

Real world designs in economic evaluation : bridging the gap between clinical research and policy making

Citation for published version (APA):

Baltussen, R. M. P. M. (1998). *Real world designs in economic evaluation : bridging the gap between clinical research and policy making*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.19981009rb>

Document status and date:

Published: 01/01/1998

DOI:

[10.26481/dis.19981009rb](https://doi.org/10.26481/dis.19981009rb)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 19 Apr. 2024

REAL WORLD DESIGNS IN ECONOMIC EVALUATION

**BRIDGING THE GAP BETWEEN CLINICAL RESEARCH AND
POLICY MAKING**

Real World Designs in Economic Evaluation
Bridging the gap between clinical research and policy making
Copyright R. Baltussen 1998
Proefschrift Universiteit Maastricht

ISBN 90-9011-882-9

Printed by: Datawyse | Universitaire Pers, Maastricht

REAL WORLD DESIGNS IN ECONOMIC EVALUATION

BRIDGING THE GAP BETWEEN CLINICAL RESEARCH AND POLICY MAKING

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. Dr. A.C. Nieuwenhuijzen Kruseman,
volgens het besluit van het College van Decanen, in het openbaar te verdedigen
op vrijdag 9 oktober 1998 om 12.00 uur

door

Robertus Martinus Petrus Maria Baltussen
geboren te Rijkevoort op 24 april 1968

Promotor:

Prof. Dr. R.M. Leidl

Co-promotor:

Dr. A.J.H.A. Ament

Beoordelingscommissie:

Prof. Dr. W.N.J. Groot (voorzitter)

Dr. J.C. Jager (RijksInstituut voor Volksgezondheid en Milieuhygiëne)

Prof. Dr. R.T.J.M. Janssen (Katholieke Universiteit Brabant)

Prof. Dr. J.A.M. Maarse

Dr. G.G. van Merode

CONTENTS

CHAPTER 1	General introduction	1
CHAPTER 2	Potential biases of economic evaluation for policy making	15
CHAPTER 3	Interpreting potential biases: coping with the impact of age	39
CHAPTER 4	Interpreting potential biases: coping with interactions with other diseases	61
CHAPTER 5	Evaluating the TUMT-TURP trial at the level of the Dutch population	79
CHAPTER 6	The use of modeling in economic evaluation: Pneumococcal vaccination	93
CHAPTER 7	Reporting the results from economic evaluation to policy makers	113
CHAPTER 8	Real world designs in economic evaluation	133
CHAPTER 9	Conclusions	155
SUMMARY	167
SAMENVATTING	171
DANKWOORD	177
CURRICULUM VITAE	181

CHAPTER 1

GENERAL INTRODUCTION

INTRODUCTION

The health care system is continuously being confronted with promising new medical interventions. In the past, most of these new interventions were adopted into clinical practice without much *a priori* knowledge of their performance, except in terms of safety and efficacy. In order to find the most favourable set of interventions for a community, rigorous evaluations must also be made for economic efficiency. Economic evaluation of a proposed intervention is a fairly recent phenomenon, and its importance is growing. This thesis examines how economic evaluation can enhance the efficiency of the health care system. The basic idea of economic evaluation is to compare the costs of different treatment alternatives with their various consequences, in order to determine whether a treatment offers 'good value for money'.¹

BACKGROUND

The principles of economic efficiency and cost-benefit analysis have been detailed by the economists Pareto, Kaldor² and Hicks.³ Economic evaluations of medical interventions began about 30 years ago, as rather crude comparisons of the costs and benefits of health care programs using the human-capital approach, in which benefits were defined as reduced health care costs and increased labour production. Weisbrod provided some pioneering evaluations of public health programs such as measles vaccination for children. He compared the costs of these programs with the benefits, which were defined as the sum of the estimated reduction in subsequent treatment costs plus increases in productivity due to better health and increased survival.⁴ The first cost-effectiveness analysis was published in 1968 by Klarman *et al.*, on the treatment of chronic renal disease.⁵ QALYs were developed in the early 1970s in an attempt to develop a measure of effectiveness incorporating effects on both quantity and quality of life. Some of the pioneering work in developing QALYs was carried out by Bush *et al.* in 1973⁶, in an analysis of a phenylketonuria screening program, and by Weinstein and Stason in 1976⁷ in their analysis of the cost-effectiveness of treatment for hypertension. The development of the methodology of economic evaluation is an ongoing process. The bearings of this development have recently been formulated in a set of recommendations by the Washington Panel, a group of leading researchers in the area of economic evaluation.¹

In the 90s, the application of the technique of economic evaluation has become very popular and the literature in this field has grown exponentially. The results of economic evaluations are now being used as the basis for major policy decisions about pricing and reimbursement of drugs, and, initially, for rationing in the Oregon experiment.⁸ A number of governments now require submission of the results of economic evaluation of new pharmaceuticals to inform decisions about drug

reimbursement. In 1992, the Australian government was the first to formally require evidence on cost-effectiveness for public reimbursement for drugs, in addition to proof of efficacy and safety. These requirements have also been introduced in Ontario, Canada.⁹ Recently, the Ministry of Health in the Netherlands has taken up the issue. The Ministry is now in the process of developing guidelines for reimbursement for new pharmaceuticals, in which the pharmaceutical industry has to show proof of cost-effectiveness before a new drug can be included in the basic health package.¹⁰ In the US, the Health Care Financing Administration is now considering the introduction of cost-effectiveness as a criterion for Medicare reimbursement.^{11,12} These requirements, and the likelihood that other countries will emulate Australia's and Canada's lead, has led to further attention on the area of economic evaluation.¹³

THE DESIGN OF ECONOMIC EVALUATION

Since the first applications of economic evaluation to health care, as noted, more refined methods are becoming available to measure the costs and benefits stemming from the use of specific health care programs. Despite these methodological advances, much work remains to be done before the results of economic evaluations can be used with great confidence in decisions about the allocation of health care resources. This thesis aims to be a further step in preparing economic evaluations for use by policy makers.

There is a wide variety of instruments for measuring the costs and benefits of interventions. Generally, the RCT is considered to be the 'gold standard' for the purpose of economic evaluation.¹ However, trial-based economic evaluations of competing therapies have manifold drawbacks and present many methodological difficulties. These were first identified in the early nineties by Drummond and Davies¹⁴, and by Rutten *et al.*¹⁵ They touched upon issues in design, in the collection of resource use data, the collection of outcome data, and in the interpretation of results from RCT-based economic evaluations. For example, such factors as inadequate sample size and the inadequate duration of trials were noted as limitations to adequate policy making. Somewhat later, issues on the generalisibility of results from RCT-based economic evaluation arose, illustrated in a study focussing on the representativeness of cost-effectiveness results as estimated in one country for decision making in other countries.¹⁶ In 1994, Leidl identified a number of factors that may cause cost-effectiveness ratios at the population level to deviate from what would be expected when looking at RCT-derived patient cost-effectiveness ratios. The factors identified include the presence of economies of scale and externalities.¹⁷

The question that emerges, after identifying these limitations of economic evaluation, is how these issues might be overcome in designing future economic evaluation

studies. How, for example, to interpret an economic analysis based on an RCT that excludes patient groups which may however be an integral part of any treatment population, such as patients with co-morbidities? And how to handle issues of protocol-driven resource patterns in a clinical trial? In other words, which approach to economic evaluation will render results that can be used for appropriate and effective policy making?

This methodological challenge has only recently received attention. A decade ago, the use of checklists has been proposed to test the internal validity of RCT-based economic analyses.¹⁸ Later on, these checklist have been extended to also address the policy makers' need to be informed on the value of economic evaluation studies for their specific decision making context. This has e.g. been applied for deciding on the use of two alternative treatments for acute myocardial infarction. The analysis was found to be greatly uncertain on the cost-side of the analysis in particular, due to the mutual influence of various interfering factors.^{19,20} Other researchers have suggested the application of sensitivity analysis. They note that by applying sensitivity analysis to those factors that seem to most affect the cost-effectiveness ratio, and by screening the range of parameters relevant to the specific decision context, policy makers can assess the specified impact.²¹

Some authors argue for a trend away from trial-based studies and towards the use of observational data or modeling studies.^{16,22} They note that observational data describe ordinary medical practice and may therefore encompass a high real world relevance. The economic evaluation of thrombolytic therapy is an example of a study in which observational data have been applied to an RCT. In this study, estimates drawn from observational databases were used instead of estimates of unit costs, as observed in the trial.²³

Others suggest that modeling might be applied as a substitute for or supplement to RCT-based economic analysis.^{12,24,25} Proponents of modeling argue that using it in economic evaluation is an 'unavoidable fact of life'²⁴ which will result 'in richer analyses with greater validity than a single trial can provide'.¹² Models applied as *substitutes* for RCT-based economic analysis derive their input data from a variety of sources with varying degrees of scientific rigour. Effectiveness data can be RCT-based, as applied e.g. in the cost-effectiveness study of the treatment of hypertension²⁶, but can also be derived from the less scientifically rigorous evidence of observational studies, as applied e.g. in the cost-effectiveness study of mammographic screening strategies²⁷, or can even be based on expert opinion. Modeling, as *additional* to RCT-based economic evaluation, has been advocated because it allows a simultaneous consideration of the effects of various factors that may all affect the cost-effectiveness of a medical intervention as estimated in an RCT-based economic analysis.²⁵

However, this trend away from RCT-based economic analysis is experiencing strong opposition, mainly from the field of clinical medicine. This perspective is typically characterised by the notion that, if an outcome is not directly measured, 'one can obtain any desired results, depending upon the model and the assumption used'. For example, in an editorial in the *New England Journal of Medicine*, great concern was expressed towards the use of (additional) modeling in economic evaluation: 'bias can compromise even original scientific studies, but we believe that the opportunities for introducing bias into economic studies are far greater given the discretionary nature of model building and data selection in these analyses'.²⁸ Opponents of modeling point to a much cited analysis of antidepressant use, which claimed to show that paroxetine was more cost-effective than imipramine, and which caused considerable debate.²⁹ Reworking by other analysts who revisited the key assumptions of the model led to opposite conclusions.³⁰

While economists would consider that such a hesitant view reflects a lack of appreciation of the role of models in economic analysis, it is clear that there is a major difference between the clinical and medical economics perspectives. This probably stems from the background of economic evaluation in health care. Although it originates in clinical medicine, which has a tradition of experimental research, it is also influenced by the field of economics, in which modeling has long been practised.

Basic to these diverging opinions is the worrying question of which instruments best inform policy makers on the costs and benefits of medical interventions. The choice of which instruments should be applied in economic evaluation is an issue of increasing importance.³¹ As the pharmaceutical industry, encouraged by powerful financial incentives, enters into the provision of economic information, concerns about bias have been raised. There is concern that assumptions and selection of data sources can be manipulated in order to influence reimbursement decisions.³²

This issue regarding the choice of instruments for economic evaluation is also reflected in the formulation of national guidelines for the registration of pharmaceuticals. In 1992, the Australian government released the first guidelines for the pharmaceutical industry on the preparation of submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). The guidelines established strict evidentiary standards and reporting structures. The PBAC recognised only economic evaluations that were carried out alongside randomised trials. In 1995, a revised set of guidelines was released, containing a number of interesting changes.³³ Although every submission still requires trial-based evidence, the revised guidelines recognise the often limited nature of the information from randomised trials, in terms of judging the full clinical and economic performance of a drug. Modelled economic evaluation is considered to be potentially useful to the PBAC. Modeling can be useful to transfer or generalise from trial-based evidence, for example to modify resource use patterns,

particularly for non-Australian based trials. Hereby they acknowledge that trial-based evidence is the most internally valid but may be inadequate for policy making.

OBJECTIVES

This thesis elaborates on the informative value of economic evaluation to policy makers and explores the methodological means by which an optimal informative value can be achieved. The informative value of economic evaluation is defined as the ability to adequately inform policy makers on the consequences of adopting a medical intervention. This thesis examines the characteristics of the various instruments of economic evaluation in terms of fulfilling this objective, and formulates options for improving their informative value. In other words, the thesis addresses the question of how economic evaluations can be prepared so that their results can be successfully used by policy makers. Underlying this aim is the need to link economic evaluations and policy making. The aim of the economic evaluation of health care programs is to help policy makers in making appropriate decisions. If the economic evaluation of health care programs were to have no impact on decisions about the allocation of resources for health care programs, carrying out such evaluations would be a meaningless activity. Thus, it is crucial to investigate further what evidence policy makers require to adequately define health care policy.

As a first step, this requires information on costs and benefits that is valid from both an internal and external point of view. Internal validity addresses whether an economic analysis truly determines which of the medical interventions under study would provide the most benefit for the available resources. Just as with other types of studies, the validity of economic analysis is primarily determined by the strengths of the methods used. External validity of results from economic analysis refers to whether the results can also be applied in other policy making contexts and whether they accurately reflect costs and benefits as occurring in daily practice. In this thesis, external validity is also labelled as 'real world' relevance. This concept implies that the results from economic evaluation must be adapted as closely as possible to the real world, where patients are actually treated. This real world has been described by Spilker as:

"the manner in which a particular medicine or treatment is usually used, given, or administered (which) includes an incredible diversity and complexity of factors that may interact with any medicine or treatment prescribed and influence a patient's clinical outcome".³⁴

The nature of the development of interventions implies that economic evaluation is limited in its capacity to produce both internally and externally valid information.³⁵ Figure 1 shows the typical diffusion path of an intervention.³⁶ Most economic evaluations are Phase 1 or Phase 2 studies, carried out prior to the licensing of an

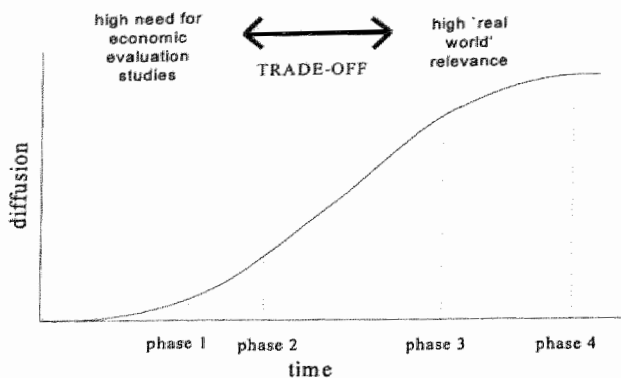


FIGURE 1. The timing of economic evaluation studies and the trade-off between the need for economic evaluation and the 'real world' relevance of results

intervention. External validity, however, is limited as observations are typically done in an artificial design that does not necessarily reflect daily practice. Preferably, economic evaluations should be carried out during Phase 4, when the economic profile of the intervention can be observed in the real world environment. However, with regard to the importance of licensing and reimbursement decisions, the need for economic evaluation is much more urgent during Phases 1 and 2. The figure shows the trade-off between the need for economic evaluation studies and the real world relevance of results of these studies. This situation in economic evaluation produces sub-optimal information to policy makers. This thesis considers how the informative value of (RCT-based) economic analysis can be enhanced so that it better reflects the real world environment of a policy maker. This includes an in-depth assessment of the context-specific system variables.

As a second step, a formulation of the aggregate impact, i.e. the total costs and effects at the population level, is required. This serves the need of e.g. health care insurers who may want information on total costs related to the adoption of an intervention before making decisions on coverage. Such analyses are in line with the traditional economic approach that dictates that the impact of a new (medical) intervention is assessed from a macro-perspective. This perspective takes account of the present pattern of use of medical interventions and outcomes (an initial equilibrium position) and models the impact of a new medical intervention in terms of substitution patterns and the resultant new equilibrium in the pattern of use of the medical intervention, given the characteristics of the population being treated. Thus, the net impact of the new medical intervention is examined in an equilibrium framework and compares the cost-outcomes profiles before and after the new medical intervention is introduced.

Depending on the level of detail needed, such system impact analyses may be simple extrapolations of results obtained at the individual level, but they may also involve more sophisticated approaches like public health modeling. These latter analyses comprise modules on the incidence of disease, on disease development, on patterns of costs of interventions, and on the morbidity and mortality effects of interventions. This thesis discusses the possibilities and limitations of transposing results from the individual (patient) level to the aggregate (population) level.

RESEARCH QUESTIONS

The central question to be addressed in this thesis is:

How can economic evaluation studies be prepared to optimally support policy making?

This central question is distinguished into four sub-questions. As these sub-questions are very much related to each other, they are not examined chapterwise but rather discussed throughout the whole thesis. The following sub-questions are distinguished:

- 1) *What are the requirements for economic evaluation to optimally support policy making?* This question is addressed in chapters 2, 7, and 8.
- 2) *What are the existing approaches to economic evaluation?* This question is addressed in chapters 2, 5, 6, and 8.
- 3) *Do these approaches fulfill the requirements for economic evaluation?* This question is addressed in chapters 2,3,4,5,7, and 8.
- 4) *What can be done to bridge the gap between the existing approaches and the requirements?* This question is addressed in chapters 2,3,4,5,7, and 8.

HOW TO READ THIS BOOK

This book consists of a number of studies that analyse the sub-questions as stated above. On the basis of these individual contributions, the thesis works towards a general framework to bridge the gap between clinical research and policy making. Although the book is best read from cover to cover, the time-constrained reader might find chapter 8 of most interest. In a general framework on the design of economic evaluation studies, this chapter presents an overview of the methodological achievements of this thesis. This chapter can also be read as an independent contribution to the further development of the methodology in economic evaluation. This thesis is further organised as follows:

Chapter 2 examines the requirements for economic evaluation to optimally support policy making, and tests the characteristics of RCT-based economic analysis against

these requirements. Particular focus is given to a critical assessment of the real world relevance or external validity of results from RCT-based economic analyses. A systematic overview is provided on the presence of potential confounding factors which must be considered when translating results from RCT-based economic analyses to the real world context of a policy maker. This results in a checklist: a set of recommendations to researchers on how to report RCT-based economic analysis, and to policy makers how to interpret its results.

Chapter 3 examines the consequences of disregarding the potential biases of economic evaluation in policy making. As an example, the impact of age is analysed. It is shown that disregarding age as an important variable in cost-effectiveness analysis may seriously bias allocation decisions regarding health care resources. One reason for this is that the age structure in the RCT may be much different from the age structure of the population in which the intervention will be applied. Another reason is that - because of demographic and epidemiological developments - the age structure of the population may alter over the period for which the allocation decision is relevant. Some suggestions are presented to control the influence of these biases on policy making.

As another example of a potential bias, in Chapter 4, the potential impact of ignoring interactions with other diseases within an economic evaluation is addressed. This is illustrated in a study on influenza-related hospital admissions in the context of the cost-effectiveness of influenza vaccination. The hypothesis is that some hospitalisations which are attributed to other diseases on hospital discharge certificates are in fact related to influenza. From an economic perspective, this implies that the cost-effectiveness of influenza vaccination may be severely underestimated when looking at the changes achieved in the primary diagnoses of the related diseases alone. Regression analysis is applied to control for this potential confounding factor.

Chapter 5 explores additional modeling as an approach to enhance the informative value of RCT-based economic analyses. The economic analysis of an intervention against benign prostatic hyperplasia (BPH) is taken as an example. By combining RCT-based data and observational data on resource use related to BPH treatment, costs as occurring in a real world environment are estimated. Furthermore, by performing analyses at the aggregate level, the system impact of introducing the intervention is assessed.

Chapter 6 reports on the informative value of economic analyses based on models. As an example, a decision-tree model on the cost-effectiveness of pneumococcal vaccination has been designed. The use of modeling is highly important for decisions in situations that cannot be grasped by an RCT-experiment. This chapter illustrates the relevance and usefulness of this methodology.

The reporting of costs, effects and cost-effectiveness ratios by researchers is critically assessed in Chapter 7. There is confusion on the decision-criteria as based on results from economic evaluation. Moreover, an incorrect reporting of results may easily lead to inadequate policy making. A number of suggestions are formulated to confront the confusion.

The methodological achievements of this thesis are presented in Chapter 8. The chapter addresses the research questions as proposed above. A general framework is proposed to increase the validity (both internal and external) of economic evaluation at the individual level, as well as to allow the transposition of results from the individual level to the aggregate level.

Chapter 9 comprises a conclusive overview of this thesis. The results of the individual studies are carefully examined within the context of the thesis, and the achievements of the thesis are assessed in context of the overall methodological development of economic evaluation in past decades. Furthermore, some areas for further research are identified.

