioritization sometimes requires giving priority to those with the lowest levels of overall well-being, even at some cost to aggregate health benefits produced *and* at the cost of not treating sicker persons whose overall well-being is much higher. A preference for health interventions that raise the level of well-being of those who are worst off

in overall well-being, instead of giving priority to the sickest, might be justified in order not to increase the unjustified disadvantage suffered by those with the lowest overall level of well-being. If, instead, the priority to the worst off in health resource prioritization should focus only on health states and so on the sickest, a justification of this narrowed focus is needed.

SEVENTH ISSUE: When should small benefits to a large number of persons receive priority over large benefits to a small number of persons?

Cost effectiveness and utilitarian standards require minimizing the aggregate burden of disease and maximizing the aggregate health of a population without regard to the resulting distribution of disease and health, or *who* gets what benefits. The issue about priority to the worst off focuses on who gets the benefits. A different issue concerns *what* benefits different individuals get. Some would argue that health benefits are often qualitatively different and so cannot all be compared on a single scale like the HUI, or in turn by a single measure like QALYs, but that is not the issue of concern now. In its most general form the issue about aggregation concerns what ethical limits there are, if any, on aggregating together different size benefits for different persons in comparing and prioritizing different health interventions; CEA accepts no such limits. There are many forms in which this issue can arise which cannot be pursued here *(15),* but the version that has received the most attention, and which Daniels has called the aggregation problem, is when, if ever, large benefits to a few individuals should take priority over greater aggregate benefits to a different and much larger group of individuals, each one of whom receives only a small benefit. This issue arises when a very serious disease or condition for those affected that is also very costly to prevent or treat is compared with a much more prevalent disease or condition that both has a very small impact on each individual affected and is very inexpensive to treat or prevent in any one individual. Applying cost effectiveness or utilitarian standards, preventing or treating the very prevalent but low impact disease or condition at a given cost will receive higher priority when doing so produces greater aggregate benefits than using the same funds to treat or prevent the disease or condition that has a very great impact on each individual affected. The example that received considerable attention in the United States arose in the Oregon Medicaid priority setting process where capping teeth for exposed pulp was ranked just above an appendectomy for acute appendicitis, a potentially life-threatening condition. Because an appendectomy is approximately 150 times as expensive as capping a tooth for exposed pulp, the aggregate benefit of capping a tooth for 150 patients was judged to be greater than the benefit of an appendectomy for one patient. Since Medicaid coverage decisions were to be made according to the list of

treatment/condition pairs ranked in terms of their relative cost effectiveness, it could have turned out, depending on the overall level of resources available to the Medicaid program, that tooth capping would have been covered but appendectomies not covered.

This result, and other less extreme cases like it, was highly counter-intuitive and unacceptable to most people, whose intuitive rankings of the relative importance or priority of health interventions are based on one-to-one comparisons, for example of one tooth capped as opposed to one appendectomy performed. In the face of these results Oregon made a fundamental change in its prioritization methodology, abandoning the cost effectiveness standard in favor of a standard that did not take account of differences in costs. This was not a minor problem requiring tinkering at the margins of the CEA standard, but a fundamental challenge to it and so required a fundamental revision in it.

Yet it is by no means clear that no such aggregation can be ethically justified. The very case that precipitated Oregon's Medicaid revision was a 12 year old boy in need of a bone marrow transplant as the only effective chance to save his life. Oregon denied coverage under its Medicaid program on the grounds that it could do greater good by using its limited resources to improve prenatal care for pregnant women, in this case giving higher priority to small benefits to many over a potentially much larger benefit to a few. Moreover, many public policy choices appear to give higher priority to small benefits to many over even life saving benefits to a few; for example, governments in the United States support public parks used by tens or hundreds of thousands of persons, while reducing funding for public hospitals resulting in quite predictable loss of life.

The cost effectiveness or utilitarian standard that permits unlimited aggregation of benefits might be defended by distinguishing between the clinical context in which physicians treat individual patients and the public health and health policy context in which health resource allocation decisions are made that will affect different groups in the population. In the clinical context, physicians forced to prioritize between individual patients typically will first treat the patient who will suffer the more serious consequences without treatment, or who will benefit the most from treatment, even if doing so will prevent her treating a larger number of less seriously ill patients. But from a public health or health policy perspective, it could be argued that the potential overall or aggregate effects of alternative interventions on population health is the appropriate perspective. However, the Oregon experience makes clear that even when allocating public resources for interventions to improve the health of a population it is ethically controversial whether always giving priority to producing the maximum aggregate benefits, even when that is done by giving small benefits to many at the cost of forgoing large benefits to a few, is justified.

Just as with the problem of what priority to give to the worst off, part

of the complexity of the aggregation problem is that for most people some, but not all, cases of aggregation are ethically acceptable and equitable. The theoretical problem then is to develop a principled account of when, and for what reasons, different forms of aggregation satisfy requirements of equity and when they do not *(15)*. There is no consensus on this issue either among ordinary persons or within the literature of health policy or ethics and political philosophy. As with the problem about priority to the worst off, the complexities of this issue have received relatively little attention in bioethics and moral and political philosophy, and there is much difficult but important work to be done.

EIGHTH ISSUE: The conflict between fair chances and best outcomes

The third ethical issue in the use of CEA for health resource utilization that I will mention here has been characterized as the conflict between fair chances and best outcomes *(19)*. The conflict is most pressing when the health intervention is life saving and not all those whose lives are threatened can be saved, but it arises as well when threats are only to individuals' health and well-being. In the context of health care, this issue first received attention in organ transplantation where there is a scarcity of life saving organs such as hearts and lungs resulting in thousands of deaths each year of patients on waiting lists for an organ for transplant; an abstract example from transplantation can illustrate the issue most clearly and succinctly *(23)*.

   Suppose two patients are each in need of a heart transplant to prevent imminent death, but there is only one heart available for transplant. Patient A has a life expectancy with a transplant of ten years and patient B has a life expectancy with a transplant of nine years (of course, precise estimates of this sort are not possible, but the point is that there is a small difference in the expected benefits to be gained depending on which patient gets the scarce organ), with no difference in their expected quality of life. Maximizing health benefits or QALYs, as a CEA standard requires, favors giving the organ to patient A, but patient B might argue that it is unfair to give her no chance to receive the scarce heart. Just as much as A, she needs the heart transplant for life itself and will lose everything, that is her life, if she does not receive it. It is unfair, B might argue, to give the organ to A because the quite small increment in expected benefits from doing so is too small to justly determine who lives and who dies. Instead, she argues, each of them should receive a fair chance of getting the organ and having their health needs met; in this case, that might be done by giving each an equal chance of receiving the transplant through some form of random selection between them, or by a weighted lottery that gives the patient who would benefit more some greater likelihood of being selected to receive the organ, but still gives the

patient who would benefit less some significant chance of getting it instead *(15;23;24)*.

Most prioritization and rationing choices arise not from physical scarcity of the needed health resource, as in organ transplantation, but from economic scarcity, limits in the money society devotes to health care. Will this issue of equity arise in health resource prioritization and allocation choices forced by economic scarcity? Two considerations will often mitigate the force of the ethical conflict between fair chances and best outcomes there. First, allocation of resources in health care is typically not an all or nothing choice, as in the case of selecting recipients for scarce organs, but is usually a matter of the relative priority for funding to be given to different health programs or interventions. That one health program A promises a small gain in aggregate health benefits over a competing program B need not entail that A is fully funded and B receives no funding, but only that A should receive higher priority for, or a higher level of, funding than B. Persons with the disease or condition that A treats will have a somewhat higher probability of being successfully treated than will those who have the disease or condition that B treats; in the case of prevention, those at risk of A will have a somewhat higher probability of successful prevention than will those at risk of B. When there is significant resource scarcity this will involve some sacrifice in aggregate health benefits that might have been produced by always preferring the more cost effective alternative. But doing so means that individuals who are served by B have no complaint that the small difference in expected benefits between programs A and B unfairly prevents them from having their health needs met at all. Instead, the small difference in expected benefits between programs A and B need only result in a comparably small difference in the resources devoted to A and B; it is not obvious that this is unfair to those patients served by B, whose needs are somewhat less well served than patients in program A because of B's lower priority and level of funding.