REFERENCES

1. Gold M.R., Siegel J.E., Russell L.B., Weinstein M.C. (editors). *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996
2. Kaldor N. Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal* 1939; 49: 549-52
3. Hicks J.R. The foundation of welfare economics. *Economic Journal* 1939; 49: 696-712
4. Blumenschein K., Johannesson M. Economic evaluation in health care: a brief history and future directions. *Pharmacoeconomics* 1996; 10: 114-22
5. Klarman H.E., Francis J.O.S., Rosenthal G. Cost-effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care* 1968; 6: 48-54
6. Bush J.W., Chen M., Patrick D.L. Cost-effectiveness using a health status index: Analysis of the New York State PKU screening program. In: Berg R., editor. *Health status indexes*. Chicago: Hospital Research and Educational Trust, 1973
7. Weinstein M.C., Stason W.B. *Hypertension: a policy perspective*. Cambridge (Mass.): Harvard University Press, 1976
8. Tengs T.O. An evaluation of Oregon's Medicaid rationing algorithms. *Health Economics* 1996; 5: 171-81
9. Drummond M. Cost-effectiveness guidelines for reimbursement of pharmaceuticals: is economic evaluation ready for its enhanced status? *Health Economics* 1992; 1: 85-92
10. Innovator. Annual overview health care 1998 (in Dutch). 1997; 3: 8-9
11. Leaf A. Cost-effectiveness as a criterion for Medicare coverage. *New England Journal of Medicine* 1989; 321: 898-900
12. Luce B.R. Policy implications of modeling the cost-effectiveness of health care technologies. *Drug Inf J* 1995; 29: 1469-75

13. Maynard A. Economic evaluation techniques in health care: reinventing the wheel? *Pharmacoeconomics* 1997; 11: 115-8
14. Drummond M.F., Davies L. Economic analysis alongside clinical trials. *International Journal of Technology Assessment in Health Care* 1991; 7: 561-73
15. Rutten F.F.H., Bonsel G.J., Uyl-de Groot C.A. Economic evaluation and clinical trials- methodological and practical aspects. *Abstract Fifth International Conference on System Science in Health Care*. Omnipress Publishing, Prague 1992
16. Drummond M.F., Bloom B.S., Carrin G., Hillman A.L., Hutchings H.C., Knill-Jones R.P., Pouvoirville de P., Torfs K. Issues in the cross-national assessment of health technology. *International Journal of Technology Assessment in Health Care* 1992; 8: 671-82
17. Leidl R. Some factors to consider when using the results of economic evaluation studies at the population level. *International Journal of Technology Assessment* 1994; 10: 467-78
18. Drummond M.F., Stoddart G.L., Torrance G.W. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press, 1987
19. O'Brien B.J., Heyland D., Richardson S., Levine M., Drummond M.F. Users' Guide to the Medical Literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? *JAMA* 1997; 277: 1802-6
20. Drummond M.D., Richardson S., O'Brien B.J., Levine M., Heyland D. Users' Guide to the Medical Literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? *JAMA* 1997; 277: 1552-7
21. Overkamp A., Ament A.J.H.A. *Uncertainty on the optimal use of medical technologies: the role of sensitivity analysis* (in Dutch). Working Paper Department of Health Economics, Maastricht University WP 94.01, 1994
22. Sheldon T.A. Problems of using modeling in the economic evaluation of health care. *Health Economics* 1996; 5: 1-11
23. Kalish S.C., Gurwitz J.H., Krumholz H.M., Avorn J. Cost-effectiveness model of thrombolytic therapy for acute myocardial infarction. *J Gen Intern Med.* 1995; 10: 321-30
24. Buxton M.J., Drummond M.F., van Hout B.A., Prince R.L., Sheldon T.A., Szucs T., Vray M. Modeling in economic evaluation: an unavoidable fact of life. *Health Economics* 1997; 6: 217-27
25. Mason J. The generalisability of pharmacoeconomic studies. *Pharmacoeconomics* 1997; 11: 503-14
26. Kawachi I., Malcolm L.A. Treating mild to moderate hypertension: cost-effectiveness and policy implications. *Journal of Cardiovascular Pharmacology* 1990; 16 (Suppl.7): S126-8
27. Lindfors K.K., Rosenquist C.J. The cost-effectiveness of mammography screening Strategies. *JAMA* 1995; 274: 881-4
28. Kassirer J.P., Angell M. The journal's policy on cost-effectiveness analyses (editorial). *New England Journal of Medicine* 1994; 331: 669-70
29. Jönsson B., Bebbington P. What price depression? The cost of depression and the cost-effectiveness of pharmacological treatment. *British Journal of Psychiatry* 1995; 166: 398-99

30. Woods S.W., Rizzo J.A. Cost-effectiveness of antidepressant treatment reassessed. *British Journal of Psychiatry* 1997; 170: 257-63
31. OHE Briefing. Towse A., Drummond M.F. (editors). Luce B.R. in: *The pros and cons of modeling in economic evaluation*. London: Office of Health Economics, no 33, 1997
32. FDA Division of Drug Marketing, Advertising and Promotion. *Principles for the review of pharmaceutical promotion*. Washington DC: FDA, 1995
33. Langley P.C. The November 1995 revised Australian guidelines for the economic evaluation of pharmaceuticals. *Pharmacoeconomics* 1996; 9: 341-52
34. Spilker, B. *Guide to Clinical Trial.*, New York: Raven Press, 1991
35. Bloom B.S., Fendrick A.M. Timing and timeliness in medical care evaluation. *Pharmacoeconomics* 1996; 9: 183-7
36. Banta D., Luce B.R. *Health care technology and its assessment: an international perspective*. Oxford; New York: Oxford University Press, 1993

CHAPTER 2

POTENTIAL BIASES OF ECONOMIC EVALUATION FOR POLICY MAKING

Published as: Baltussen R.M.P.M.¹, Ament A.J.H.A.¹, Leidl R.M.^{1,2} Making cost assessments based on RCTs more useful to decision makers. *Health Policy* 1996; 37: 163-83

¹ From the Department of Health, Organisation, Policy and Economics, Maastricht University

² From the Department of Health Economics, University of Ulm, Germany

SUMMARY

The appropriateness of results from economic evaluation for allocation-decisions in health care is a point of major concern for decision makers. This chapter concentrates on the real world relevance of results from economic evaluation as an additional step towards making results more useful to decision makers. Three classes of biases are examined. The first relates to the limited scope that economic analysts sometimes choose in RCTs. The second class involves the methodological aspects of RCTs and questions the real world relevance of the tools with which economic analysts estimate costs on the basis of RCTs. The third class concerns the representativeness of RCT results, i.e. the generalizability of these results and their usefulness in other treatment contexts. The chapter discusses options for limiting the potential confounding influences of these biases and proposes a checklist which should be applied by policy makers when using results from economic analyses in their own decision context and by economic evaluators when constructing and describing economic analyses. This will enhance the relevance of the results of economic evaluation in decision making and improve the information basis for actual allocation decisions in health care.

INTRODUCTION

In the clinical evaluation of medical interventions, there is no doubt that today the 'gold standard' design for determining the efficacy of treatment is the randomized clinical trial (RCT). Reliance on this system is beginning to be countered by a growing body of outcome research which is based on observational studies and non-randomized designs.¹ At the core of these diverging perspectives on analysis — experimental or observational — lies the worrying question of whether the results of experimental assessments will provide us with sufficient information on actual treatment effectiveness. RCTs are also considered an appropriate basis for the economic evaluation of interventions. Because this design is increasingly used for economic evaluations, it is important to consider whether the results of such studies will provide us with sufficient information for our economic decisions in health care.

There are several impediments to the unrestricted use of economic evaluation results that have emerged from RCT-based studies. For a given decision context, the appropriateness of the trial design has to be checked from an economic perspective, and the representativeness of the trial for the actual treatment of patients has to be considered. Decision makers have to look at the economics of actual treatment in a given decision context, not the circumstances of the original trial: decision making on the allocation of health care budgets requires a comparison of the most realistic cost-information of various medical interventions rather than their costs as incurred under very specific conditions. Because the RCT is the accepted standard for experimental design and because of the need to use information appropriate for a decision context, results from RCTs will continue to be used, but they must be adapted as closely as possible to the 'real world', where patients are actually treated. This real world has been described by Spilker as:

'the manner in which a particular medicine or treatment is usually used, given, or administered. Although this term is used as if it represents a single entity, this is obviously not true. Except in rare circumstances, the real world includes an incredible diversity and complexity of factors that may interact with any medicine or treatment prescribed and influence a patient's clinical outcome'.²

In this definition, Spilker refers primarily to the outcome of treatment, that is, the effect side in economic evaluation. This paper focuses on the cost side of economic evaluation. A wide variety of factors may cause the costs as estimated through the RCT to deviate from the costs in the real world. A good deal of research has been done to date in this area. Drummond and Davies³, Petrou *et al.*⁴, Luce⁵ have discussed issues such as the choice of discount rate, perspective, inadequate sample size, inadequate measurement and valuation of cost data, and neglect of relevant costs. This article adds to this existing knowledge and concentrates on the real world relevance of results from economic evaluation.

The analysis carried out in this article may be considered as a checklist for decision makers who are extrapolating cost estimations from RCTs to their own setting. Moreover, it may be beneficial to those constructing an RCT by identifying shortcomings which can then be better anticipated. Hence, this analysis can be regarded as a further step in the communicative process between researchers and decision makers to make results from economic evaluation more useful for decision making. All of the factors discussed in this article may have a confounding influence on the estimation of costs in connection with RCTs, as opposed to the costs actually accruing in the real world. Some factors are relevant at all stages of the diffusion of medical technologies, others only at later stages.

The factors, or potential biases, can be defined in three classes. The first relates to the limited scope that economic analysts sometimes choose in RCTs. Economic analysts may adopt a reductionist stance which does not fully account for factors like adverse effects and interactions with other diseases, and this may bias the cost estimation involved. The second class of biases involves the methodological aspects of RCTs and questions the real world relevance of the tools with which economic analysts estimate costs in connection with RCTs. The third class of biases relates to the representativeness of the RCT, i.e. the generalizability of its results and its transferability and usefulness in treatment contexts other than the RCT setting. Even when the scope and the methodology of economic evaluation do meet the standards so as to produce unbiased cost figures, decision making in another treatment context than that in which the RCT took place may be biased.

Costs can be defined as the resource consumption which is associated with an intervention. It should include all resource consumption, comprising not only costs directly associated with providing the technology, but also the cost flows that are saved or spent as a result of the course of the disease, changes in treatment, and the use of the technology. However, there is much confusion as to which consequences to consider as effects, and which as costs.⁶ In this article, 'costs' are considered to be those economic consequences which typically appear in the numerator of the cost-effectiveness ratio.

It should be noted that this article does not aim to question the design of RCTs, but to discuss issues which could cause decision makers to derive the wrong implications from trial studies. Moreover, while the analysis focuses on the RCT as the vehicle which is regarded as the most appropriate for economic evaluation, several of the issues raised are relevant to other kinds of economic evaluation too.

The paper begins with a systematic outline of the factors belonging to these three classes of bias. In the following section, options to limit the potential confounding influence of these biases are discussed and a check-list is presented. This check-list adds to existing check-lists and is developed for decision makers to test the relevance

of results from economic evaluation for their decision making context. Finally, conclusions are drawn.

SCOPE OF RCTs

RCTs are typically limited in scope in three respects, any one of which may seriously flaw the adequate estimation of actual treatment costs. First, evaluators limit the breadth of analysis of RCTs to the disease under study; this means that interactions with other diseases may be overlooked. Second, the impact of external effects may be overlooked. Third, RCTs may not be designed to cover a period long enough to adequately measure the full economic impact of medical technologies.

Interactions with other diseases^a

Interactions with other diseases may bias cost estimations derived from RCTs. Evaluators typically estimate the direct effect of an intervention on a single disease, while assuming that the intervention will not change mortality or morbidity from other causes. This assumption is rarely tested because RCTs usually lack the statistical power and length of observation period to detect associations between interventions, risk factors and apparently unrelated morbidity and causes of death.⁷

Two types of interactions with other diseases can be distinguished. The first relates to diseases with 'concurrent' or 'dependent' determinants. In the case of two diseases with a concurrent determinant, an intervention which focuses on this determinant in order to fight one disease will change the risk of acquiring the other disease. This risk may be reduced, if the determinant was a risk factor for the other disease. For example, assuming smoking is a risk factor for both coronary heart diseases (CHD) and lung cancer, a reduction in smoking will reduce the incidence of both CHD and lung cancer.^{8,9} On the other hand, the risk of suffering another disease may be increased, if a determinant is a risk factor for one disease (which is the subject of an intervention), but at the same time is a protective factor with respect to another disease. It has been suggested, for example, that smoking has a protective function with respect to dementia. The positive effects of smoking cessation on CHD and lung cancer would then be offset by an increase in dementia.^{9,10} Disregarding the indirect impact on other causes of death will lead evaluators to miscalculate the survival gain from an intervention. In the estimation of costs of an intervention aimed at one specific disease, ignoring any costs or savings stemming from interactions with other diseases will lead to biased estimates of the real world costs.⁷

The second type of interaction with other diseases is in the case of diseases with 'independent' determinants. Assume that the risk of dying of a specific disease will be reduced by intervening on its determinants. However this has no effect on the risks of acquiring diseases that were not correlated with the determinants of the initial

disease. The likelihood that the treated individual will develop any of these substitute diseases is the same as for the rest of the population. There is no consensus among economic researchers whether the extra health care costs related to these diseases should be included in the analyses. In the Canadian guidelines for economic evaluation it is recommended to include these future costs¹¹; in the Australian guidelines the opposite is declared.¹² However, because of the process of discounting, the present value of future costs and effects is considerably reduced and is considered as safe to be ignored in the analysis.¹³ This can be shown for the economic evaluation of tuberculin screening and treatment in which two separate analyses were carried out. An 'augmented' analysis, which included the costs of medical care incurred during extended survival and a 'direct' analysis which did not include these costs. In the augmented analysis the extra health care costs were found to amount to one-third of the total costs of screening and treatment of tuberculin. However, when discounted, the influence of this factor on the cost-effectiveness ratio was reduced to an almost negligible level, since the extra health care costs would be incurred only in the distant future.¹⁴

External effects

External effects may produce costs or savings that fall on non-patients and which are typically not considered in RCTs. For example, external effects may be involved in the case of early intervention in human immuno deficiency virus (HIV) infection.¹⁵ At the individual level early intervention may be beneficial but at the population level this may, given certain conditions, also provoke negative effects; the intervention may cause an increase in the spread of HIV in the population because people at an infectious stage are kept alive longer and may consequently infect more people. Decision making in the presence of external effects is bound to be sub-optimal if based solely on information provided by RCTs.

Inadequate duration

The definition of the endpoint of the RCT and hence the duration chosen for the RCT may vary for clinical research and economic evaluation. Clinical research tends to define endpoints as the state of affairs at a fixed follow-up time, such as the *n*-year survival rate difference. Economic evaluation requires an endpoint which represents the cumulative life-time effect. For example, economic evaluations of medical interventions aiming at chronic diseases should cover a period long enough to ensure that the full economic impact of the intervention has been measured. In the case of economic evaluations of medical interventions directed at acute diseases, the estimations of costs and quality of life also require a longer time interval than the measurement of clinical parameters only.¹⁶ Decision makers should be aware of any unreliability caused by this factor. For example, Petrou *et al.*⁴ have calculated cost-per-QALY estimates after sensitivity analyses were performed on the assumed survival periods accruing from seven interventions. For two procedures, the analysis has little effect on the cost-per-QALY estimates since the recurring annual costs of

treatment and the annual QALYs gained per patient tend to cancel each other out. For the remaining five procedures, wide variations in cost-per-QALY ratios emerged, largely explained by variations in the potential survival periods. This is the result of the one-off cost for surgery reducing the average annual cost per patient as the survival period increases.

METHODOLOGY OF RCTs

There are a number of features of the methodology of economic evaluation based on RCTs, which may potentially bias the estimation of costs. The underlying cause is the difference between the objectives of RCTs and economic evaluations. The main purpose of RCTs is to assess the efficacy of a therapy rather than to demonstrate its efficiency; economic evaluation is often viewed as ancillary work. Moreover, economic analysts are frequently involved only at a latter stage, when most decisions with respect to the methodology have already been made.

A large number of these methodological shortcomings of economic evaluation have been extensively reported elsewhere.^{3,4,5} However, two factors which may typically limit the real world relevance of results from economic evaluation and which have not attracted much attention in the context of economic evaluation are artificial design of RCTs and patient non-compliance.

Artificial design

RCTs often have an artificial design, which means that cost figures obtained from an RCT may very well deviate from the costs that would occur in the real world. For example, more safety and efficacy tests are usually performed in an RCT than in normal medical practice.² More attention is also paid to consistently following-up every complaint or potential adverse event, with increased patient monitoring and consultations. This may cause cost estimates based on RCTs to be higher than might be the case in medical practice.⁶ One must also consider the *Hawthorne effect*, which is the tendency of people to change their behaviour because they are the target of special interest and attention in a study: patients may improve because of this attention and not because of the treatment itself.¹⁷ The presence of the *Hawthorne effect* may cause the effects, and hence cost-effectiveness, of an intervention to be more favourable in the RCT than in the real world. On the other hand, certain clinical problems may be detected more frequently in RCTs than would be the case in the real world, they may be detected earlier, or may even be prevented from occurring. Depending on the problem being considered, this may increase or decrease costs in the long run.

Patient non-compliance

RCTs rarely show the effects of poor compliance and so reveal only clinical cost-effectiveness. For example, it has been argued that both the treatment benefit and cost vary with compliance in the case of a drug therapy.¹⁸ The consequence of inadequate drug usage is likely to be clinical deterioration of the factors the drug was prescribed to prevent or manage, e.g., epileptic seizures for anti-epileptic drugs, hypertension for anti-hypertensive drugs etc. The consequences may take the form of an acute problem resulting in hospitalization costs, or of a more chronic form requiring costs for prescription changes, diagnostic tests, consultations, and more frequent visits. In an estimation of non-compliance and its economic consequences, based on a group of ostensibly cooperative patients, it was calculated that savings of approximately one-third of the total costs of drug treatment can result from rather conservative improvements in drug usage.¹⁸ Compliance with therapy can be expected to be higher in RCTs, since patients are closely supervised. If non-compliance is not considered when results are extrapolated to the real world, costs may turn out to be higher than expected.

REPRESENTATIVENESS OF RCTs

The third class of bias concerns the representativeness of RCTs. The costs of treatment may well vary depending on which patients are treated, which physicians treat them and where the treatment takes place. Consequently, decision makers who wish to extrapolate RCT results to their own context should bear in mind that the patient group which was studied, the treating physicians and the setting of the RCT may not be representative, causing costs to differ.

Patient case-mix

The patient case-mix refers to the mixture of patients with varying capacities to benefit. RCTs tend to include cases which are most likely to lead to effect detection. However, this case-mix may not be representative of the situation in the real world, which can have a large impact on the costs involved. The case-mix in an RCT may vary from the medical practice context because of differences in the patients' severity of illness, because of co-interventions or because of differences in patient characteristics.

Severity of illness. There are two dimensions to the severity of illness which may impact on the costs of interventions. First, across settings, the introduction of a procedure in a hospital may attract patients with a different severity of illness than the patients involved in the RCT. There is evidence that patient resource use, for patients within individual diagnostic related groups (DRGs), varies significantly depending on the patient's severity of illness.¹⁹ In this case, the severity of illness was calculated from patients' laboratory, radiology, and diagnostic test results, vital

signs, medical history and physical exam and secondary diagnosis (co-morbidity). It has also been shown that the costs of heart transplants depend on the level of risk or severity of illness of the patient involved; for a low-risk male patient the calculated cost of a heart transplant is \$81,297, whereas for a high-risk male patient these are \$103,897.²⁰ For acute myocardial infarction (AMI) patients, it has also been demonstrated that the variance of costs can be largely explained by the severity of illness on admission.²¹

Second, the indications for patients eligible for treatment may expand over time. Once an intervention is adopted, the mix of patients receiving treatment may change over time, and it may be that the first patients to be treated have a higher capacity to benefit than the patients treated later.²² This can also be regarded as technological change, in the sense that physicians increasingly master the technology to treat 'more difficult' patients. For example, in a study on the increase in physician inpatient spending from 1985 to 1988, a slight worsening of case severity has been found to be an important contributor, both for medical and surgical procedures.²³ In a study on the indications for coronary artery bypass graft (CABG) surgery, it has been shown that extensions of these indications lead to a patient case-mix including patients whose condition is worse, and hence to worse results. The costs may be expected to vary accordingly.²⁴ Failure to recognize these potential cost differences could bias the extrapolation of results from RCTs into the real world.

Co-interventions. As noted, RCTs tend to include cases which are most appropriate for effect detection. Since treatments other than the intervention under scrutiny might confuse the estimation of the costs and effects of that intervention, patients undergoing these so called co-interventions will typically be excluded from the analysis. However, this constitutes a bias as patients in the real world do undergo other treatments.²⁵

Patient characteristics. Differences in patient characteristics may also give rise to variations in costs. For the treatment of AMI patients, it has been shown that specific background characteristics of patients may be responsible for differences in medical care costs, given the severity of illness.²¹ Patients may, e.g., differ in their pre-hospital medical history or in individual characteristics such as age²⁶ or gender.