The second consideration that may mitigate some the conflict between fair chances and best outcomes in health resource prioritization forced by economic scarcity is that often, probably usually, the diseases and health problems to be treated or prevented are not directly life threatening, but instead only impact on individuals' quality of life, and often for only a limited period of time. In these cases, the difference in health benefits between individuals who receive a needed health intervention that is given a higher priority and individuals who do not receive a needed health intervention because their condition is given lower priority, is much less, making the unfairness arguably less compelling.

These two considerations may mitigate, but they do not fully avoid, the conflict between fair chances and best outcomes in prioritization decisions about health interventions forced by economic scarcity. When a more cost effective health program is developed for one population instead of a different less cost effective health program for a different

population, individuals who would have been served by the second program will have a complaint that they did not have a fair chance to have their needs served only because of a small gain in the benefits that are produced by the first program. The fair chances versus best outcome conflict will arise in prioritizing health interventions in health policy; how this conflict can be equitably resolved is complex, controversial, and unclear.

NINTH ISSUE: Does use of CEA to set health care priorities unjustly discriminate against the disabled?

In several contexts using CEA to set health care priorities will result in assigning lower priority to both life extending and quality of life improving treatment for disabled than nondisabled persons with the same health care needs *(8;25)*. Here are five such contexts. First, since already disabled persons have a lower HRQL from their disability than nondisabled persons, treatment that extends their life for a given number of years produces fewer QALYs than the same treatment that extends the life of a nondisabled person for the same number of years. Second, if two groups of patients with the same HRQL have the same need for a life sustaining or quality of life improving treatment, but one will be restored to normal function and the other will be left with a resultant disability, more QALYs will be produced by treating the first group. Third, persons with disabilities often have a lower life expectancy because of their disability than otherwise similar nondisabled persons. As a result, treatments that prevent loss of life or produce lifetime improvements in quality of life will produce fewer QALYs when given to disabled than to nondisabled persons with the same health care needs. Fourth, disabilities often act as comorbid conditions making a treatment less beneficial in QALYs produced for disabled than for nondisabled persons with the same health care needs. Fifth, the presence of a disability can make treatment of disabled persons more difficult and so more costly than for nondisabled persons with the same health care needs; the result is a lower cost effectiveness ratio for treating the disabled persons.

In each of the five cases above, disabled persons have the same medical and health care need as nondisabled persons, and so the same claim to treatment on the basis of their needs. But treating the disabled person will produce less benefit, that is fewer QALYs, *because of their disability* than treating the nondisabled. Thus, their disability is the reason for their receiving lower priority for treatment. This at least arguably fails to give equal moral concern to disabled persons' health care needs and is unjust discrimination against them on grounds of their disability. Indeed, United States Health and Human Services Secretary Louis Sullivan denied

Oregon's initial request for a waiver of federal regulations for its proposed revisions to its Medicaid plan on the grounds that Oregon's method of prioritization of services was in violation of the Americans with Disabilities Act (ADA).[3] Sullivan cited some of the five kinds of cases I noted above in support of that position, and Oregon in turn made essentially ad hoc revisions in its ranking to avoid the putative violation of the ADA.

Disabled persons charge that in cases like the first I cited above concerning life saving treatment, the implication of use of CEA to prioritize health care is that saving their lives, and so their lives themselves, have less value than nondisabled persons' lives. They quite plausibly find that implication of CEA threatening and unjust. There are means of avoiding these problems about discrimination against persons with disabilities, but they involve abandoning fundamental features of CEAs. For example, one response to the first case cited above would be to give equal value to a year of life extension, whatever the quality of that life, so long as it is acceptable to the person whose life it is *(15)*. But that has problematic implications too since, for example, a small percentage of persons in surveys say they would want their lives sustained even if they were in a persistent vegetative state. I cannot pursue the issues further here, but I believe the problem of whether CEA unjustly discriminates against the disabled is a deep and unresolved difficulty for use of CEA and QALYs to prioritize health care.

The sixth, seventh, and eighth issues above all raise possible criticisms of the maximization standard embodied in CEA; in each case, the claim is that equity requires attention to the distribution of health benefits and costs to distinct individuals. Steadfast utilitarians or consequentialists will reject the criticisms and hold fast to the maximization standard. But most people will accept some departure from the maximization standard of CEA; there are two broad strategies for how to do so. The first and probably most common is to propose CEA as an aid to policy makers who must make prioritization and allocation choices in health care, but then to remind those policy makers that they must take account of these considerations of equity as well in their decision making; this may be, but usually is not, accompanied by some guidance about alternative substantive positions, and reasons in support of them, on the equity issues. Moreover, some use of CEA in health policy and health program evaluation does not raise these last three issues of equity; for example, CEA of alternative treatments that each have uniform but different benefits for a group of patients with a particular medical condition. And outside of a CEA, either QALYs or DALYs can be used for evaluating alternative interventions, or for monitoring changes over time in health status or the burdens of disease, in a given group or population.

The second strategy for responding to concerns about equity seeks to develop a quantitative tool that measures the specific weight people give to different equity concerns in comparing interventions that raise issues

of distributive justice because they serve different individuals or benefit individuals differently. The most prominent and promising example is the "person trade-off" approach which explicitly asks people how many outcomes of one kind they consider equivalent in social value to X outcomes of another kind, where the outcomes are for different groups of individuals (26). For example, people can be asked, as in our earlier example, to compare treatment A for very severely ill patients who are at .25 on the HUI without treatment and who can be raised only to .45 with treatment, with treatment B of less severely ill patients who are at .60 and can be raised to .90 with treatment; filled out detailed examples, of course, will make the comparisons more understandable. Respondents are then asked how many patients treated with A would be equivalent in social value to treating 100 patients with B. Answers to questions of this form will tell us in quantitative terms how much importance people give to treating the sickest when doing so conflicts with maximizing aggregate health benefits.

The person trade-off approach is designed to permit people to incorporate concerns for equity or distributive justice into their judgments about the social value of alternative health programs. There has been relatively little exploration and use of this methodology in health care evaluation in comparison with the mass of methodological work on and studies of aggregate QALYs and CEAs, in part because many health policy analysts and health economists assume, often with little or no argument, that the social value of health programs is the sum of the individual utilities produced by the program. As I noted in the introduction to the paper, the early stage we are now at in the development and use of the person trade-off approach is a reason for caution at the present time about using it to settle issues of equity in health resource prioritization. While the utilitarian assumption in CEA is rejected in most philosophical work on distributive justice, as well as in the preferences most ordinary people express for different health outcomes and programs, I also noted in the introduction a second more important reason for caution about bringing considerations of equity into health policy decision making through a quantitative methodology like the person trade-off methodology—the issues of distributive justice that must be addressed by equitable health resource prioritization represent deep and long-standing divisions in moral and political philosophy about which there is not now, and may never be, anything approaching consensus. There is a strong case to be made, though I cannot pursue it here, that important value conflicts about justice of this sort should be addressed in public, democratic political processes, or in fair, participatory and accountable procedures within private institutions like managed care organizations (1). The person trade-off method can be a useful aid to those deliberative decision making processes in providing more structure and precision to different people's views about equity in health care resource prioritization and trade-offs, but it is not a substitute

for that deliberation. Despite these briefly noted reservations, I do emphasize that for purposes of resource prioritization and allocation, the person trade-off approach is the proper perspective, in comparison with CEA, because it correctly reflects that the choices are typically about how health benefits and costs are distributed to different individuals.

CONCLUSION

I have distinguished above nine distinct issues about equity and justice that arise in the construction and use of cost effectiveness analysis to minimize the burdens of disease and to maximize health outcomes. In each case the concern for equity is in my view valid and warrants some constraints on a goal of unqualified maximization of health outcomes. There has not been space here to pursue at all fully any of these nine issues regarding equity and justice—each is complex, controversial, and important. In each case, my point has been that there are important ethical and value choices to be made in constructing and using the measures; the choices are not merely technical, empirical, or economic, but moral and value choices as well. Each requires explicit attention by health policy makers using CEA. In a few cases I have indicated my own view about how the potential conflict between equity and utilitarian maximization might be resolved, but in other cases I have simply summarized briefly some arguments for giving the particular concern about equity some weight when it conflicts with maximization of utility. For some of these issues, the literature and research is at a relatively early stage and one cannot be confident about how the issues should be resolved or even about the range of plausible positions and supporting reasons on them. However, this is not grounds for ignoring the issues, but instead for getting to work on them and for ensuring that they receive explicit attention and deliberation in decisions about health resource prioritization and allocation.