Physician

Different physicians may treat similar patients differently, which may lead to a variance in costs. Variances in physician characteristics and physician learning effects may cause the costs estimated from the RCT to deviate from costs observed in the real world.

Physician characteristics. In a study of the impact of medical staff composition on costs it was found that higher percentages of specialists and board-certified physicians

increase costs significantly. Using physicians employed at a medical school was found to be significantly related to lower costs.²⁷ In another study, it was observed that a higher percentage of staff physicians under age 45 reduced hospital costs.²⁸

Physician learning effects. In a number of industries it is well known that the average cost per unit produced tends to decrease with increased production volume²⁹; there is good evidence that this volume-cost relationship is due, in part, to a learning curve whereby efficiency of production increases with experience.³⁰ Physician learning effects are increases in skill levels which take place when more operations are performed: the effect has been termed 'practice makes perfect'³¹ or 'skill maintenance'.³² In the case of heart transplants, the specific experience of individual transplant surgeons has been found to have no significant effect on costs.²⁰ On the other hand, there is some evidence that physicians with more years of experience turn out to be significantly more costly, with cost differences much larger than expected based on salary differentials.²⁷ Decision makers must consider that when clinical approach of the physician involved in the RCT and the physician in the real world differ, costs may differ accordingly.

Setting

The difference of setting between RCTs and the real world represents another potential bias. This can be seen most clearly in the different costs for the same intervention carried out in a developed country, and in a less developed country. Costs may deviate not only across countries, but also across hospitals or other levels of organisation. Interventions which are introduced in hospitals other than that in which the RCT was conducted may generate costs that are different from those estimated from the RCT, because of factors such as hospital characteristics, learning effects and economies of scale.

Hospital characteristics. There has recently been ample research on the relationship between hospital characteristics and costs. It has been observed that hospitals which spend higher percentages of time in teaching, administrative, or research activities, relative to patient care activities, generate higher costs. For example, the cost of discharging a patient from a teaching hospital (academic hospital) is \$338 higher than in a non-teaching hospital.²⁷ This implies that an RCT of an intervention conducted at a teaching hospital - which is certainly not an irrelevant case - may yield cost estimates that are relatively high compared to the costs which will be incurred at non-teaching hospitals. The extrapolation of results from such an RCT to non-teaching hospitals may therefore be seriously biased.

Hospital learning effects. Hospital learning effects in heart transplants are defined as:

'the product of the collective experience of the organization as a whole, not merely the experience of individual surgeons, and as the efficiencies (which) are learned at a system level as the entire transplant team, including the surgeon, gains experience in the successful coordination of activities and resources'.²⁰

A study on the costs of heart transplants disclosed that the reduction in cost over time can be explained to a substantial degree by hospital learning effects. Using a reference case, the study predicts a cost of \$81,297 for the first heart transplant procedure performed at a hospital. For the fifth, the predicted costs would be only \$48,431. For the twenty-fifth case, the predicted cost would be \$35,352, and for the fiftieth case the cost would be \$25,458 (31 percent of the original cost).²⁰ The implication is that, over time, costs in the real world may decrease, so that an intervention which seems to be cost-ineffective based on the figures from an RCT may very well become cost-effective.

Capacity utilization. Capacity utilization observed during the RCT and capacity utilization in the real world may differ considerably.³³ If there is a higher utilization of medical equipment in the real world than in the RCT, and the economic analyst does not take this into account, the actual costs in the real world will be lower than the costs estimated from the RCT. One study of the costs and effects of laparoscopic cholecystectomy recognized this issue and provided information on costs for different levels of utilization. The critical number of cholecystectomies required before it becomes cost-efficient to invest in electro cautery laparoscopic equipment has been estimated to range from 60 to 140.³⁴

A related problem arises when an intervention pushes production beyond capacity thresholds. In such cases, additional equipment has to be purchased, causing costs - if the additional equipment is not fully utilized - in the real world to be much higher than the costs estimated through the RCT. If full capacity utilization is achieved, however, this problem vanishes.

Economies of scale. Economies of scale result where greater capacity leads to lower average costs in the long run, at a given level of capacity utilization.³³ This may cause costs in the real world to deviate from costs estimated from the RCT, with the direction of the difference depending on the difference between the scale of the setting of the RCT and the scale of the setting of the real world. In a study on the costs of U.S. teaching hospitals, the average incremental costs of the outputs were estimated to be generally higher than their marginal costs at sample means, indicating the presence of some product-specific economies of scale. Strong economies of scale were found for the numbers of residents, discharges, and inpatient days, and near-constant returns to scale for outpatient visits.²⁷

Joint production. Joint production is said to occur if the function of hospital costs is sub-additive, that is, if the sum of the cost of producing outputs separately is greater than the cost of producing the sum of these outputs jointly.²⁷ This implies that one should know whether there were joint production effects in the RCT if the results are to be extrapolated to other settings. If there is joint production in the RCT and not in the real world, costs in the real world may be higher. Conversely, if there is no joint

production in the RCT, but it does occur in the real world, real world costs may be lower. Positive product-specific joint production has been found, for example, for residency training and outpatient visit outputs.²⁷ That is, the resource costs of producing these services jointly was found to be less than the sum of the individual costs of producing each service separately.

Health service / country characteristics. The substantial variations in the organisation of health services across different local areas may also limit the generalization of cost estimations from RCTs. Context-specific factors such as the availability of outpatient or community services or the hospital reimbursement mechanism may very well influence the length of inpatient stay and consequently the costs. For example, the success of a method to reduce hospitalization for people with AIDS might depend on the availability of community care and the extent to which informal care networks have been developed.³

Decision makers who wish to extrapolate cost estimates from RCTs carried out in another country to their own setting should be extremely careful. Several factors have been reported which vary from country to country and which may affect the costs of an intervention to a considerable extent. These include differences in general and relative price levels for health care resources, and in community attitudes, clinical practice and financial systems. For example, the costs of ulcer patients receiving surgery in the United States were calculated to be 6 times higher than those in the United Kingdom.³⁵

A VIEW UPON THE IMPACT OF BIASES

As shown above, cost assessment in decision making may depend on a number of factors that can deviate between the RCT and the decision context. These factors may affect costs in various directions and ways, and they may do so specific to the medical intervention under study: a factor may constitute a major bias in an economic evaluation of one intervention, while being unimportant in another. Therefore, the full range of potential factors has to be specifically considered when using RCT evaluation results in concrete decision making.

In order to illustrate the potential impact of the various factors on cost assessment, a tentative example is discussed in the following. It is based on preliminary but plausible hypotheses which must be specified in the concrete utilization of RCT evaluation results in actual decision making. Such hypotheses should also be formulated by decision makers to assess which potential biases are most relevant to economic evaluations of a particular intervention. In the example, two types of interventions, such as surgical procedures and drug treatments, are considered (Table 1). The strength of the potential impact is reported in a range from '0' expressing no

impact to '***' denoting large impact. Only those factors which are hypothesized to play a role for one class of medical interventions, but not for the other class, are discussed below.

Most biases stemming from the (limited) scope of RCTs are equally pertinent to surgical and drug interventions. However, the issue of external effects seems irrelevant to surgical procedures. External effects typically occur in the case of infectious diseases which normally involve drug treatments. The issue of patient compliance clearly appears to be more important for drug treatments than surgical procedures, as the latter requires relatively little active participation from the patient. Regarding the methodology of RCTs, it can be stated that the potential biases arising from selective patients' withdrawal are likely to be more important for drug interventions than surgical procedures, as the latter offer fewer opportunities for withdrawal.

With respect to the representativeness of results from RCTs, physician characteristics and physicians' learning-effects are most important for complicated, risky and difficult procedures.^{20,36,37} Moreover, hospital characteristics or hospital learning effects are most relevant to procedures which are dependent on relatively complicated equipment and/or intensive cooperation of the operating theatre staff.^{36,37} This means that physician/hospital characteristics and learning effects may be a more important source of bias for surgical procedures than drug interventions. The issue of capacity utilization seems more applicable to surgical procedures than to drug treatments, because the former is more likely to involve expensive equipment and high fixed costs. Economies of scale are most likely to be found for interventions with high overhead costs, e.g. costs of hospital administration. This seems to be more applicable to surgical procedures than to drug treatments, as the former are relatively more imbedded in institutional care.

The factors associated with the representativeness of economic evaluations seem to be more pertinent to drug interventions compared to surgical interventions. Factors related to the limited scope and the methodology of an RCT seem to be equally pertinent to both types of medical interventions. For most categories, the direction of the potential bias is ambiguous and has to be specified for the particular intervention under study. However, for some factors the general direction of the bias can be stated. The issue of non-compliance, which seems to be more of a problem in the 'real world' than in the RCT, will cause the costs in the 'real world' to be higher than estimated from the RCT. Also, the 'patient case-mix' probably causes RCTs to underestimate costs; RCTs tend to include patients who are most suitable for effect detection and who therefore generate relatively low costs. It is important, however, to remember that these factors are relevant only if their impact on the cost estimates differs between the experimental programme and the comparison programme.

TABLE 1 Potential impact of potential biases in the cost estimation of two classes of medical interventions^a

Class of bias	Source of bias	Potential impact of bias	
		Surgical intervention	Drug intervention
Scope of RCT	interactions with other diseases	***	***
	external effects	0	***b
	inadequate duration	**	**
	patient non-compliance	*	***
Methodology of RCT	artificial design	**	**
	patient non-compliance	*	***
Representativeness of RCT	patient case-mix	***	***
	physician characteristics	**	*
	physician learning effects	**	*
	hospital characteristics	**	*
	hospital learning effects	**	*
	capacity utilization	**	*
	economies of scale	**	*
	joint production	*	*
	health service / country characteristics	***	***

a) The impact of the various factors ranges from '0' expressing no impact to '***' denoting large impact.

b) This is only applicable in the case of infectious diseases.

For example, if patient non-compliance causes costs in the real world to be higher than estimated at the RCT, but by the same factor for the experimental and comparison programme, the observed cost difference between the RCT results and the real world will not be important to decision makers. The cost-effectiveness of the experimental programme remains the same. Therefore, it is only necessary to control for those factors which are expected to respond differently in the economic evaluations of the interventions studied in the RCT. For example, there will be less need for intensive control in an economic evaluation involving two almost identical drug interventions than in an evaluation of a surgical and a drug intervention.

ACCOUNTING FOR POTENTIAL BIASES

Accounting for potential cost biases is not solely the responsibility of decision makers. There are a number of things which economic evaluators and decision makers can do to increase the relevance of results of economic evaluation based on RCTs for decision making.

These issues, which stem directly from the three classes of biases as identified above, are presented in Table 2. The table can be considered as a checklist which decision makers can use to assess the relevance of economic evaluations for their own decision context and which economic evaluators may review and use in their documentation. Many of these matters are related to the collection and use of additional data which will not always be available. However, these issues being explicitly defined, researchers will become more aware of their importance and can - especially in early stages of the design of the economic evaluation - make more efforts to produce this additional data. The checklist adds to standard checklists for economic evaluation studies like has been proposed by Drummond, Stoddart and Torrance.¹³ Our list also has a different focus: whereas the Drummond, Stoddart and Torrance check-list concentrates on examining the appropriateness of the methodology applied in economic evaluations, our checklist concentrates on the real world relevance of results from economic evaluations as an additional step towards making results useful to decision makers.

1. Is epidemiological information reported?

To allow decision makers to adjust the cost analysis for any differences in capacity utilization between the RCT and their own decision contexts, information is needed regarding the number of patients who are medically eligible for intervention. Researchers should provide incidence and prevalence figures regarding the disease the medical technology is focussed on.

Furthermore, decision makers should be able to account for risk reductions or increases due to interactions with diseases with dependent determinants. Although little is known about these kinds of interactions, decision makers should at least be informed about any indications of possible impacts and epidemiological information regarding these factors should be indicated. Furthermore, researchers must provide additional information with regard to the presence of external effects, which requires advanced modeling in the case of communicable diseases. Indications about epidemiology and linked costs are available for some diseases, such as HIV-related diseases.³⁸ Of course, epidemiological information regarding interactions with other diseases and external effects needs to be reported only if relevant.

2. Are context-specific factors explicitly reported?

It will never be possible to fully transfer the results of an economic evaluation from one setting to another. However, to allow decision makers to assess, and perhaps improve, the representativeness of the results, economic evaluators should at least outline a number of issues.

First, the various procedures undertaken in the RCT should be reported explicitly. With this information, decision makers can compare the use of resources in their

particular decision context with that of the RCT and can account for potential biases that may arise from any artificiality in the design of the RCT.

TABLE 2. Check-list cost-assessments based on RCTs

1. Is epidemiological information reported?
 - * Are the patient streams (as determined by incidence and prevalence figures) indicated?
 - * Are indications about interactions with diseases (risk reduction or risk enhancements) indicated?
 - * Is additional information with respect to the presence of externalities provided?
 2. Are context-specific factors explicitly reported?
 - * Are the various procedures undertaken in the RCT explicitly reported?
 - * Are volume and price components reported separately?
 - * Are fixed and variable costs related to the intervention specified?
 - * Is the extent to which the case-mix of the patients participating in the RCT is representative of the patient case-mix in the real world indicated?
 - * Are the specific physician and hospital characteristics (such as experience) that seem most likely to most influence the cost estimation reported?
 - * Are the characteristics of the health care system in which the RCT is carried out and that seem most likely to most influence the cost estimation reported?
 - * Are any learning effects indicated?
 3. Is there an indication of the impact of non-compliance?
 4. Is real world sensitivity analysis applied?
 - * Is sensitivity analysis carried out regarding those factors which are most likely to cause deviations between RCT results and the real world?
 5. Are future developments indicated?
 - * Are future patient flows (as determined by the incidence and prevalence of the disease) indicated?
 - * Is there an assessment of what kinds of patient groups will receive treatment in the future, as the intervention is diffused?
 - * Is there an indication of possible future price changes for the equipment?
-

Second, volume and price components should be specifically reported, to enable decision makers to adjust the economic analyses for any price differences that may occur between countries or over time. Third, fixed and variable costs related to the intervention should also be specified. This will allow the decision maker to adjust the results of the RCTs for any differences in capacity utilization and economies of scale in the real world context. Fourth, there should be an indication of the extent to which the case-mix of the patients participating in the RCT is representative of the patient case-mix in the real world. Any changes in cost-effectiveness which are expected as the medical technology is applied to other groups of patients should be reported. Fifth, specific physician, hospital and health care system characteristics which seem to most influence cost estimates, should be reported. Sixth, any learning effects should be indicated; learning effects can be identified by checking data for a time trend while controlling for confounding factors.

3. Is an indication of the impact of non-compliance given?

Although difficult to quantify, indications of expected levels of non-compliance and their impact on costs should be provided by researchers. To accomplish this, first, levels of non-compliance must be predicted. Non-compliance parameters have been established regarding some interventions, but obviously these may be different in other therapeutic areas.³⁹ Second, the possible consequences of inadequate drug usage (e.g. side-effects) need to be predicted. The effects of non-compliance are, of course, specific to the intervention under study and can often only be guessed at. However, observational studies have been carried out for some interventions, and consequences predicted with some degree of certainty.⁴⁰ Third, the costs of the various consequences need to be assessed. Depending on the severity of the consequences, this may involve costs of hospitalization, extra diagnostic tests or more frequent office visits. Overall, it is clear that estimating costs of non-compliance is a highly speculative matter.¹⁸ Nevertheless, these crude indications may very well improve the usefulness of RCT results for the decision maker.

4. Is real world sensitivity analysis applied?

The various components of an economic evaluation are bound to be surrounded by uncertainty. To reduce this uncertainty and to increase the usefulness of the results for decision making in the real world, researchers should carry out real world sensitivity analysis. Subject to analyses should be those factors that are most likely to cause divergences between RCT results and the real world situation. For example, sensitivity analyses should be applied if it is suspected that the duration of RCTs tends to be too short to provide certainty about the full economic consequences of medical technologies. This involves a test of the assumptions regarding the survival period of the individuals treated and the durability of the equipment involved. Furthermore, sensitivity analysis may be applied to factors like capacity utilization or discount rate.

5. Are future developments indicated?

The rapid developments taking place in medical technologies demand economic evaluations which consider dynamic processes. Decision makers who are implementing medical technologies need to consider not only current cost-effectiveness but also the impact of future developments on cost-effectiveness. Therefore a number of time-related processes should be indicated.

First, future patient streams (incidence and prevalence) regarding the medical technology under study should be forecasted and reported. This makes it possible to analyse the differences regarding capacity utilization between the RCT and the decision making context. Second, and closely related, there should be an assessment of what kinds of patient groups may receive treatment in the future, when the medical technology under study diffuses, i.e. other potential indications for the medical intervention should be identified. These patient groups may be less suitable for

treatment - thereby involving less benefits (and higher costs) - which would cause cost-effectiveness to decrease.

Third, researchers should indicate any major changes which can be anticipated in the price of the equipment involved. This is often rather hard to predict, but in other cases e.g. when economies of scale are involved (when the medical technology is introduced on a large scale) it may be easier to anticipate.

The information required is specific to the medical intervention under scrutiny; only those elements which are relevant should be included. It may be beyond the scope of articles in scientific magazines to report all of this information. Therefore the information should be included in 'laboratory books' which provide background information on the results presented in articles and which can be employed by decision makers.³⁹

However, even when the results of an economic evaluation are provided in such a transparent form, decision makers may need additional analyses to account for other potential confounding aspects. For example, when relevant costs are unknown, or the mode of comparison or the perspective is inadequate, additional data needs to be collected or intensive modeling is required.

CONCLUSION

There has been a good deal of discussion on the appropriateness of cost-effectiveness information estimated from RCTs, and its transferability to other contexts. This paper has contributed to this discussion by detailing and amplifying the issues on the cost side and by improving the documentation of relevant factors, such as patient case-mix, learning effects and patient compliance. There are many shortcomings to cost estimations based on RCTs. Various factors have been discussed which are linked to the limited perspective of RCTs, the methodological features of economic evaluations conducted together with RCTs, and the limited generalizability of the results of RCTs. Many of the factors do not only exist in the context of an RCT but also in other research designs. They are specific to the disease and the related intervention studied; a factor which constitutes a major bias regarding one intervention may not affect another. Moreover, not all factors are relevant at all stages of the diffusion of a health care technology; for example, learning effects and indication expansion may be relevant only at a later stage of diffusion. We have indicated which classes of interventions are most prone to effects from these factors. Finally, options to account for the factors have been defined and a checklist for both researchers and decision makers has been developed.

The issue raised in this paper - that RCTs do not necessarily generate reliable data for decision making - does not imply that linking cost studies to RCTs should be

discouraged. On the contrary, if conditions are satisfied to successfully extrapolate its results to other decision contexts, economic evaluations based on RCTs fulfill an important role in supporting allocation decisions in health care. By creating a checklist, a first attempt is made in this paper to define these conditions. This checklist and further explorations in this paper are intended to support decisions where the results of RCT-based studies have to be transferred to another decision context. Decision makers should apply the checklist to the intervention which is being considered. This should help them to gain a better insight into the relevance of the cost estimations for their own setting, although the overall effect of the various factors will remain difficult to predict. Moreover, the checklist should also be considered by economic evaluators when constructing and reporting RCTs.

Many of the factors discussed as to why costs in the real world may differ from those as estimated alongside an RCT likewise hold for the effect data. Indeed, in medical research, deviations of effectiveness from efficacy have been elaborated on systematically.^{2,41} Research, however, on how to 'translate' efficacy data into effectiveness data is limited. One explanation for this little research is that decisions in the medical world are mostly taken at the individual patient level, where population effectiveness is less relevant. The issue is, on the other hand, approached in outcome research.⁴² In cost-effectiveness research, the aim is to support allocation decisions at the population level which implies an important role for research on how to transpose costs and efficacy data from RCTs to more reliable real world cost-effectiveness data for decision making.

This paper emphasizes the special need to consider real world issues in the standardization discussion. High-quality studies should make it easy for decision makers to transfer the study results to their own context or at least to test their appropriateness for their decision context. A number of issues raised in this paper have been recognized in the process of standardization. For example, Australian guidelines on economic evaluation urge that protocol-associated resource use should not be included in the analysis.¹² Ontario guidelines recommend that indications of the generalizability of results to other subgroups should be reported.⁴³ Still, many other important real world issues, such as the impact of non-compliance, are not yet included in the guidelines. This analysis may be an aid to the further development of guidelines for economic evaluation.