NOTES

1 This paper draws heavily on my "Considerations of Equity in Relation to Prioritization and Allocation of Health Care Resources," in *Ethics, Equity and Health for All,* eds Z. Bankowski, J.H. Bryant and J. Gallagher (Geneva: CIOMS, 1997) and "Ethical Issues in the Development of Summary Measures of Population Health States" in *Summarizing Population Health: Directions for the Development and Application of Population Metrics* (Washington DC: National Academy Press, 1998).

2 Interventions that would improve health should be understood broadly, and in particular extend substantially beyond health care. It is widely agreed that other factors such as improved sanitation and economic conditions have contributed more to the health gains of the past century than has health care. However, in this paper I shall largely confine myself to health care interventions.

3 Unpublished letter from Secretary of Health and Human Services Louis Sullivan to Oregon Governor Barbara Roberts, August 3, 1992.

REFERENCES

1. Daniels N, Sabin J. Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Philos Public Aff* 1997; 26(4):303-350.

2. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care* 1981; 19(8):787-805.

3. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992; 30(6):473-483.

4. Brock DW. Quality of life measures in health care and medical ethics. In: Sen A, Nussbaum M, eds. *The quality of life*. Oxford: Oxford University Press; 1992.

5. Torrance G, Zhang Y, Feeny D, Furlong W, Barr R. *Multi attribute preference functions for a comprehensive health status classification system.* Hamilton, Ontario: McMaster University, Center for Health Economics and Policy Analysis; 1992. Working Paper No. 92-18.

6. Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Services Research* 1988; 23(2):203-235.

7. Murray CJL. Rethinking DALYs. In: Murray CJ, Lopez AD, eds. *The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020.* Cambridge, MA: Harvard University Press; 1996, p. 1-98.

8. Brock DW. Justice and the ADA: does prioritizing and rationing health care discriminate against the disabled? *Soc Philos Policy* 1995; 12(2):159-185.

9. *International classification of impairments, disabilities and handicaps.* Geneva: World Health Organization; 1980.

10. World Bank. *World development report* 1993: Investing in health. New York: Oxford University Press; 1993.

11. Murray CJL. Quantifying the burden of disease: the technical basis for disability-adjusted life years. In: Murray CJL, Lopez AD, eds. *Global comparative assessments in the health sector: disease burden, expenditures and intervention packages.* Geneva: World Health Organization; 1994.

12. Daniels N. Am I my parents keeper? *An essay on justice between the young and the old.* New York: Oxford University Press; 1988.

13. Williams A. *Intergenerational equity: an exploration of the 'fair innings' argument.* Health Economics 1997; 6(2):117-132.

14. Brock DW. *Justice, health care, and the elderly.* Philos Public Aff 1989; 18(3):297-312.

15. Kamm FM. *Morality/mortality. Volume One. Death and whom to save from it.* Oxford: Oxford University Press; 1993.

16. Walzer M. Spheres of justice: *Defense of pluralism and equality.* New York: Basic Books; 1983.

17. Gold MR, Siegel JE, Russel LB, Weinstein MC. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996.

18. Keeler EG, Cretins. Discounting of life saving and other nonmonetary effects. *Management Science* 1983; 29(3):300-306.

19. Daniels N. Rationing fairly: programmatic considerations. *Bioethics* 1993; 7(2-3):224-233.

20. Brock DW. Priority to the worst off in health care resource prioritization. In: Battin M, Rhodes R, Silvers A, eds. *Health care and social justice*. New York: Oxford University Press (forthcoming); 2003.

21. Nord E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. *Health Policy* 1993; 24(3):227-238.

22. Parfit D. Equality or priority. The Lindley Lecture. Copyright, Department of Philosophy, University of Kansas.

23. Brock DW. Ethical issues in recipient selection for organ transplantation. In: Mathieu D, ed. *Organ substitution technology: Ethical, legal, and public policy issues*. Boulder and London: Westview Press; 1988.

24. Broome J. Selecting people randomly. *Ethics* 1984; 95(1):38-55.

25. Brock DW. Health care resource prioritization and discrimination against persons with disabilities. In: Francis L, Silvers A, eds. *Americans with disabilities*. New York: Rutledge; 2000.

26. Nord E. *Cost-value analysis in health care: Making sense of QALYs*. New York: Cambridge University Press; 1999.

Index