Various factors analysed in this paper that cause deviations between costs as estimated through the RCT and as occurring in the real world, like the age-structure of patients treated, can be regarded as context-specific differences between RCTs and the real world. Other factors should be considered as efficiency differences, e.g. physician experience; at the lower end of the learning curve relative young physicians may incur higher costs than their older counterparts. It depends on the type of decision context whether these inefficiencies in health care provision should be

considered when extrapolating results from RCTs: in an allocation decision context, these inefficiencies should be included not to underestimate the total costs stemming from the use of a medical intervention. However, in a reimbursement context, to furnish incentives for an efficient provisions of medical interventions, these inefficiencies should not be included when deciding at the fee levels.

For improved decision making, supplying decision makers with economic evaluations that lead to more accurate interpretation is only one step: the next step is to consider further requirements. A decision maker who also has budget responsibility and who allocates funds at the population level must also consider the total budget impact and the total achievable health gains. This requires two types of analyses: first, an analysis at the patient level, as described in this article, to increase the internal and external validity of economic evaluations. Second, it requires an analysis at the population level, i.e. the formulation of the epidemiologic aspects of the medical intervention, to assess how many patients are, and will become, medically eligible. Several factors which may play a role at this second aspect have been documented³³, such as the distribution of patients over different stages of disease severity. However, a good link between results from RCTs and the allocation of health care resources at the population level has not yet been established. This paper is a first attempt to provide the baseline information essential for the development of such a framework.

REFERENCES

1. U.S. Congress, Office of Technology Assessment. *Identifying Health Technologies that Work: Searching for evidence*. OTA-H-60, Washington DC: U.S. Government Printing Office, 1994.
2. Spilker, B. *Guide to Clinical Trial.*, New York: Raven Press, 1991.
3. Drummond, M.F. and Davies, L. Economic analysis alongside clinical trials: Revisiting the methodological issues. *International Journal of Technology Assessment in Health Care* 1991; 7: 561-573.
4. Petrou, S., Malek, M. and Davey, P.G. The reliability of cost-utility estimates in cost-per-QALY league tables. *Pharmacoeconomics* 1993; 3: 345-353.
5. Luce, B.R. Cost-effectiveness analysis: Obstacles to standardisation and its use in regulating pharmaceuticals. *Pharmacoeconomics* 1993; 3: 1-9.
6. Luce, B.R. and Elixhauser, A. Estimating costs in the economic evaluation of medical technologies. *International Journal of Technology Assessment in Health Care* 1990; 6: 57-76.
7. Garber, A.M. Competing risks and the cost-effectiveness of health interventions (1989 Annual Meeting Program and Abstracts). *Medical Decision Making* 1989; 9: 319.
8. Hoogenveen, R.T., Verkley, H. and Jager, J.C. *Verkennde studie betreffende concurrerende en vervangende doodsoorzaken in verband met kwantitatieve gezondheidsdoelen* (Explanatory study concerning concurrent and replacing causes of death in relation to quantitative health goals). Report nr. 958606 004, RIVM, Bilthoven 1990 (in Dutch).

9. Rijksinstituut voor Volksgezondheid en Milieuhygiëne (RIVM). *Volksgezondheid Toekomst Verkenning: De gezondheidstoestand van de Nederlandse bevolking in de periode 1950-2010* (The health status of the Dutch population in the period 1950-2010), chapter 4, The Hague: Sdu Uitgeverij Plantijnstraat, 1993 (in Dutch).
10. Duin, van C.M. and Hofman, A. Relation between nicotine intake and cancer conference. *British Medical Journal* 1991; 302: 1491-1494.
11. Blaker, D., Detsky, A., Firestone, T., Hubbard, E., Kennedy, W., Koncinski, R., Menon, D., Schubert, F., Olver R. and Tugwell, P. *Guidelines for Economic Evaluation of Pharmaceuticals*. Ottawa: Canada, 1994.
12. Commonwealth Department of Health. *Guidelines for the pharmaceutical industry on preparation of submission to the pharmaceutical benefits advisory committee*. Canberra: Australian Government Publishing service, 1992.
13. Drummond, M.F., Stoddart, G.L. and Torrance, G.W. *Methods for the economic evaluation of health care programmes*, Oxford: Oxford University Press, 1987.
14. Schechter, C.B., Rose, D.N., Fahs, M.C. and Silver, A.L. Tuberculin screening: Cost-effectiveness analysis of various testing schedules. *American Journal of Preventive Medicines* 1990; 6: 167-175.
15. Paltiel, A.D. and Kaplan, E.H. Modeling Zidovudine therapy: A cost-effectiveness analysis. *Journal of the Acquired Immune Deficiency Syndrome* 1991; 4: 795-804.
16. Rutten, F.H., Bonsel, G.J. and Uyl-de Groot, C.A. Economic evaluation and clinical trials — methodological and clinical aspects. In M. Chytil, G. Duru and W. van Eimeren (Eds.), *The Challenge of Change*. Fifth International Conference on System Science in Health Care, Prague: Omnipress, 1992.
17. Fletcher, R.H., Fletcher, S.W. and Wagner, E.H. *Clinical Epidemiology: The Essentials*, Baltimore: Williams and Wilkins, 1988.
18. Urquhart, J. and Chevalley, C. Impact of unrecognized dosing errors on the cost and effectiveness of pharmaceuticals. *Drug Information Journal* 1988; 22: 363-378.
19. Averill, R.F., McGuire, T.E., Manning, B.E., Fowler, D.A., Horn, S.D., Dickson, P.S., Coye, M.J., Knowlton, D.L. and Bender, J.A. A study of the relationship between severity of illness and hospital costs in New Jersey hospitals. *Health Services Research* 1992; 27: 587-605.
20. Woods, J.R., Saywell, R.M., Nijhuis, A.W., Jay, S.J., Lohrman, R.G. and Halbrook, H.G. The learning curve and the costs of heart transplantation. *Health Services Research* 1992; 27: 219-238.
21. Voss, G.B.W.E. *Severity of illness and costs of medical care in patients with acute myocardial infarction*. Thesis, Maastricht: Universitaire Pers Maastricht, 1993.
22. Gerard, K. and Mooney, G. QALY league tables: Handle with care. *Health Economics* 1993; 2: 59-64.
23. Mitchell, J.B. Time trends in inpatient physician spending. *Health Services Research* 1993; 28: 641-660.
24. Horneffer, P.J., Gardner, T.J. and Monolio, T.A. The effects of age on outcome after coronary bypass surgery. *Circulation* 1987; 76: 6.
25. Guyatt, G., Drummond, M., Feeney, D., Tugwell, P., Stoddart, G., Haynes, R.B., Bennett, K. and Labelle, R. Guidelines for the clinical and economic evaluation of health care technologies. *Social Science and Medicine* 1986; 22: 393-408.

26. Baltussen, R., Leidl, R. and Ament, A. The impact of age on the cost-effectiveness of medical interventions and its control in decision making. *Health Economics* 1996; 5: 227-239.
27. Custer, W.S. and Willke, R.J. Teaching hospital costs: The effects of medical staff characteristics. *Health Services Research* 1991; 25: 831-857.
28. Grannemann, T.W., Brown, R.S. and Pauly, M.V. Estimating hospital costs: A multiple-output analysis. *Journal of Health Economics* 1986; 5: 107-127.
29. Lieberman, M.B. The learning curve and pricing in the chemical pricing industries. *RAND Journal of Economics* 1984; 15: 213-228.
30. Belkaoui, A. *The learning curve*. Westport CT: Quorum Books, 1986
31. Hannan, E.L. Kilburn, H., O'Donnell, J.F., Bernard, H.R., Shieds, E.P., Lindsey, M.L. and Yazici, A., A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. *Health Services Research* 1992; 27: 517-542.
32. Burns, L.R. and Wholey, D.R. The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Medical Care* 1991; 29: 251-271.
33. Leidl, R. 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.
34. Kesteloot, K. and Penninckx, F. The costs and effects of open versus laparoscopic cholecystectomies. *Health Economics* 1993; 2: 303-312.
35. Drummond, M.F., Bloom, B.S., Carrin, G., Hillman, A.L., Hutchings, H.C., Knill-Jones, R.P., Pouvoirville de P. and Torfs, K. Issues in the cross-national assessment of health technology. *International Journal of Technology Assessment in Health Care* 1992; 8: 671-682.
36. Hannan, E.L., O'Donnell, J.F., Kilburn, H., Bernhard, H.R. and Yazici, A. Investigation of the relationship between volume and mortality for surgical procedures performed in New York States hospitals. *Journal of the American Medical Association* 1989; 262: 503-510.
37. Kelly, J.V. and Hellinger, F.J. Heart disease and hospital deaths: An empirical study. *Health Service Research* 1987; 22: 369-395.
38. Leidl, R. A survey of the economic evaluation of early drug intervention in HIV infection. In E. Kaplan and M. Brandeau (Eds.), *Modeling the AIDS Epidemic: Planning, Policy, and Prediction*. New York: Raven Press Ltd., 1983.
39. Leidl, R. Standardisierung und Flexibilisierung ökonomischer Evaluationsstudien (Standardising and flexibilisation of economic evaluation studies). In Trampisch J.H., Lange S., (Eds.) *Medizinische Forschung - Ärztliches Handeln*; Bochum, September 1995, München: MMV Medizin, 1995 (in German).
40. Sclar, D.A., Tartaglione, T.A. and Fine, M.J. Overview of issues related to medical compliance with implications for the outpatient management of infectious diseases. *Infect. Agent. Dis.* 1994; 3: 266-273.
41. Tugwell P., Bennett K. J., Sackett D. L. And Haynes R. B. The measurement iterative loop: A framework for the critical appraisal of need, benefits and costs of health interventions. *Journal of Chronic Diseases* 1985; 38: 339-351.
42. Goldberg, H.J., Cummings, M.A., Steinberg, E.P., Ricci, E.M., Shannon, T., Soumerai, S.B., Mittman B.S., Eisenberg, J., Heck, D.A., Kaplan, S. et al.

- Deliberations on the dissemination of PORT products: Translating research findings into improved patient outcomes. *Medical Care* 1994; 32: JS90-110.
43. Drugs Programs Branch. *Ontario Guidelines for Economic analysis of Pharmaceutical Products*. Ontario: Ministry of Health, 1993.

CHAPTER 3

INTERPRETING POTENTIAL BIASES: COPING WITH THE IMPACT OF AGE

Published as: Baltussen R.M.P.M.¹ Leidl R.M.^{1,2} Ament A.J.H.A.¹ The impact of age on cost-effectiveness ratios and its control in decision making. *Health Economics* 1996; 5: 227-39

1 From the Department of Health, Organisation, Policy and Economics, Maastricht University

2 From the Department of Health Economics, University of Ulm, Germany

SUMMARY

This chapter examines the consequences of disregarding a specific bias in policy making based on economic evaluation: the impact of age. First, at the clinical level, the relation between age and the cost-effectiveness of medical interventions is analysed. Second, at the population level, a framework is presented which allows researchers and decision makers to assess the impact of these effects on the decision making process. It is shown that the allocation of health care resources at the macro-level is seriously impaired when age is ignored as a variable in cost-effectiveness analysis. Because clinical trials typically employ 'young' populations, when the data is extrapolated to the whole population the attractiveness of medical interventions in terms of cost-effectiveness may be considerably overestimated. Furthermore, the cost-effectiveness ratio may vary across countries or over time as a result of demographic or epidemiologic variation. Economic evaluators should describe the impact of age, which should then be considered by decision makers to control for age effects.

INTRODUCTION

In recent years, there has been much discussion on the effects of population aging on the health care system. Many see the issue of population aging as the most significant cause of the growth of health care expenditure.¹ Other studies have found that population aging should not be regarded simply as a major cost-driving factor but rather as a threat to the financing of the health care system.² It is evident that the consequences of the increasing share of the elderly in the total number of patients and hence in the use of medical interventions warrant closer attention.³ Some authors have suggested age-based rationing of high-technology utilization.⁴ There is therefore a need for adequate information on the impact of age on the cost-effectiveness of medical interventions, as suggested recently.⁵

This article will show that age can have an important impact on the cost-effectiveness of medical interventions, and will demonstrate that decision makers' expectations concerning the effects of population aging will be seriously impaired if this relation is ignored. First, a theoretical examination of the impact of age on the various elements of cost-effectiveness ratios derived from clinical trials is provided. This is supported with empirical evidence from economic evaluations which report age-specific cost-effectiveness ratios and a detailed analysis of the impact of age on the components of these cost-effectiveness ratios.

The section at the population level analyses the effects of these relationships on decision making. It demonstrates that disregarding age as an important variable in cost-effectiveness analysis may seriously bias allocation decisions regarding health care resources. One reason is that the age structure in the clinical trial may be much different from the age structure of the population in which the intervention will be applied. Thus decision makers who wish to extrapolate results from a clinical trial to their own situation should be very cautious. Another reason is that - because of demographic and epidemiologic developments - the age structure of the target population may alter over time. In both cases, the cost-effectiveness at the population level may be affected. Age should thus be considered as a variable to be controlled for when making allocation decisions in health care. The final section of this article discusses the implications of these findings. The reference point of this paper is the maximization of QALYs (quality adjusted life years) within a given budget.

THE IMPACT OF AGE ON COST-EFFECTIVENESS

Age may influence the effects as well as the costs involved in medical interventions, as illustrated in Figure 1. Medical interventions range from short-term (acute) treatments to long-term, and sometimes life-long, therapies.

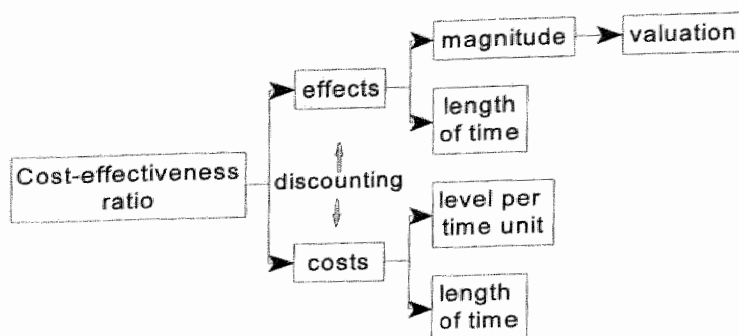


FIGURE 1. The elements of the cost-effectiveness ratio which age may affect

It is necessary to consider the consequences of age over the whole period in which the intervention produces effects and generates costs. Unless costs and effects are concurrent, the chosen discount rate will also affect the impact of age on the cost-effectiveness of medical interventions.

Effects

It has been suggested that, for many interventions, older patients tend to respond less well to therapy than younger patients, with the difference of course depending on the type of intervention.⁶ Normal physiological aging is characterized by a gradual loss of reserve capacity which may be caused by older persons' decline in cardiac and renal function, muscle strength (including the respiratory muscles) and mental acuity. This decline in reserve capacity may slow down the rehabilitation of older patients.^{7,8} For example, it has been demonstrated that the probability of death or non-ambulation at one year after hip fracture increases markedly with age above the age of 60. While 80% of patients who are less than 60 years old will survive and be able to walk, only 6% of those older than 90 will have a similar outcome.⁹ Not only may older people respond less well to therapy, the risk of complications also increases. It has been found, for example, that the risk of thrombosis increases with age.¹⁰ As regards drug treatments, it has been noted that the elderly have a much more variable physiological response and a much less predictability of drug action than younger people, leading to an increased incidence of adverse drug reactions and therapeutic failure for the elderly.¹¹ Relationships between increasing age and risk of complications have also been found for other medical interventions, such as lower limb angioplasty for ischemic diseases¹², the treatment of hypertension¹³ and the use of bioprosthetic heart valves.¹⁴

Besides the impact of age on the magnitude of effects per se, age may also have a significant influence on the valuation of health states by patients. It has been found, for example, that older patients assign a lower utility to dialysis and transplant therapies, indicating that they experience a lower gain in quality-adjusted life years (QALYs) from these interventions than their younger counterparts, other things being equal.¹⁵ On the other hand, other studies have found that older patients assign higher utilities to certain interventions or have found no differences in preferences attributable to age.¹⁶

Age may also affect the duration of the effects. In the case of life-extending interventions, younger patients are more likely to have a favourable outcome than older patients because the young have a higher remaining life expectancy^a than older people. For example if patients of various ages are undergoing a life-saving intervention, and the efficacy of the intervention is equivalent in both young and old, the potential life-years gained because of the intervention for male patients aged 0 years (newborns), 15, 45 and 65 years will be 73, 58, 30 and 14 years respectively (figures based on life expectancy in Germany in 1990).¹⁷ It is evident that the benefits of the intervention, measured in life-years saved, are much greater for the younger patients than for the older patients. However, for interventions producing only relatively short-lasting benefits, gains in life-years will be approximately the same for both age groups, other things being equal.

Costs

It has been argued that costs are generally positively related to age; older patients cost more to treat.^{5,18-20} A complete understanding of the relationship between age and the costs of medical interventions per time unit (e.g. per admission) would require a comprehensive analysis of all such effects across medical interventions for all disease categories. As an example, we use some evidence of the average hospitalization duration for pneumonia treatment in the Netherlands. In 1994 the average length of stay of all admissions was found to increase gradually with age; for patients in the age groups 55-64 years, 65-74 years, 75-84 years and 85+ years old the average length of stay was 14.6, 17.1, 19.2 and 19.9 days, respectively.²¹ This indicates an increasing resource consumption with age. The exact magnitude of this relationship between age and costs, of course, depends on the type of intervention.

Furthermore, age may influence the length of the period in which recurring costs are generated, with respect to a particular illness. Since younger patients have a higher remaining life expectancy, they can be expected to generate higher costs for illnesses whose treatment involves long-term recurring expenses - other things being equal. This has, for example, been demonstrated for hypertension treatment.²²

Discounting

Incorporating discounting in a cost-effectiveness study may very well affect the relationships which are found between age and effects and between age and costs. Regarding the relation between age and effects, discounting tends to reduce the greater effect which is otherwise found for younger patients. Their gains in life years are diminished more by discounting because they occur in a more distant future than the gains of the elderly. Welch¹⁸ presented an example of a life-saving intervention in which the gains of saving the life of a 5-year old and a 40-year old were respectively 70.6 and 37.2 years without discounting; a discount rate of only 5% made this difference, however, almost negligible, respectively 19.5 and 16.8 years.¹⁸

For costs, the impact of discounting depends on the length of the period during which the costs are generated. In the case of costs which are incurred over the whole life-span of a patient, the net present value of the costs of younger people will be reduced more by discounting than will the costs of older patients. However, discounting will have only a minor impact on the relationship between age and costs in the case of interventions that only involve short-term costs.

The net impact of discounting on the relationship between age and the cost-effectiveness ratio depends on the intervention under scrutiny. However discounting generally has a unfavourable impact on the cost-effectiveness of medical interventions, because the effects typically occur in a more distant future than the costs. This effect is more profound for younger patients than for older patients, because the effects for the former will be gained in a more distant future, but this does not apply to the costs to the same extent. Kawachi et al.²² showed for the case of treatment of hypertension that the use of a discount rate of 10% instead of 5% implies a more than 50-fold increase of the cost-effectiveness ratio for the 40-year old and a slight 1.6-fold increase for the 60-year old. Thus when a higher discount rate is applied, the cost-effectiveness of medical interventions generally become less favourable, with the magnitude of the reduction being larger for young than for old patients.

It is important to consider that the examples discussed above are illustrative only and that the nature of the impact of age on costs and effects may vary across medical interventions. As various age-effects may reinforce or counterbalance each other, the adjustment that age will have on cost-effectiveness ratios is particularly uncertain.

REPORTING OF AGE EFFECTS IN ECONOMIC EVALUATIONS

A MEDLINE search^b for the period 1991-1994 found 307 abstracts of articles on cost-effectiveness and cost-utility studies, of which 43 (14%) reported that age had been considered in the analyses. As results of cost-effectiveness studies are not

homogeneously reported - sometimes as costs per life saved, as costs per life-year saved or as costs per infection averted - a full comparative analysis of age effects for this type of economic evaluation is difficult. For this reason, the analysis of the impact of age is restricted to cost-utility studies alone - which uniformly report cost per QALY ratios. Of the 40 cost-utility studies found in the Medline search, the abstracts of 8 reported that age had been considered in relation to the cost-utility ratio under scrutiny. These studies reported on early surgery in the management of small abdominal aortic aneurysms (A.A.A., a health problem caused by a widening or rupture of an artery)²³, screening and treatment of A.A.A.²⁴, radiofrequency ablation in the Wolff-Parkinson-White syndrome (a health problem caused by a rhythm disturbance of the heart)²⁵, chemotherapy for node-negative breast cancer²⁶, treatment of mild-to-moderate hypertension²² and tuberculin screening.²⁷ One article written in Chinese and another which reported overlapping age-cohorts have not been considered further. To illustrate the importance of age effects, the size of the impact for the six selected studies is presented in Figure 2.

For each intervention, a ratio index has been calculated to indicate the sensitivity of the cost-utility ratio to patient age. The unweighted average of the upper and lower age-related bounds of the cost-utility ratio has an index value of 100. For example, early surgery in the management of A.A.A. has a cost-utility ratio of \$17,406 for a 60-year old, and \$140,972 for an 80-year old. This more than eight-fold increase is represented in the figure by a corresponding shift of the ratio index from 22 to 178. The figure shows that, in the cost-utility studies selected, age has a large impact. Taking the lowest value as the basis, the variations in the cost-utility ratios range from 44% in the case of breast cancer screening to a 810% in the case of early surgery in the management of A.A.A.

Identification of variation

The selected studies will be used to ascertain whether the mechanisms which were theoretically identified above are indeed present and relevant, and to compare the impact of age on the costs, effects and cost-effectiveness of these interventions.

In the case of chemotherapy for node-negative breast cancer, the benefits decline steadily from age 60 to 80, not only because of the lower remaining life expectancy of the older patients, but also because they suffer a higher mortality from other causes and experience a higher decline in health status stemming from the toxic effects of chemotherapy. The lower remaining life expectancy means that 80-year old patients generate less follow-up costs and have fewer recurrences than the younger age categories (the article does not explicitly report whether the costs of chemotherapy are affected by age). However increasing age causes the effectiveness of chemotherapy to decrease more rapidly than costs, which results in an increasingly unfavourable cost-utility ratio with age. Discounting may diminish this result, but does not eliminate it.²⁶

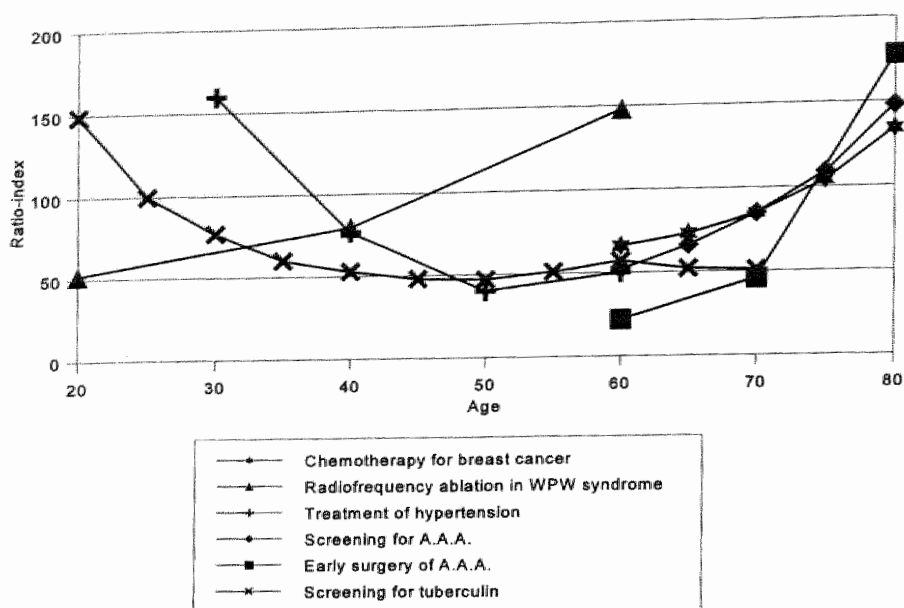


FIGURE 2. Age-related cost-effectiveness ratios

In the case of radiofrequency ablation in the Wolff-Parkinson-White syndrome, the older patients have a lower gain in QALYs compared to the younger patients because of a lower remaining life expectancy. Whether the magnitude of the effects is also affected by age is not explicitly reported. The bulk of the costs of the intervention are associated with the operations involved. As these operations cost more or less the same for all ages and the costs are only incurred once, costs can be considered the same for all age groups. Overall, the cost per utility gained increases with age, although discounting may reduce this effect somewhat.²⁵

For the treatment of hypertension, the elderly gain more in terms of magnitude of effects than the young. However, because younger patients have a higher remaining life expectancy, they gain more in terms of QALYs. The costs of the treatment per time unit (e.g. per year) are not age-related but young patients generate more costs than the elderly because costs are incurred over the whole remaining life-span of the patients. Although the total effects gained for the young are higher than for the older patients, their higher costs and the effects of discounting imply that, in present-value terms, the cost-utility ratio of the treatment of hypertension becomes more favourable with age in the range from 30 to 50 years. For higher ages, the effect of discounting diminishes; the cost-utility ratio becomes less favourable with age due to the lower net gain in remaining life expectancy at age 60 years compared with age 50 years.²²

With respect to screening and treatment of A.A.A., the effects of the intervention decline with age as the operative mortality of elective repair rises and the life expectancy falls. As the costs of the treatment are incurred only once and are reported to be the same for all age-groups, age has no impact on the size of the total costs. Consequently, the cost-utility ratio becomes less favourable with age, although discounting may reduce this effect.²⁴

Regarding early surgery in the management of A.A.A., increasing age has a negative impact on the effects gained. Elective operative mortality rate increases with age and, because remaining life-expectancy declines with age, the elderly benefit less than the young in terms of the number of life-years gained. No age-related effects on the costs of the operation are reported but it is reported that a significant proportion of the patients will require long-term skilled nursing care. So, assuming that the nursing needs of younger and older patients are the same, the former will, because of their longer remaining life expectancy, generate higher average costs. Overall, the cost-utility ratio becomes less favourable with age. Discounting reduces this effect to some extent.²³

It is not reported whether the magnitude of the effects of the treatment for tuberculin depend on age, but the number of QALYs gained is greater for the younger patients because they have a higher remaining life expectancy. Total lifetime costs do not differ across age groups because the costs per treatment may be considered to be equal across all age groups and the treatment costs are incurred for only a relatively short time. At the programme level, however, there are age-related differences in costs: screening all 20-year old costs much more than screening all 70-year old because the former group is twice as numerous. Because of this phenomenon, the cost-utility ratio is nearly the same for screening all cohorts between 35 and 70 years; both the costs incurred and the QALYs gained decrease with age. The cost-utility ratio for screening patients aged between 20 and 30 years is relatively unfavourable because the large screening programme generates relatively high costs compared to the benefits gained.²⁷ The results of the above discussion are summarized in Table 1.

We can conclude that the remaining life expectancy of the patients was a major factor in determining the impact of age on the cost-utility ratio of the medical interventions. On the effect side, young patients may experience a higher gain in effects than the elderly in case of interventions producing long-lasting benefits. However, on the cost side they may generate higher costs where life-time costs are involved. It was also found that in most but not all cases, increasing age has a negative influence on the magnitude of effects. None of these evaluations explicitly reported whether patients valuations are subject to age effects. Furthermore, the level of costs incurred per period was either found to be not affected by age, or was not explicitly reported. Discounting the costs and effects was in all cases found to favour the elderly compared to the young.

TABLE 1. The impact of increasing age on effects, costs, and cost-utility ratios as reported in the evaluation of six medical interventions^{a, b}

intervention/element	effects ^c		costs		cost-utility ratio
	duration	magnitude	duration	level	
chemotherapy for breast cancer	+	+	-	N.A.	+
radiofrequency ablation in WPW syndrome	+	N.A.	0	0	+
treatment of hypertension	+	-	-	0	+/-
screening and treatment of A.A.A.	+	+	0	0	+
early surgery of A.A.A.	+	+	-	N.A.	+
tuberculin screening	+	N.A.	0	0	+/-

- a) This figure shows the direction of the impact of age on the various elements of the cost-utility ratios. In cases where the impact of greater age on the specific element causes the cost to utility ratio to rise, this is denoted by '+'. Where the impact of age on the specific element causes the cost-utility ratio to decline, this is denoted by '-'. Where age has no effect, this is denoted by '0'. If the effect is not explicitly reported, this is expressed by 'N.A.' (Not Available).
- b) In all six interventions, the use of a higher discount rate will diminish the change of the cost-utility ratio that is induced by increasing age.
- c) None of the six studies reported the impact of age on the valuation of effects.

The empirical evidence confirms that age may affect the cost-effectiveness of medical interventions in various ways and that this influence does not necessarily feature a linear relationship. Decision makers should thus consider possible age effects when using the results of economic evaluation studies, but should also be cautious when making assumptions on the relation between age and the cost-effectiveness of medical interventions.

THE POPULATION LEVEL

While economic evaluations are usually conducted at the clinical level, allocation decisions are typically taken at the population level. A budget holder who allocates funds at the population level will not be primarily interested in patient-based cost-effectiveness ratios but rather in the total costs, total effects and cost-effectiveness ratios which are relevant for a given population. Several factors may cause these to deviate from expectations based on patient-level cost-effectiveness ratios.²⁸ The age structure of the clinical trial is another factor. To illuminate possible biases in decision making stemming from this factor, we will extend the analysis from the clinical level to the population level. The methodological framework is first presented below, followed by a close analysis of two types of potential biases.

Assessing the impact on decision making

To describe the full impact of the relationship between age and the cost-effectiveness ratios of medical interventions on decision making at the population level, it is necessary to link demographic, epidemiologic and economic aspects. This can be done by adapting a methodology that has been used to describe the future impact of diseases on the health care system.²⁹ The approach distinguishes a demographic module, an epidemiologic module and a utilization module which is in this case replaced by the cost-effectiveness ratio.

The cost-effectiveness ratio from a clinical trial is represented in equation 1, in which the indicator i denotes the various age-groups (e.g. 0-20, 21-50, 51-70 and 70+ years). Indicator n represents the number of people in the various age groups, with N as the total number of people.

$$(1) CER_{RCT} = \frac{1}{N} \sum n_{POPRCT, i} * CER_{RCT, i}$$

The ratio (CER_{RCT}) which is traditionally reported is - explicitly or implicitly - calculated as the weighted average of the number of patients belonging to various age-groups enrolled in the trial ($n_{POPRCT, i}$) times the age-specific cost-effectiveness ratio of these age-groups ($CER_{RCT, i}$). It is obvious that the resulting cost-effectiveness ratio (CER_{RCT}) is an average figure that disregards the underlying age-structure if only reported as such.

When examining the biases for decision making which arise from the impact of age on the cost-effectiveness of medical interventions, the most appropriate level for analysis is the population level. Here, demographic, epidemiologic and economic aspects must be linked. This approach is illustrated in equation 2.

$$(2) CER_{POP} = \frac{1}{N} \sum n_{POP, i} * EPI_i * CER_{RCT, i}$$

The product of the first two elements, the number of patients belonging to various age-groups in the population ($n_{POP, i}$) and the age-specific epidemiological features of the population (EPI_i), gives the number of people medically eligible in the population. The cost-effectiveness ratio at the population level is the weighted average of the number of people medically eligible times the age-specific cost-effectiveness ratio as estimated at the clinical trial ($CER_{RCT, i}$). Applying this equation, one can account for the impact of age in estimating the cost-effectiveness ratio at the population level, but it requires economic evaluations which report age-specific cost-effectiveness ratios.

However, economic evaluations usually do not report age-specific ratios. In that case, the cost-effectiveness ratio stemming from the clinical trial and used in the equation

is independent of age (CER_{RCT}) which implies that the cost-effectiveness ratio at the population level cannot be corrected for the impact of age.

It is evident that, in accounting for the impact of age on the cost-effectiveness of a medical intervention, the key figure for the decision maker is the age composition of those actually medically eligible for the intervention. This age composition may be subject to various sources of variation: it may differ from the age-composition of the patients enrolled in the clinical trial, and it may alter because of demographic or epidemiologic variations, either across countries or over time. These issues are discussed in more detail in the following sections.

Age-effects related to the representativeness of the age composition in clinical trials
Clinical trials tend to include cases which are most likely to produce detectable effects, i.e. relatively young patients, and are rarely conducted on adults over 65 years of age³⁰, although the interventions under study may very well be applied to this age-category. It has been indicated above that age may very well affect the cost-effectiveness of an intervention since older patients may respond less well to therapy, their gain in potential life-years may be less, their valuation of health states may be different and their costs of treatment may be higher. Consequently, the cost-effectiveness ratio estimated in the clinical trial, for a patient case-mix consisting of a specific age-group, may not be representative when extrapolated to other age-groups.³¹ This can be illustrated using the case of early surgery in the management of A.A.A. for the example of Germany, the European country in which aging, and consequently the impact of age on the cost-effectiveness of interventions, is most significant.³² Table 2 shows the age-related variations in the cost-utility ratio, and cost-utility ratios at the population level calculated according to the equations above.

If only the base-case analysis is taken into account, thus ignoring age-related variations, the cost-utility ratio at the population level equals \$17,406. Taking the age-specific cost-utility ratios into account (equation 2), the cost-utility ratio at the population level is \$60,640. This demonstrates that ignoring age-related variations in the cost-utility ratio may seriously bias decision making when results from economic evaluations are extrapolated to the population as a whole.

The impact of demographic and epidemiologic variations

By definition, population aging will change the age-composition of a population and, given constant age-related incidence and prevalence of disease, will also change the age-composition of those medically eligible for an intervention. Even in a population with a constant age-structure, changes in epidemiology such as increased smoking of young women may change the number of patients with a certain disease, which will also change the age-structure in the long term.

TABLE 2. Early surgery in the management of A.A.A.: representativeness of the age-composition in clinical trials

type of analysis/age	eligible for treatment (x1,000) ^a			age-specific cost-utility ratio (\$) ^b			cost-utility ratio at the population level (\$)
	60	70	80	60	70	80	all ages
base case analysis	18.0	22.7	16.5	17,406	-	-	17,406
age-controlled analysis	18.0	22.7	16.5	17,406	36,589	140,972	60,640

a) The number of patients eligible for treatment is assumed to be equal to the prevalence of disease (in percentages)³³ times the size of the male population in Germany¹⁷

b) For men, 5% discount rate; 1992 US\$²³

These demographic and epidemiologic developments may lead to changes in the age-composition of the patients medically eligible, so that the cost-effectiveness of medical interventions at the population level may diverge both across countries and over time. The potential divergence of the cost-effectiveness ratio at the population level across countries is illustrated in Table 3, using the results from the study on early surgery in the management of A.A.A. as an example. Equation 2, which allows both the demographic and the epidemiologic module to change, has been applied^c.

TABLE 3. Early surgery in the management of A.A.A.: demographic variations across countries and the cost-utility ratio

country/ age	population estimate 1995 (x1,000)			eligible for treatment 1995 (x1,000) ^a			cost-utility ratio ^b			cost-utility at the population level
	60	70	80	60	70	80	60	70	80	all ages
Japan	781 ^c	447	227	31.4	29.8	20.4	17,406	36,589	140,972	55,288
Germany	449 ^d	340	184	18.0	22.7	16.5	17,406	36,589	140,972	60,640
U.S.	1,058 ^e	940	558	42.5	62.7	50.1	17,406	36,589	140,972	65,006

a) The number of patients eligible for treatment is assumed to be equal to the prevalence of disease (in percentages)³³ times the size of the male population in the respective countries.

b) For men, 5% discount rate²³

c) Source: Japan Statistical Yearbook 1993/1994³⁴

d) Source: Council of Europe Press 1994¹⁷

e) Source: Statistical Abstracts of the United States 1993³⁵

It is shown that the numbers of patients eligible for treatment per age cohort and thus the cost-per-QALY ratios at the population level clearly differ between Germany, Japan and the United States. The United States, with a relatively large number of 80-years old patients in the total number of patients medically eligible, has the highest cost to utility ratio at the population level. It lies almost 20% above that of Japan.

This indicates the need for caution regarding the age component in results from economic evaluations when transferring study results to different countries.

Demographic and epidemiologic structures can vary not only across countries but also over time. This involves a change in the number and distribution of patients across various age cohorts. For example there may be a general aging (demographic change) or the incidence of diseases such as cancer in certain age groups may change due to changing smoking or dietary behaviour (epidemiologic change). The implication is that the demographic and epidemiologic modules of equation 2 are subject to change over time; this is accomplished by combining a projection of the demographic module, a projection of the epidemiologic module and the constant, current cost-effectiveness ratio (thus assuming constant medical technology).

Figure 3 shows the results, using data from the two economic evaluations of early surgery in the management of A.A.A. and the treatment of mild-to-moderate hypertension^d. The figure graphically presents the changes in the cost-utility ratios at the population levels of the two medical interventions as the result of demographic and epidemiologic change. The size and direction of these variations depend on the age pattern of the cost-utility ratios^e. Prior to the year 2013, early surgery in the management of A.A.A. has a more favourable cost-utility ratio at the population level than the ratio for the treatment of hypertension. Beyond the year 2013, the treatment of hypertension involves a lower cost to utility ratio at the population level, so that intervention will then be preferred. Thus, these demographic and epidemiologic changes in cost-effectiveness ratios may have important consequences where medical interventions are prioritized to ensure the optimal allocation of health care resources. The rank order of interventions in the League Tables may alter, and allocation decisions, at least at the margin, may change.

The time frame chosen (25 years) may seem too long to produce relevant results, given the assumption of constant medical technology. It could be considered likely that technological progress will produce better or cheaper medical interventions for the disease being considered within that time-span. However, the example has shown that even in the short-run - in which medical technology may be assumed constant - considerable deviations in the cost-effectiveness of medical interventions at the population level can be expected. Moreover, the purpose of the above exercise was to focus attention on the importance of possible future variations in the age composition of the medically eligible population and the impact of such variations on decision making.

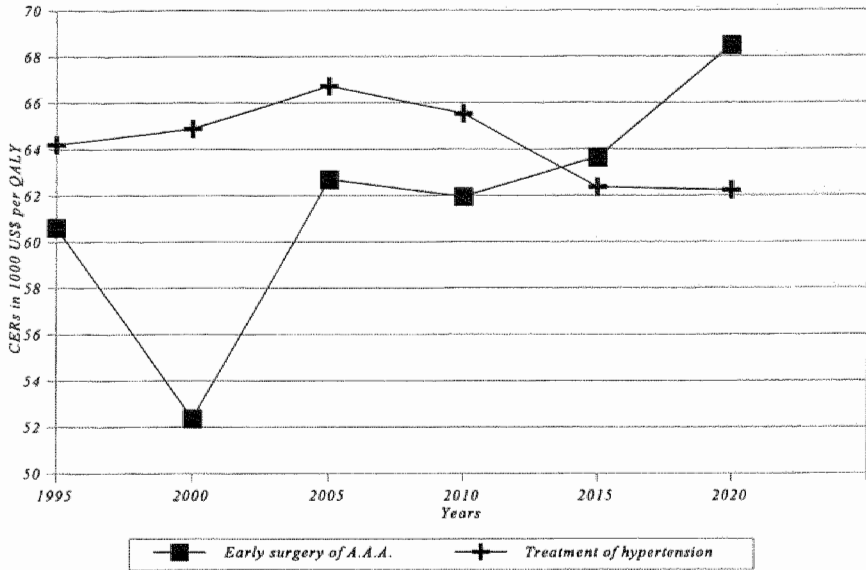


FIGURE 3. The impact of demography on cost-effectiveness ratios

DISCUSSION

This study has shown the need to investigate and report the age relationship of cost-effectiveness ratios. It has shown that age-specific ratios may be important information when deciding on the allocation of health care resources. Ignoring this information when extrapolating data from a clinical trial to the population as a whole may lead to a considerable underestimation of the costs involved, because of the typically 'young' trial populations. Transferring cost-effectiveness ratios which are not specified by age to other country contexts may also produce a substantial bias. Moreover, it has been noted that because of demographic change and the relation between age and cost-effectiveness ratios, the prioritizing of medical interventions over time may alter. This indicates the importance of age-specific ratios for the optimal allocation of health care resources over time.

Economic evaluators should provide information on age effects. If age is not found to have a significant impact on the cost-effectiveness of a medical intervention, this should be explicitly reported. In other cases, economic evaluators should provide the information which decision makers need if they are to foresee and control potential biases. First, age-specific costs, effects and cost-effectiveness ratios need to be reported. Second, indications of the age-groups to which the medical intervention

will be applied in real practice need to be included in the analysis. Third, expected epidemiological developments of the disease concerned need to be described. This information should be provided in a form suited to the needs of decision makers: that is, the cohorts which are expected to be most relevant to decision makers should be analysed.

Where reports of economic evaluations do not include specific information on age effects, decision makers should check whether the cost-effectiveness of the medical intervention under scrutiny is sensitive to age. To this end, a number of suggestions regarding the impact of age have been formulated. These can be applied to any intervention, and should indeed be applied when making decisions about a particular intervention. They enable decision makers to gain a better understanding of the impact of age on the intervention, and how to control for this. First, decision makers should note that those interventions which generate benefits over the whole life-span of the patients involved are more subject to age effects than interventions which generate only short-term benefits.

Second, decision makers should verify whether age has any impact on response to therapy for the medical intervention they are considering; it is not evident, *a priori*, which interventions are most prone to age effects. Third, it should be noted that the valuation of effects may differ across age groups. The impact of this phenomenon, and which interventions are most affected, is difficult to determine or anticipate but it should be considered. Fourth, with regards to costs, decision makers should recognize that interventions which involve recurring costs over a long period are affected more by age effects than interventions which generate costs only once.

Fifth, it should be realized that the relation between age and the costs generated per period is apt to vary between interventions, although it is uncertain which interventions are most sensitive to the impact of age. Sixth, it should be recognized that the results of economic evaluations which employ relatively high discount rates are relatively more prone to age effects than are results from economic evaluations using lower discount rates. However, the various effects may counterbalance each other; for example, the effectiveness of a medical intervention may be unfavourably affected by increasing age, for example where its effects last the whole life-span of the patients, but this might be outweighed by the favourable effect of lower costs with increasing age.

The need to control for age effects should also be considered in the discussion on the standardization of economic evaluation studies. High-quality studies should make it easy for decision makers to control for age-effects or at least to identify them. Some issues raised in this paper have indeed been recognized in the process of standardization. For example, the Ontario Guidelines for Economic Analysis of Pharmaceuticals Products recommended that indications on the generalizability of

results to other subgroups should be reported.³⁷ Still, this is not in itself sufficient to allow decision makers to control for age effects. The analysis in this paper may be an aid to the further development of guidelines for economic evaluations.

Of course, age is not the only factor which should be controlled for when extrapolating results from clinical trials. The cost-effectiveness of treatment may vary well depending on which patients are treated, which physicians treat them and where the treatment takes place. Consequently, decision makers who wish to extrapolate clinical trial results to their own context should bear in mind that the patient group which was studied, the treating physicians and the setting of the clinical trial may not be representative, causing the cost-effectiveness to differ. For example, in one study it was found that the cost-effectiveness of cholesterol-lowering treatments varies considerably with the presence of risk factors. For men, the costs per year of life saved (most-favourable ratios) without risk factors is US\$58,000, but with multiple risk factors only US\$13,000.³⁸ To assess the relevance of economic evaluations for their own decision context, decision makers should apply a check-list like that developed by Baltussen et al.³¹

If the goal of allocation in health care is the maximisation of health effects within a limited budget, the prioritizing of medical interventions takes place according to their cost-effectiveness ratio. Strictly applied this means that the explicit differentiation to age, in the case that the cost-effectiveness ratio increases with age, will lead that for the same interventions, younger individuals will be given priority over older individuals. Implicitly age-effects also play a role when different medical interventions focus on different age-groups. When maximizing health effects within a given budget, the younger may be advantaged in the sense that their potential gains in effects are larger. For example, when comparing the cost-effectiveness of heart-transplantation with bone-marrow transplantation, it appears that the ratio of the latter is only a quarter of the former. However, this is partly due to the fact that the recipients of bone-marrow transplantations are on average 30-year old, but that recipients of heart transplantations are on average 50-year old.¹⁸ Again, to achieve efficiency, it would have to be decided not to treat those with higher ages.

Thus, the explicit as well as the implicit allocation of resources by age raises questions about the age-related consequences on equity when maximizing health effects within a given budget.^{39,40} The latter, however, is but one rule of allocating funds, which simultaneously implies the equity concept familiar under the term 'QALY-egalitarianism'. A society may choose, however, also for different allocation rules, for example to allocate funds in proportion to the absolute health gains possible, or to the losses induced by illness. Different allocation rules will have different impacts for the various age-groups. In any case, the impact of age on the results of economic evaluations, and its confounding influence on the interpretation of these economic evaluations, clearly need to be analysed before one attempts to

evaluate the decision making process regarding age. This paper has laid a foundation for the analysis of this problem. The issue of age-related cost-effectiveness can only increase in importance as population aging becomes more apparent.

NOTES

- a. Remaining life expectancy is defined as the life expectancy of a patient minus the age of the patient.
- b. The search selected articles containing the words 'cost-effectiveness', 'cost-utility' or 'QALY' in the title, excluding those for which no abstract was included in MEDLINE and those articles which did not refer to an explicit cost-effectiveness or cost-utility study.
- c. In this example, the epidemiologic factor is regarded as being identical across countries, leaving only demographic variations across countries to be taken into consideration. Demographic data for 1995 has been applied.
- d. The cost-utility ratios at the population level are calculated by combining the cost-utility ratios of the various age-groups and the number of people eligible in the respective age-groups. The number of patients eligible for treatment in the various age-groups is again calculated for the example of Germany. The number is assumed to be equal to the prevalence of the disease (in percentage) adjusted for expected epidemiologic change for the two medical interventions from 1995 to 2020^{33,36} times the size of the male population in Germany from 1995 to 2020.¹⁷ The age-specific cost-utility ratios of early surgery in the management of A.A.A. and the treatment of mild-to-moderate hypertension are taken from the respective economic evaluations and are considered constant over time.^{22,23} As this exercise is for illustrative purposes only, no price adjustments regarding year of study have been made.
- e. In the case of early surgery in the management of A.A.A., where the cost-utility ratio is less favourable for older patients, population aging in the future will lead to a less favourable cost-utility ratio at the population level. However in the year 2000 there is a sudden improvement in the cost-utility ratio at the population level, because the number of 80-year olds - having a relatively unfavourable cost-utility ratio - in that year decreases as compared to the number of patients of other ages. In the case of the treatment of hypertension, where the cost-utility ratio is more favourable for older patients, there is initially a smooth rise because of increasing numbers of 60-year olds - who have a relatively high cost to utility ratio - compared to the number of patients of other ages. Past the year 2005, population aging leads to more favourable cost-utility ratios at the population level.

REFERENCES

1. Idenburg, P. and van Vondelen, A. European health care careful towards managed competition. *Ziekenhuis* 1993; 23: 597-617 (in Dutch).
2. Leidl, R. Effects of population aging on health care, expenditure, and financing: some illustrations. In: Callahan, D., ter Meulen, R.H.J., Topinkova E. (editors). *A world growing old: the coming health care challenges*. Hasting Center Studies in Ethics. Washington D.C.: Georgetown University Press, forthcoming (1995).
3. Alastair, M.G. and Bosanquet, N. Forecasting the impact of demographic change: the case of the British National Health Service. *Health Policy* 1992; 20: 269-275.
4. Callahan, D. Old age and new policy. *JAMA* 1989; 261: 905-906.
5. Chelluri, L., Grenvik, A. and Silverman, M. Intensive care for critically ill elderly: mortality, costs and quality of life: review of the literature. *Archives of Internal Medicine* 1995; 155: 1013-1022.
6. Rowe, J.W. Health care of the elderly. *The New England Journal of Medicine* 1985; 312: 827-835.
7. Crosby, D. and Rees, G. Anaesthesia and surgery in the elderly. In: Pathy, M.S.J. editor. *Principles and practice of geriatric medicine*. Checheter: John Wiler & Sons, 1985, 1367-1883.
8. Evers, B.M., Townsend, C.M. and Thompson, J.C. Organ physiology of aging. *Surg. Clin. North. Am.* 1994; 74: 23-39.
9. Millar, C.W., Survival and ambulation following hip fractures. *J. Bone Joint Surg. Am.* 1978; 60: 930-934.
10. Nurmohamed, M.T., Buller, H.R. and ten Cate, J.W. Physiological changes due to age: implications for the prevention and treatment of thrombosis in older patients. *Drugs Aging* 1994; 4: 207-220.
11. Lamy, P.P. Physiological changes due to age: pharmacodynamic changes of drug action and implications for therapy. *Drugs Aging* 1991; 1: 385-404.
12. Morse, M.H., Jeans, W.D., Cole, S.E., Grier, D. and Ndlova, D. Complications in percutaneous transluminal angioplasty: relationships with patient age. *Br. J. Radiol.* 1991; 64: 5-9.
13. Lakatta, E.G., Cohen, J.D., Fleg, J.L., Frohlich, E.D. and Gradman, A.H. Hypertension in the elderly: age- and disease-related complications and therapeutic implications. *Cardiovasc. Drug. Ther.* 1993; 7: 643-653.
14. Thulin, L.I. Age-related complications and optimal choice of artificial heart valves in elderly patients. *J. Cardiovasc. Surg.* 1991; 32: 497-502.
15. Sackett, D.L. and Torrance, G.W. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases* 1978; 7: 347-358.
16. Froberg, D.G. and Kane, R.L., Methodology for measuring health sate preferences - III: population and health effects. *Journal of Clinical Epidemiology* 1989; 42: 585-592.
17. Council of Europe Press. *Recent demographic developments in Europe 1993*. Strasbourg: Council of Europe Press, 1994.
18. Welch, H.G. Comparing apples and oranges: Does cost-effectiveness analysis deal fairly with the old and young. *The Gerontologist* 1991; 31: 332-336.
19. Schneider, E.L. Options to control the rising health care costs of older Americans *JAMA* 1989; 261: 907-908.

20. Schneider, E.L. and Guralnik, J.M. The aging of America: Impact on health care costs. *JAMA* 1990; 263: 2335-2340.
21. SIG (Stichting Informatiecentrum voor de Gezondheidszorg), at request.
22. Kawachi, I. and Malcolm, L.A., The cost-effectiveness of treating mild-to-moderate hypertension: a reappraisal. *Journal of Hypertension* 1991; 9: 199-208.
23. Katz, D.A. and Cronenwett, J.L., The cost-effectiveness of early surgery versus watchful waiting in the management of small abdominal aortic aneurysms. *Journal of Vascular Surgery* 1994; 19: 980-991.
24. Russell, J.G.B., Is screening for abdominal aortic aneurysm worthwhile? *Clinical Radiology* 1990; 41: 182-184.
25. Hogenhuis, W., Stevens, S.K., Wang, P., Wong, J.B., Manolis, A.S., Mark Estes, N.A. and Pauker, S.G., Cost-effectiveness of radiofrequency ablation compared with other strategies in Wolff-Parkinson-White syndrome. *Circulation* 1993; 88 (part 2): 437-446.
26. Desch, C.E., Hillner, B.E., Smith, T.J. and Retchin, S.M., Should the elderly receive chemotherapy for node-negative breast cancer? A cost-effectiveness analysis examining total and active life expectancy outcomes. *Journal of Clinical Oncology* 1993; 11: 777-782.
27. Schechter, C.B., Rose, D.N., Fahs, M.C. and Silver, A.L. Tuberculin screening: cost-effectiveness analysis of various testing schedules. *American Journal Of Preventive Medicines* 1990; 6: 167-175.
28. Leidl, R. Some factors to consider when using the results of economic evaluation studies at the population level. *International Journal of Technology Assessment* 1994; 10: 467-478.
29. RIVM (Rijksinstituut voor Volksgezondheid en Milieuhygiene), *Volksgezondheid Toekomst Verkenning*; de gezondheidstoestand van de Nederlandse bevolking in de periode 1950-2010. The Hague: Sdu Uitgeverij Plantijnstraat, 1993: 75-102.
30. Spilker, B. *Guide to Clinical Trials*. New York: Raven Press, 1991.
31. Baltussen, R.M.P.M., Ament, A. and Leidl, R. Making cost assessments based on RCTs more useful to decision makers. *Health Policy* 1996; 37: 163-83.
32. Eurostat. *Demographic statistics 1994*. Population and social conditions Series 3A. Brussels: Eurostat, 1994.
33. Bengtsson, H., Bergqvist, D. and Sternby, N.H., Increasing prevalence of abdominal aortic aneurysms; a necropsy study. *European Journal of Surgery* 1992; 158: 19-23.
34. Statistics Bureau, Management and Coordination Agency. *Japan Statistical Yearbook 1993/94*. Tokyo: Japan Statistical Association, 1994.
35. Bureau of Statistics, Treasury Department. *Statistical abstracts of the United States 1993, 113th edition*. Washington: Government Printing Office, 1994.
36. Verschuren, W.M.M., Smit, H.A., van Leer, E.M., Berns, M.P.H., Blokstra, A., Steenbrink-van Woerden, J.A. and Seidell, J.C. *Prevalentie van risicofactoren voor hart- en vaatziekten en veranderingen daarin in de periode 1987-1991: eindrapportage Peilstationproject Hart- en Vaatziekten 1987-1991. Rapportnummer 528901011*. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieuhygiene (RIVM), 1994 (in Dutch).
37. Canadian Coordinating Office for Health Technology Assessment. *Guidelines for economic evaluation of pharmaceuticals: Canada*. Ottawa: CCOHTA, 1994.

38. Goldman, L., Weinstein, M.C., Goldman, P.A. and Williams, L.W. Cost-effectiveness of HMG-CoA reductase inhibition for primary and secondary prevention of coronary heart disease. *Journal of the American medical association* 1991; 265: 1145-1151
39. Carr-Hill, R.A. Allocating resources to health care: is the QALY a technical solution to a political problem? *International Journal of Health Services* 1991; 21: 351-363.
40. Harris, J. QALYfying the value of life. *Journal of Medical Ethics* 1987; 13: 117-123.

CHAPTER 4

INTERPRETING POTENTIAL BIASES: COPING WITH INTERACTIONS WITH OTHER DISEASES

Forthcoming as: Baltussen R.M.P.M.¹, Reinders A.², Sprenger M.J.W.², Postma M.J.², Jager J.C.², Ament A.J.H.A.¹, Leidl R.M.³ Estimating influenza-related hospitalisation in the Netherlands. *Epidemiology and Infection*, 1998; 120

1 From the Department of Health, Organisation, Policy and Economics, Maastricht University

2 From the National Institute of Public Health and the Environment, Bilthoven

3 From the Department of Health Economics, University of Ulm, Germany

SUMMARY

This chapter analyses the consequences of disregarding a potential confounder in economic evaluation, stemming from ignoring interactions with other diseases. This potential bias is assessed in a study on influenza-related hospital admissions in the context of the cost-effectiveness of influenza vaccination. Influenza infection is associated with significant morbidity and mortality. Previous research has shown that large number of deaths in the Netherlands can be attributed to influenza. The purpose of this study is to examine the impact of influenza on hospitalisation in the Netherlands. Two methods are applied to estimate this effect: a) regression analysis and b) comparison of hospitalisation in epidemic years with non-epidemic years. Hospital discharge rates in the period 1984-1993 are considered. The study shows that, during the period studied, on average, almost 2,700 people were hospitalized for influenza per annum and that influenza was diagnosed as the main cause for hospitalisation in only a fraction of these hospitalisations (326: 12%). From an economic perspective, these results imply that the cost-effectiveness of vaccination against influenza may be severely underestimated when looking only at the changes achieved in the number of hospitalisations attributed to influenza.

INTRODUCTION

Morbidity and mortality associated with influenza have long been recognized.¹⁻⁴ The total impact of influenza encompasses not only primary (direct) morbidity and mortality but also excess (indirect) morbidity and mortality. In 1848 Farr introduced the concept of excess mortality, defining it as the number of deaths over and above the expected number for the particular season in which, and the place where, an epidemic occurred.¹

Internationally, two methodological approaches have been applied to estimate excess morbidity and excess mortality associated with influenza. The most common, the comparative approach, compares morbidity in epidemic years with that in non-epidemic years, and defines excess morbidity as the difference between those.^{2,4,5,6} For example, in the state of Oregon, U.S., for some diseases, up to 50% more hospitalisations were found in epidemic years compared with non-epidemic years.⁵

Other studies have sought to explain influenza-related mortality or morbidity by regression analyses. In 1963, Serfling derived a regression function to describe normal seasonal variations in mortality as well as time trends over the longer term.⁷ This function and its several subsequent modifications have provided the basis for estimating excess mortality since used by, among others, the WHO.⁸ For the U.K., Clifford et al⁹ estimated excess morbidity using regression analysis and found that the 1969/1970 outbreak was associated with 1.5 million excess claims for sickness. Comparable research on excess mortality in the Netherlands found that in the period 1967-1989, on average, more than 2,000 people died from influenza in the Netherlands annually. However, influenza was recognized as the primary cause of death in only a fraction (565: 27%) of these deaths.^{10,11}

The mutual application of the two approaches has not been performed before in a single study, although it allows insight into the adequacy of the respective approaches to estimate excess morbidity or excess mortality. Moreover, it provides a solid basis for interpreting the results. In this paper, both approaches are applied to estimate excess hospitalisation in the Netherlands. The hypothesis is that some hospitalisations which are attributed on hospital discharge certificates to diseases like pneumonia or chronic heart disease, are in fact caused by influenza. In both approaches, excess hospitalisation is estimated by determining the influenza-related share of monthly hospitalisations for some specific diseases.

In the regression analyses, models are developed to estimate hospitalisation for these diseases in accordance with hospitalisation observed in the period 1984-1994. The models consist of variables that describe long-term trends, monthly patterns and influenza activity. Based on the impact of the latter, the number of hospitalisations linked to influenza, but attributed to other diseases, can be estimated. In the

comparative approach, excess hospitalisation is estimated by comparing annual hospitalisations for the diseases in the period 1984-1994 with those during a year of low influenza activity, 1986-1987. Observed differences are considered as excess hospitalisation.

From an economic perspective, excess hospitalisation may have a large impact on the cost-effectiveness of vaccination against influenza. If only the changes in hospitalisations attributed to influenza are considered, the economic attractiveness of vaccination may be seriously underestimated. The results of the present study apply to the economic evaluation of influenza vaccination reported elsewhere.¹²

METHODS

The question of which share of the total number of hospitalisations in the Netherlands can be attributed to influenza is central to the analyses. In other studies, a number of diseases have been identified as important contributors to influenza-related excess morbidity and excess mortality.^{2,6,9-11} These include pneumonia (ICD-9 codes 480-486), cerebral-vascular accident (CVA, ICD-9 codes 430-438), chronic heart disease (CHD, ICD-9 codes 410-414), and diabetes mellitus (DM, ICD-9 code 250). The monthly number of hospital discharges for all these diseases, as well as for influenza (ICD-9 code 487), have been obtained for the period from January 1984 through December 1994. The data was collected and provided by the National Hospital Administration (SIG) which includes 99% of all hospital discharges in the Netherlands between 1984 and 1994.¹³⁻¹⁴

Data concerning hospital discharges have been selected as these more accurately reflect patients' diseases than do hospital admission diagnoses. However, the admission dates of all hospital discharges related to the diseases mentioned are required to consider adequately the relation between influenza activity and hospitalisation. These are determined by considering the respective hospital discharge dates and the respective lengths of hospital stay. For a detailed analysis, and to allow comparisons with research on excess mortality in the Netherlands^{10,11}, these hospitalisation data are obtained age-specifically. Four age groups are distinguished: 0-59 years, 60-69 years, 70-79 years, and 80 years and older. In the analyses, influenza years are defined from July through June of the following year as this reflects the natural course of influenza epidemics.

For illustrative purposes, Figure 1 shows influenza activity defined as the monthly number of hospital discharges for influenza (all diagnoses, all ages, as allocated to the month of hospital admission) between 1984 and 1994. On average, annually 326 persons were hospitalised with influenza diagnosed as the primary cause for hospitalisation.

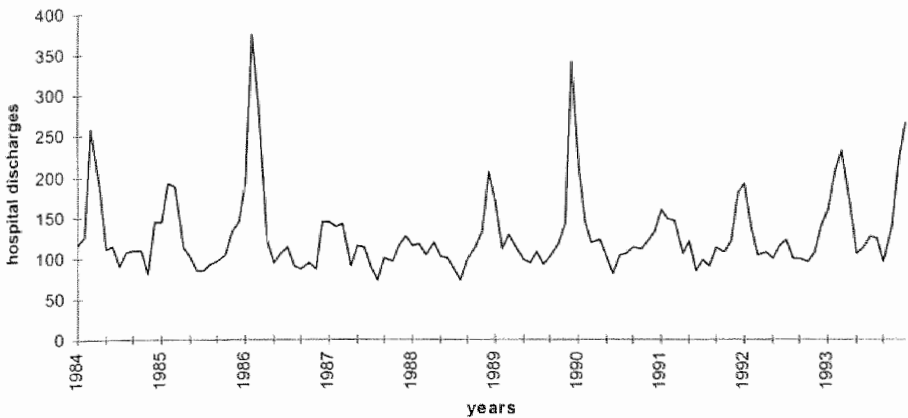


FIGURE 1. Hospitalisation for influenza (all diagnoses)

[Note: For some age-groups and diseases, December 1988 and December 1992 values are not available due to data collection problems: in those cases values are determined by interpolation. Furthermore, the presence of specialist strikes in the influenza year 1987 - 1988 has most likely influenced hospitalisation data. For this reason, this year is not considered in the definition of reference year in the regression analyses and the comparative analyses.]

Regression analysis

Regression analysis applies the technique of modeling to determine the part of the total number of hospitalisations for specific diseases that can be explained by influenza. A number of models, specific to the diseases under scrutiny and to the age groups, are constructed. Every model is based on data from January 1984 through December 1994. In every model, the monthly number of observed hospitalisations is explained by regressor variables that describe long-term trends (year variable), monthly patterns (month variable) and influenza activity (influenza indicator). Models account for long-term trends and monthly patterns to control for factors that are causally and temporally related to hospitalisation for the diseases under study but that are not linked to influenza. For example, any structural impact of temperature changes on hospitalisations is accounted for by including the month variable. Specific background information on these variables and the model in general are given in the Appendix.

A model assigns values to the parameters included; the value γ that is addressed to the variable that represents monthly influenza activity is of particular interest. This value, which is specific to disease and age group, refers to the impact of influenza on the estimated hospitalisations during the period under study, 1984-1994. This impact

in terms of hospitalisations is further estimated by a two-step approach. First, the effect of influenza is eliminated by setting the influenza activity in the model at zero, while keeping the effects of year and month the same. The model predicts in that event the number of hospitalisations if no influenza had taken place. Next, the difference between predicted hospitalisation in the situation of observed influenza activity and predicted hospitalisation in the absence of influenza activity is considered. This difference is defined as the excess hospitalisation associated with influenza (specific to disease and age group).

Various alternative models are defined. A number of alternative influenza indicators, that ideally should be sensitive and specific to influenza outbreaks, have been applied. These are: hospitalisation for influenza, primary diagnoses; hospitalisation for influenza, all diagnoses; hospital mortality due to influenza, primary diagnoses; hospital mortality due to influenza, all diagnoses. Furthermore, it is recognised in the model that hospitalisations for the diseases under scrutiny may react with a delay of one or two months following an outbreak of influenza. With this in mind, in addition to analyses ignoring the issue of time-lags, alternative analyses have been carried out including time-lags of one month and two months.

Variables are considered significant in the model when they are at least at the 5% level. The selection of alternative models (e.g. with or without time lag) is based on their explanatory power. This is indicated by R^2 , which refers to the proportion of variance in the dependent variable accounted for by the model (for more details, please see the Appendix).

Comparative analysis

In comparative analysis, hospitalisation for a number of diseases in periods with high influenza activity is compared with that in (reference) periods with low influenza activity. The hypothesis is that any detected differences in hospitalisations can be attributed to influenza. To exclude differences in hospitalisations outside influenza outbreaks, the periods usually comprise 3-4 months enveloping influenza activity in a given year.^{2,5}

In the present analysis, influenza activity is defined as hospitalisation for influenza, all diagnoses. As a reference period, we selected three adjoining months that involve the lowest influenza activity between the months of November and April in the years 1984-1994. This period, which simulates the near absence of influenza, is the period December 1986 through February 1987. Also, for each influenza year between 1984 and 1994, periods of three adjoining months are defined so that each period encompasses the highest influenza activity. Next, hospitalisations for the diseases under scrutiny in these periods are compared with those in the reference period. The difference between these hospitalisations indicates the excess hospitalisation attributed

to influenza in a given year. In the analyses, comparisons are specific to disease and age group.

This analysis focuses on differences in hospitalisation for a number of diseases between years, and attributes these differences to influenza activity. However, any differences caused by other factors (e.g. the annual increase in hospitalisations for CHD because of changes in lifestyle) may bias the estimates. Therefore, preliminary to the principal calculations, the figures regarding hospital discharges were corrected by regression analysis for trends over the years.

RESULTS

Regression analyses

The variables describing long-term trends and monthly patterns were significant at a 5% level in all models. The results are less clear with respect to the variable that describes influenza activity. Nonregarding its exact definition, this variable was significant only in the models on pneumonia (all age groups). The explanatory power of those models, as indicated by R^2 , was best when the number of hospital discharges for influenza, all diagnoses, was selected as the influenza indicator. The inclusion of time-lags did not enhance the explanatory power of the models. The value of the influenza-indicator γ in the selected models on pneumonia ranged from 0.0018 (standard error: 0.0008) to 0.0161 (standard error: 0.0032) for individuals in the age groups of respectively 0-59 years and 80 years and older.

Table 1, first column, presents the average annual number of excess hospitalisations per 100,000 persons that can be attributed to influenza, derived from the regression analyses. This is estimated as the total number of excess hospitalisations divided by the number of years considered. As noted, significant excess hospitalisation is only pertinent for the case of pneumonia (in absolute terms 2,358 hospitalisations within a population of about 15 million people). Although there were some excess hospitalisations for CVA, CHD and DM, these are not significant. Therefore, excess hospitalisation is not indicated for those diseases.

Based on these observations, further analysis is limited to pneumonia cases. Figure 2 shows that the predicted hospitalisation for pneumonia matches the observed hospitalisation fairly well. This implies that the model explains a large part of the variation in the monthly hospital discharge rates (which is also indicated by a high value of R^2 : 0.88). The figure also shows that, in general, the pattern of the predicted hospitalisation for pneumonia resembles that of influenza activity (as shown in Figure 1).

TABLE 1. Yearly excess hospitalisation per 100,000 by disease and age^a

	Regression analysis		Comparative analysis ^b
	estimate	95% confidence interval	
Pneumonia			
0-59 years	3.0 [*]	(0.3, 5.1)	1
60-69 years	27.4 [*]	(21.5, 32.9)	7
70-79 years	81.8 [*]	(68.5, 92.7)	1
80 years and older	220.2 [*]	(148.7, 272.4)	3
Cerebral-vascular accident			
0-59 years	-840.6	(-2222.3, 270.6)	471
60-69 years	-205.4	(-1109.9, 310.0)	-106
70-79 years	-204.9	(-1163.6, 832.3)	5
80 years and older	247.0	(-1802.0, 4802.4)	-132
Chronic heart disease			
0-59 years	-532.2	(-1050.4, 258.2)	-50
60-69 years	-303.2	(-833.7, 31.8)	6
70-79 years	237.0	(-314.4, 509.8)	7
80 years and older	971.0	(-913.2, 2106.6)	1
Diabetes mellitus			
0-59 years	-14.4	(-207.8, 171.2)	9
60-69 years	42.9	(-235.1, 305.1)	167
70-79 years	74.2	(-311.9, 433.3)	126
80 years and older	160.2	(-1622.1, 3896.7)	116

* significant at 5% significance level

a) denominator refers to the number of persons in the subgroup. For pneumonia, the denominator equals the total Dutch population.¹⁵ For CVA, CHD, and DM, the denominator equals the number of patients with the respective disease.¹⁶

b) Since the reference case encompasses only one year, confidence intervals are not estimated.

Excess hospitalisation is visualized in Figure 3 as the difference between predicted values including the impact of influenza, respectively assuming the absence of influenza. The figure shows that the pattern of excess hospitalisation is very similar to that of influenza activity (as shown in Figure 1), except for the year 1985/1986. In this year, influenza activity reached its highest value in February, but the predicted hospitalisations for pneumonia shows no such large peak. This can be explained by the absence of peak values of influenza activity in February in other years which causes the model to assign a relatively low value to this month variable. Consequently, this has a decreasing effect on predicted hospitalisations in these February months.

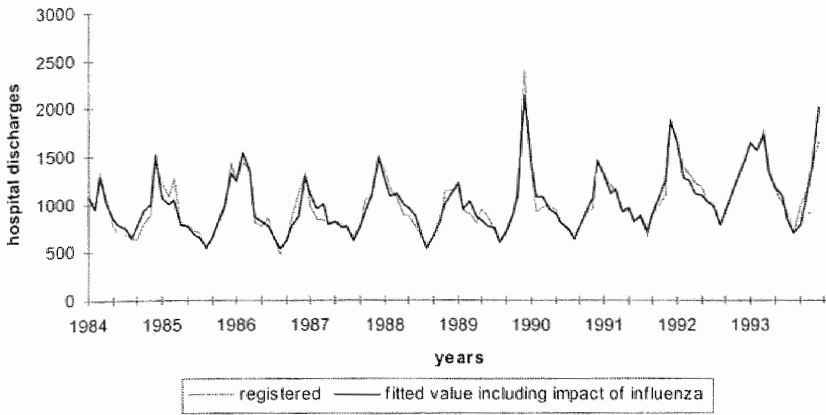


FIGURE 2. Registered and predicted hospitalisation for pneumonia, all age groups

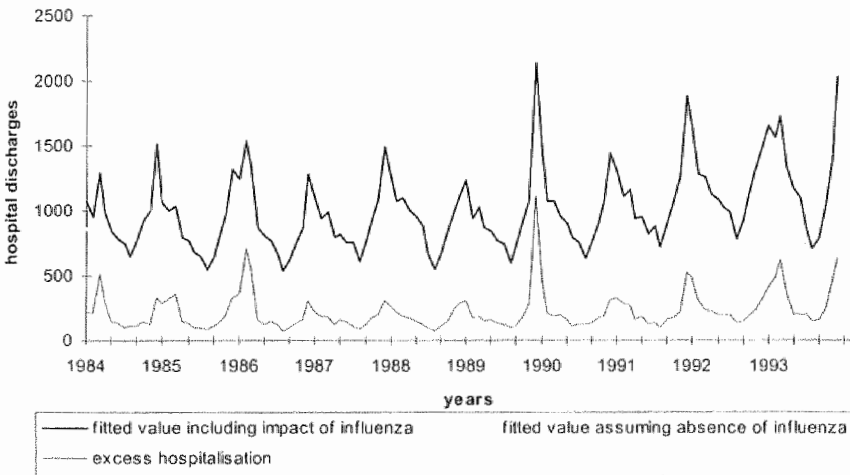


FIGURE 3. (Excess) Hospitalisation for pneumonia, all age groups

The presence of some excess hospitalisation during summer seasons can be explained by the diagnoses of some cases of influenza during these seasons (as shown in Figure 1).

Comparative analysis

As noted, regression analysis was carried out preliminary to the comparative analysis to correct for trends in the yearly number of hospitalisations that may bias the estimates. These analyses have been carried out on hospitalisation data for all

diseases and all age groups. Significant yearly trends regarding hospital discharges for all diseases and all age groups, except for CVA, were identified. These trends were corrected for with the most recent year, 1993/1994, as base year. The results of the comparative analyses are presented in the second column in Table 1. The table shows large numbers of excess hospitalisations for all age groups regarding pneumonia, and for most age groups regarding CHD and DM. For CVA, smaller and, for two out of four age groups, negative numbers of excess hospitalisations are found.

Excess hospitalisation by risk groups and age

Policy making regarding influenza vaccination is predominantly based on the classification of individuals according to age and risk status. A typical age distinction is between those below 65 years of age and those 65 years or older. Individuals are labelled as high-risk when they have one of the diseases that predispose for influenza or that aggravate the course of disease once infected. High-risk individuals are defined by the Health Council of the Netherlands as individuals having one or more chronic illnesses such as IHD, CVA, DM, other heart diseases (ICD-9 codes 415-417) or bronchitis, emphysema and asthma (ICD-9 codes 490-493).¹⁷ The allocation of the numbers of excess hospitalisations to these subgroups of individuals is relevant in context of the feasibility of influenza vaccination. The economic attractiveness of vaccination may very well depend on the extent to which it can prevent (excess) hospitalisations.

For these reasons, the results of the present analysis are also classified according to these distinctions. The classifications are based on results from the regression analysis which, for reasons explained in the discussion section, is preferred to the comparative analysis. Only significant excess hospitalisation is included. This implies that from the high-risk conditions mentioned above, no excess hospitalisation estimated is allocated to high-risk groups. A large part of the excess hospitalisation that is attributed to pneumonia should, however, be allocated to high-risk groups. From research in the US, it can be derived that, for individuals 65 years and older, 77% of all excess hospitalisation for pneumonia occurs in high-risk patient-groups. Below the age of 65 years, 93% of all excess hospitalisations for pneumonia occurs in high-risk patients.² In the present analysis, these figures have been applied to allocate the excess hospitalisation for pneumonia to high-risk and low-risk groups.

Table 2 presents the various hospitalisation rates per 100,000 individuals, as assigned to risk and age groups. The third column shows that the excess hospitalisation for pneumonia - attributed to low-risk and high-risk groups - increases with age and is more important for high-risk compared to low-risk groups.

TABLE 2. Yearly excess hospitalisation per 100,000 by risk group and age

	direct hospitalisation ^{a,b} (1)	excess hospitalisation ^{b,c} (2)	total hospitalisation (3) = (1) + (2)
low-risk			
0-64 years	0	0	0
65 years and older	2	38	40
high risk			
0-64 years	28	72	100
65 years and older	10	175	185

a) defined as the number of hospital discharges with influenza as the primary diagnosis.

b) allocation to high- and low-risk groups is based on U.S. research (2). For the age group 60-69 years, the average of the relevant figures regarding individuals younger than 65 years and those of 65 years and older is taken.

c) includes only significant numbers of excess hospitalisation.

DISCUSSION

The present study indicates that, on average, almost 2,700 people were hospitalized for influenza annually and that influenza was diagnosed as the main cause for hospitalisation in only a fraction of these hospitalisations (326: 12%). This means that a large proportion (88%) of all influenza-related hospitalisations were not recognized as such. Excess hospitalisation seems to be more relevant for the elderly than for the young and more pertinent in the high-risk population than in the low-risk population. All excess hospitalisations identified were hospitalisations diagnosed as pneumonias.

The fact that a number of diseases are not included in the present analysis has most likely caused an underestimation of the magnitude of excess hospitalisation. These diseases have been shown to account for about one-third of all influenza-associated deaths in the Netherlands.^{10,11} The impact of these diseases on excess hospitalisation is unknown and will be topic of further research. Furthermore, during the period studied, only a few moderate outbreaks of influenza have occurred in contrast to the large number of more intense epidemics prior to this period (especially from 1971-1978). The relative absence of large clusters of influenza activity in the present study makes detection of influenza-related hospitalisation more difficult.

In this study, two approaches have been addressed to estimate the magnitude of excess hospitalisation. For pneumonia, both the regression analysis and the comparative analysis yield (significant) excess hospitalisation. For the other diseases, the results are less clear. Nevertheless, the vast majority of all results stemming from the comparative analyses lie within the (in some cases very broad) 95% confidence interval as indicated by the regression analyses (although figures do not correspond

for the youngest age groups). As we see it, regression analysis is preferable to comparative analysis for estimating excess hospitalisation. In the regression approach, influenza activity is set to a level indicating the absence of influenza (keeping the effects of year and month the same).

In contrast, in the comparative approach, the reference year is not likely to be entirely free of influenza activity; this may decrease observed hospitalisation differences across diseases between this reference year and years with influenza activity. This may explain the relatively small numbers of excess hospitalisation found in the comparative approach compared with those found in the regression approach. This also implies that, in the comparative approach, results are likely to be dependent on the choice of reference year(s) and they should be interpreted with caution. This is especially the case if comparisons are limited to only a few years. In this context, we decided not to report the results of an alternative comparison of two epidemic years with one reference year (compare ²) due to the large sensitivity of the results in relation to the definition of the respective periods. Furthermore, while the main advantage of the comparative analysis seems its simplicity, regression analysis, as applied prior to the principal analysis, is still required to control for confounding trends over years.

Some issues should be considered critically when interpreting the results from regression analysis. It should be noted that excess mortality and excess morbidity are statistical concepts and cannot prove a causal relationship between influenza and non-registered influenza hospitalisations. Ideally the analysis should be carried out by distinguishing the different strains (H3N2, H1N1, B) that are mainly responsible for the different epidemics. This, however, is not feasible due to the lack of quantitative information on the causative subtype of influenza. However, the strong statistical correlation observed in the regression analysis suggests that the relation between influenza activity and non-registered hospitalisations is more than just a matter of coincidence.

A number of plausible biological relationships further support this relation. Diabetes mellitus patients are assumed to have an impaired immune response to the influenza virus¹⁸ and are especially endangered by 'Staphylococcus aureus' skin infection during influenza epidemics; this has previously been demonstrated to be a major risk factor in the development of secondary staphylococcal pneumonia.¹⁹ Furthermore, it has been suggested that certain influenza strains increase platelets' stickiness, thus making the formation of platelet aggregates more likely and resulting in the possibility of precipitating ischemic heart disease.^{20,21} Influenza has also been recognized as the most important viral infection of the respiratory tract, partly because of complications which have been shown to include exacerbations of pre-existing diseases as asthma.^{22,23}

The presence of external factors may bias the results stemming from regression analyses. For example, monthly hospitalisation rates may be influenced by external factors that limit the identification of excess hospitalisation. For example, regarding hospitalisation for CHD, we found structural decreases in hospital discharges in the months of December followed by increases in discharges in the months of January, possibly caused by the low number of working days in December or the reluctance of potential patients to be admitted to hospital during the December holidays. If structural, these differences across months are accounted for in the regression analysis by including a month variable. However, temporal events that are not caused by influenza but that do affect hospitalisation may certainly bias the estimates. Ignoring such events may cause an incorrect estimation of the role of influenza in hospitalisations. For example, some excess hospitalisation attributed to influenza may instead be caused by the respiratory syncytial (RS) virus. This virus, which also shows high activity during winter seasons, also causes respiratory infections like pneumonia. However, as the virus is mainly predominant in children, its confounding impact on this analysis will be limited.

Another important issue in regression analysis is the choice of hospital discharges, all diagnoses, as the indicator for influenza activity. Although this resulted in the best fit, it should be noted that it suggests the presence of influenza activity during summer months - which may partly be caused by over-diagnosing. This may result in an over-estimation of excess hospitalisations identified during these months.

In our analyses, as in other studies⁹⁻¹¹, it is assumed that the effect of influenza activity on hospitalisation for a certain disease, as represented by γ , is constant over different time periods. From a theoretical point of view, the value of γ should be period-specific to include changes in the prevailing types of influenza viruses. However, as noted, this is not feasible due to the lack of quantitative information on the causative subtype of influenza. Moreover, differentiating periods would considerably diminish the number of observations to be used in each analysis and hence its power to produce significant results.

The results are in accordance with those found in other research. In studies on excess hospitalisation in other countries as well, the vast majority of all identified excess hospitalisations were found to be cases registered as hospitalisations for pneumonias. In the state of Oregon, U.S., eight years (up to 4 months) with influenza activity were compared to one reference year comprising a (near) absence of influenza. Regarding the high-risk population, more hospitalisations were found for pneumonia and influenza (up to 50%) for all epidemic years considered.⁶ Another study in the state of Oregon found that hospitalisation for pneumonia and influenza in two epidemic years exceeded that in a (non-epidemic) reference year by 140-150%. Excess hospitalisation for acute cardiac failure and acute respiratory diseases other than pneumonia could not be proven.² In the Netherlands, hospitalisations of diabetes

mellitus patients because of pneumonia in epidemic years were found to exceed those in non-epidemic years by 45%-300%.²⁴

Research on excess mortality in the Netherlands found that the identified excess deaths were deaths registered as due to various disease categories like CHD (34%), COPD (17%) and other diseases (22%).^{10,11} This indicates that the nature of hospital discharges data is different from that of mortality data. During periods of influenza activity, influenza-related hospitalisations seem to be classified relatively often as pneumonias, while influenza-related deaths may more often be certified as due to a wider range of diseases.

Acknowledgement: We would like to thank N.J.D. Nagelkerke and C. de Lezenne Coulander for their statistical support for this study.

Appendix

The analyses have been carried out by using the procedure GENMOD of the statistical package SAS 6.03.²⁵ The GENMOD procedure fits generalized linear models.²⁶ The class of generalized linear models is an extension of traditional linear models which allows the mean of a population to depend on a linear predictor through a nonlinear link function and allows the response probability distribution to be any member of an exponential family of distributions. In this study a log-linear link function is preferred to a traditional linear model because of the nature of hospital discharge data. While predictors in a traditional linear model can take on any value, a log-linear function does not allow the predicted hospital discharge values to take on negative values. Furthermore, in this analysis, as hospital discharge data is of discrete nature, a Poisson distribution is assumed.

This implies that the number of observed monthly discharge data is assumed to be Poisson distributed random variable with mean and variance equal to a parameter λ specified as

$$\log \lambda_i = \log N_i + \left(\sum_{j=1}^{12} \alpha_j \cdot M_j + \sum_{k=2}^{10} \beta_k \cdot J_k + \gamma \cdot F_i \right)$$

or equivalently:

$$\lambda_i = N_i \cdot \exp \left(\sum_{j=1}^{12} \alpha_j \cdot M_j + \sum_{k=2}^{10} \beta_k \cdot J_k + \gamma \cdot F_i \right)$$

i = 1...120 (monthly figures)

N_i = size of the considered population in month i

M_j = 1 for calendar month j

= 0 elsewhere

- $j = 1, \dots, 12$ (July - June)
 $J_k = 1$ for the k -th year considered
 $= 0$ elsewhere
 $k = 1, \dots, 10$
 $F_i =$ influenza-activity indicator in month i

The coefficients α_j , β_k , and γ have to be estimated. The coefficient γ represents the effects of influenza activity on hospitalisation for a certain disease. The quantity $1 - \exp(-\gamma F_i)$ represents the excess hospitalisation in month i as a proportion of λ_i . The presence of excess hospitalisation is tested for by applying a two-tailed student t -test at a 5% significance level. The null-hypothesis is defined as $H_0: (1 - \exp(-\gamma F_i)) \lambda_i = 0$.

The decrease in deviance indicates the capacity of the model to explain the variation of the monthly hospital discharge rates. The figure R^2 , which indicates the 'goodness of fit' of the model is estimated as:

$$R^2 = 1 - \exp [\{ \text{deviance (fitted model)} - \text{deviance (unfitted model)} \} / n]^{27}$$

It is assumed that the monthly observed hospital discharges are mutually independent, given the explanatory variables year, month and influenza activity. To control for over-dispersion, the scale parameter was defined as the deviance divided by the number of degrees of freedom.

REFERENCES

1. Farr, W. In: *Vital Statistics*. London, Office of the Sanitary Institute 1885; 330-3.
2. Barker W.H., and Mullooly J.P., Impact of epidemic type A influenza in a defined adult population. *American Journal of Epidemiology* 1980; 112: 798-811
3. Schoenbaum S.C. Economic impact of influenza: the individualistic perspective. *The American Journal of Medicine* 1987; 82: 26-30
4. McBean A.M., Babish J.D., Warren J.L. and Meldon E.A. The effect of influenza epidemics on the hospitalisation of persons 65 years and older. In: *Options for the control of influenza II*. Hannoun C. et al. Excerpta Medica, Amsterdam Elsevier Science Publishers, Amsterdam, The Netherlands 1993
5. Mullooly J.P., Bennett, M.J., Hornbrook M.C., Barker W.H., Williams W.W., Patriarca P.A. and Rhodes P.H. Influenza vaccination programs for elderly persons: cost-effectiveness in a health maintenance organisation. *Annals of Internal Medicine* 1994; 121: 947-952
6. Fleming D.M., The impact of three influenza epidemics on primary care in England and Wales. *Pharmacoeconomics* 1996; 9: 38-45
7. Serfling R.E. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Report* 1963; 78: 494-506
8. Assaad I., Cockburn W.C., Sundaresan T.K. Use of excess mortality from respiratory disease in the study of influenza. *Bulletin WHO* 1973; 49: 219-233

9. Clifford R.E., Smith J.W.G., Tillett H.E., and Wherry P.J. Excess mortality associated with influenza in England and Wales. *International Journal of Epidemiology* 1977; 6: 115-128
10. Sprenger M.J.W., Beyer W.E.P., Kempen B.M., Mulder P.G.H. and Masurel N. Risk factors for influenza mortality? in: *Options for the control of influenza II*; Hannoun C. et al. Excerpta Medica, Amsterdam Elsevier Science Publishers, Amsterdam The Netherlands 1993
11. Sprenger M.J.W., Mulder P.G.H., Beyer, W.E.P., Strik R. van, and Masurel N. Impact of influenza on mortality in relation to age and underlying disease, 1967-1989. *International Journal of Epidemiology* 1993; 22: 333-339
12. Reinders A., Postma M.J., Govaert T.M., Sprenger M.J.W. Kosten-effectiviteit van influenza vaccinatie in Nederland (Cost-effectiveness of influenza vaccine in The Netherlands). *Nederlands Tijdschrift voor Geneeskunde* 1997; 141: 93-97
13. Stichting Informatievoorziening Gezondheidszorg (SIG). Data on patients, discharged from hospital in the period 1984-1994 with CVA, CHD and DM as primary diagnosis, and with pneumonia and influenza as primary and secondary diagnosis (data tape). Utrecht: SIG, 1996
14. Stichting Informatievoorziening Gezondheidszorg (SIG). *Toelichting klinische data* (specification hospital data). Utrecht: SIG, 1997
15. Centraal Bureau voor de Statistiek (CBS, Central Bureau of Statistics). *Age composition of the Netherlands 1984-1994*. Maandstatistiek bevolking. Voorburg: Hoofdafdeling bevolkingsstatistiek, 1984-1994
16. Stuurgroep Toekomstscenario's Gezondheidszorg (STG, Steering Committee on Future Scenarios in Health Care). *Scenario-onderzoek chronische ziektes* (Scenario-analysis chronic diseases). Rijswijk: STG, 1991
17. Gezondheidsraad (Health Council of the Netherlands): committee on influenza vaccination. *Influenza vaccination season 96-97*, publication no. 1996/13. Rijswijk: Health Council of the Netherlands, 1996
18. Diepersloot R.J.A., Bouter K.P., Hoekstra J.B.L. and Masurel N. Humoral immune response and delayed type hypersensitivity to influenza vaccine in patients with diabetes mellitus. *Diabetologia* 1987; 30: 397-401
19. Gosling W.R.O., Mulder J., Djajadiningrat J. and Masurel N. Staphylococcal pneumonia in influenza. Relation to antecedent staphylococcal skin infection. *Lancet* 1959; 2: 428-430
20. Bainton D., Jones G.R. and Hole D. Influenza and ischemic heart disease - as possible trigger for acute myocardial infarction? *International Journal of Epidemiology* 1978; 7: 231-239
21. Spodick D.H., Inflammation and the onset of myocardial infarction. *Annals of Internal Medicine* 1985; 102: 699-701
22. Connolly A.M., Salmon R.L., Lervy B. and Williams D.H. What are the complications of influenza and can they be prevented? Experience from the 1989 epidemic of H3N2 influenza A in general practice *BMJ* 1993; 306: 1452-1454
23. Nicholson K.G., Kent J. and Ireland D.C. Respiratory viruses and exacerbations of asthma in adults. *BMJ* 1993; 307: 982-986
24. Bouter K.P., Diepersloot R.J.A., Romunde L.K.J. van, Uitslager R., Masurel N., Hoekstra J.B.L., Erkelens D.W. Effect of epidemic influenza on ketoacidosis,

- pneumonia and death in diabetes mellitus: a hospital register survey of 1976-1979 in The Netherlands. *Diabetes Research and Clinical Practice* 1991; 12: 61-68
25. SAS/STAT Software: *The GENMOD Procedure*, Release 6.09. SAS Institute Inc., US, 1993
 26. Nelder J.A. and Wedderburn R.W.M. Generalized Linear Models. *Journal of the Royal Statistical Society* 1972; 135: 370-384
 27. Nagelkerke N.J.D. *Maximum likelihood estimation of functional relationships*. Berlin: Springer 1992

CHAPTER 5

EVALUATING THE TUMT-TURP TRIAL AT THE LEVEL OF THE DUTCH POPULATION

Published as: Baltussen R.M.P.M.¹, Wielink G.², Stoevelaar H.J.³, Van der Wilt G.J.², Severens J.L.², Ament A.J.H.A.¹ The economic impact of introducing TUMT in the treatment of BPH: A scenario analysis. *World Journal of Urology*, 1998; 16: 142-147

¹ From the Department of Health, Organisation, Policy and Economics, Maastricht University

² From the Department of Medical Informatics, Epidemiology, and Statistics, University of Nijmegen

³ From the Institute for Health Care Policy and Management, Erasmus University, Rotterdam

SUMMARY

This chapter reports on a study on the economic impact of the introduction of transurethral microwave therapy (TUMT) in the treatment of benign prostate hyperplasia (BPH). In the study, an modeling approach is developed to enhance the informative value of RCT-based economic analyses to policy makers. The approach combines results from an RCT-based economic analysis on TUMT with observational data to get a more realistic real world view of the economics aspects following the introduction of TUMT. Furthermore, analyses at the system level are performed to assess its budget impact. Several scenario's are constructed, including a baseline scenario, demand scenarios reflecting the number of men that will be treated by TUMT following its introduction, and supply scenarios reflecting the number of hospitals that will provide TUMT following its introduction. The analysis leads to important conclusions on the overall impact of TUMT. It reveals that the organisation of providing TUMT hardly affects costs at the system level, and also makes clear that the most important determinant of the total costs of BPH-treatment is an increased demand for BPH-treatment following the introduction of TUMT. The study shows that modeling techniques, when applied in addition to RCT-based economic evaluation, can be a useful tool to enhance the real world relevance of results from RCT-based economic analyses. Furthermore, additional modeling proves to be effective in transferring results from the individual level to the aggregate level.

INTRODUCTION

In economic evaluation, randomised clinical trials (RCTs) are often regarded as the optimal vehicle for estimating costs related to the introduction of a medical intervention. However, this belief is beginning to be countered by criticism regarding the limited real world relevance of their results. Moreover, traditional RCT-based economic evaluations are criticised because they do not report on the impact of adopting a medical intervention on the aggregate level.¹

This study combines results from an RCT with observational data to get a more realistic view of the economic impact of the introduction of a new minimally invasive therapy in the treatment of benign prostate hyperplasia (BPH): transurethral microwave therapy (TUMT). The concept of TUMT is to create high temperatures deep inside the prostatic lobes while preserving the urethral surface from the worst effects of heat, by using conductive cooling from water circulating within the applicator.² In patients with relatively moderate symptoms and mild obstruction, a significant improvement may follow from the use of low-energy thermotherapy. The WHO mention in their recommendations that in such patients, TUMT offers a suitable alternative to transurethral resectioning of the prostate (TURP) because of its minimally invasive character.³ High-energy TUMT is likely to be more effective for severely obstructed patients, although its precise value has not yet been determined.

Recently, an RCT is performed comparing the cost-effectiveness of high-energy TUMT with that of the most common surgical treatment for BPH, TURP. The two-year follow-up results show that high-energy TUMT results in durable outcome in 93% of patients, with an objective improvement of 42% and subjective improvement of 65%.⁴ Furthermore, the use of TUMT generates large cost-savings: a TUMT is typically performed on an outpatient basis whereas treatment with TURP typically encompasses a hospital stay of about 7 days.^{4,5}

However, the results of this RCT-based economic analysis may not be entirely adequate for policy making. Costs measured in the RCT may not reflect the costs that can be expected when TUMT is adopted in daily practice. In addition, it is uncertain which and how many patients will be eligible for TUMT in the future, and which treatments will be replaced by TUMT. Another important issue is that many men in the community who are symptomatic and who currently reject surgical treatment might consider treatment with TUMT because it is relatively non-invasive. This may lead to considerable increases in the number of BPH-treatments and related costs. Therefore, scenario analysis is conducted to explore possible diffusion paths related to the introduction of TUMT, and associated health care costs.

METHODS

In order to assess cost changes resulting from the introduction of TUMT, costs of current BPH treatments are compared with those in the hypothetical introduction of TUMT. Costs of current BPH treatments are estimated based on the diagnostic process, hospital stay (if any), and follow-up related to the various treatments. Estimating total costs related to the hypothetical situation of adopting TUMT requires a two-step approach. First, the real world relevance of the RCT results must be carefully examined. Second, the epidemiology of BPH must be assessed to arrive at an indication of the numbers of patients who will be treated with TUMT. Various scenarios are constructed to explore possible paths to the demand for TUMT and to the provision of TUMT.

The analysis does not distinguish between high-energy TUMT and low-energy TUMT. Furthermore, the analyses focus only on the most important determinants of costs in the treatment of BPH: TURP, open prostatectomy, medication (α -blockers and Finasteride), and 'no active intervention'. Other therapies, such as laser, are not considered due to their infrequent use.

Costs of current treatments

Estimates concerning the number and type of BPH treatments performed in the Netherlands in 1996 are based on National Hospital Administration figures.⁶ This database covers more than 99% of all hospital discharges in the Netherlands. All hospitalisations with BPH as a primary diagnosis are included (ICD-9 code 600), as well as those with BPH as secondary diagnosis and for whom a surgical treatment related to BPH is performed. The data show that in 1996 almost 13,000 patients were hospitalised for BPH or received a surgical treatment for BPH. Figures for 1998, as adjusted for demographic trends, are shown in Table 1.

In combination with these hospitalisation figures, data from an observational study have been applied to estimate the number of non-surgical BPH treatments in the Netherlands. This observational study of BPH patients includes 670 patients newly referred to a urologist in a sample of 13 hospitals.⁷ Additional analyses show the age-specific distribution of BPH-patients among surgical and other treatments.⁸ Of all BPH patients who have been newly referred to a urologist, 35% receive surgical procedures (TURP, open prostatectomies, or, in a few cases, other procedures such as urethrotomy). About 23% of all BPH patients are treated with medication. About 36% of all BPH patients who have been newly referred to a urologist do not undergo active intervention. These are patients who are kept on a watchful waiting regimen or who require no urological follow-up. Approximately 5% receive a therapy other than those mentioned above.⁸ Table 1 shows the corresponding figures.

TABLE 1. Annual BPH-treatments and related costs, 1998 estimates

	Therapy choice		Costs			
	% ^a	absolute ^b	clinical course of disease ^c	hospital nursing costs ^d	therapy costs ^e	total costs per patient
Surgery						
TURP	-	12,472	1,800	5,717	1,446	8,963
OP	-	1,094	1,650	12,290	4,747	18,687
Total	35	13,566	-	-	-	-
Medication	23	9,006	1,566	-	553	2,119
Watchful waiting	36	13,789	1,100	-	185	1,285
Other treatment	6	2,539	-	-	-	-
Total	100	38,900	-	-	-	-

^a = not applicable

a) Percentages are based on observational data.⁶

b) Data regarding surgical procedures are based on hospitalisation data.⁶ Data regarding medication, watchful waiting, and other therapies are extrapolations of hospitalisation data on patients receiving surgical BPH treatment, based on the distribution of BPH-treatments as retrieved from observational data.⁸

c) Volume estimates are based on observational data.⁸ Cost estimates are based on retail prices and charges^{10,11}.

d) Based on hospitalisation data.⁶ Estimates on daily hospital costs are based on information on hospital costs in the Netherlands.⁹ Presented figures are weighed costs across all groups.

e) Cost data are based on hospital and specialist charges.^{12,13}

Hospital stay costs are estimated by considering the average length of stay (ALOS) and the daily costs of hospital nursing. Age-specific ALOS data (categories 50-59 years, 60-69 years, and 70 years and older) for 1996 for the various treatments are obtained from the National Hospital Administration.⁶ The ALOS of hospitalisations with secondary diagnosis BPH and where a BPH-related surgical treatment is performed is replaced by the ALOS of hospitalisations with primary diagnosis BPH and similar surgical treatment. This is done to compensate for the possible impact of a primary diagnosis in the former type of hospitalisations requiring an extensive hospital stay that is not attributable to BPH. Daily hospital nursing costs are estimated by dividing total hospital nursing costs for all teaching and non-teaching hospitals in 1995 by the total number of hospital days.⁹ To approximate the cost of hospital days in the years after 1995, the cost data are corrected with an annual 6% rate (this equals the annual increase in average daily hospital costs in the period 1991-1995).⁹

Data regarding the diagnostic process of the various treatments are estimated based on the observational study mentioned above.⁸ Cost estimates are based on charges.^{10,11} Data on all investigations, treatments or procedures, including re-

admissions, carried out within 1 year after the initial treatment choice are based on preliminary results of the 1 year follow-up of the above mentioned observational study (including 89% of the 670 patients).⁸ Cost estimates are based on charges.¹⁰⁻¹³ It is assumed that hospital re-admissions for surgical interventions involve (the costs of) a urethrotomy encompassing three days of hospitalisation. A urethrotomy is an incision in the urethra to widen its size and is often applied following failure of a therapy. Hospital re-admissions without surgical intervention are also assumed to involve (the costs of) three days of hospitalisation. Furthermore, the TURP retreatment rate is assumed to be 10%.¹⁴ Hospital nursing costs and the costs of investigations, treatments and respective treatment procedures are provided in Table 1.

TUMT-TURP trial

The costs of TUMT, as observed in the RCT of high-energy TUMT versus TURP, are presented in Table 2. These data are based on an inclusion of 301 patients within a two-year follow-up.^{4,5} Equipment costs include that of the applicator (in the RCT, the Prostatron 2.5 is applied) and an echo. Purchase prices in 1995 were NLG 800,000 and NLG 132,000 respectively. Related annual costs include annuity costs (5% of purchase price, annually over a period of 10 years) and equipment maintenance (8% of purchase price annually). Variable treatment costs include costs of personnel and of materials. Average follow-up costs include costs of consultations as well as costs of retreatment. In total, average costs per patient treated are NLG 4,813.⁵

Costs, as observed in the RCT, are translated to costs that can be expected in a real world environment. First, to account for the artificiality of the RCT design, the diagnostic process noted in the observational study mentioned above is substituted for the diagnostic process in the RCT.⁸ Second, the current applicator price of NLG 600,000 is substituted for the costs of the applicator. Third, the utilisation of the applicator as observed in the trial (200 patients per annum) is allowed to vary up to 800 patients (100% capacity utilisation). In the baseline analysis, an 80% capacity utilisation is assumed. Finally, national statistics on daily hospital costs are applied to the follow-up costs, to substitute for the RCT specific cost-estimates.⁹

Sensitivity analysis is applied, to reflect the uncertainty associated with the cost-estimates related to TUMT. Minimal cost estimates assume the costs of the diagnostic process, the follow-up and the applicator to be 25% less than the baseline estimate. A 100% utilisation of the applicator is assumed. Maximum cost estimates assume the costs of the diagnostic process, the follow-up and the applicator to be 25% more than the baseline estimate. The utilisation of the applicator is assumed to be only 60%. The costs of TUMT, adjusted to what can be expected once TUMT is adopted in daily practice, are shown in Table 2.

TABLE 2. Costs of TUMT, 1998 prices

	Costs diagnostic process ^a	Treatment costs ^a					Costs follow up ^a	Total costs per patient		
		(1)	applicator purchase costs ^c (2)	annual applicator costs ^e (3)	annual number of patients (4)	annual applicator costs per patient (5)=(3)/(4)			variable treatment costs per patient (6)	annual costs per patient (7)=(5)+(6)
RCT	1,244 ^b		1,078,324	226,448	200	1,132	508	1,640	1,929	4,813
Daily practice	855 ^c		732,000	153,720	640 ^f	240	508	749	1,869	3,473

a) Cost estimates are based on retail prices and charges.^{10,11}b) Volume estimates are based on RCT data.^{4,5}c) Volume estimates are based on observational data.⁸

d) Including costs of Prostatron 2.5 and echo.

e) These include annual annuity costs (period of 10 years at 5%) and annual equipment maintenance costs (8% of purchase price).

f) Assumed capacity utilisation of 80%; utilisation may be less depending on the total number of patients eligible for TUMT as defined in demand scenarios.

Indication criteria of TUMT

The indication criteria of TUMT determine its role in the treatment of BPH and are approximated by the inclusion and exclusion criteria as applied in the RCT. The exceptions are patients with diabetes mellitus; these patients are excluded from enrollment in the trial because of the presence of co-morbidity which will limit effect detection, but they are nevertheless considered eligible for treatment in daily practice.¹⁵ The indication-criteria of TUMT are: a moderate to severe symptom score (WHOPPS ≥ 12), a decreased uroflow (Qmax < 15 ml/s) and an increased size of the prostate (≥ 25 cc). The exclusion criteria include the presence of a bladder stone, a large middle-lobe, or a prostate carcinoma. To quantify the role of TUMT in the treatment of BPH, these criteria are applied to observational data on the characteristics of patients currently being treated for BPH.^{8,15} It is assumed that the data, which stem from elective patients, are representative for acute patients too. (Missing values on e.g. uroflow-scores are assumed to be distributed equally among all BPH patients). The data on the presence of an enlarged middle lobe and a prostate carcinoma are based only on patients receiving surgical therapies; it is assumed that these are representative for all BPH-patients. Analysis shows that, depending on age, about 31- 43% of all patients currently receiving treatment for BPH meet the indication-criteria for TUMT and are therefore considered eligible for treatment with TUMT.

SCENARIO ANALYSIS

The study distinguishes between several scenarios. These include the baseline scenario (reflecting the situation that TUMT will not be introduced), demand scenarios (reflecting the number of patients who will be treated by TUMT when it is introduced) and supply scenarios (reflecting the number of hospitals that will provide TUMT when it is introduced)(Table 3). In all scenarios, the diffusion of TUMT starts in 1998, and costs are determined for a period of five years. This is chosen to reflect the period in which planning of health care facilities can be realised with a certain degree of certainty. All scenarios entail demographic developments - that is, for the Netherlands, an increase in the number of older men in the years to come.

Baseline scenario. The baseline scenario extrapolates current treatment patterns for BPH. The scenario entails only demographic developments.

Demand scenarios. Two scenarios are distinguished to assess the demand for TUMT in the future. The first scenario assumes that the demand for TUMT will come only from patients who are currently being treated for BPH and who meet the indication-criteria for TUMT. That is, the total number of patients treated for BPH will not increase following the introduction of TUMT (except for demographic reasons). This scenario is labelled 'constant demand'. Based on the indication-criteria for TUMT

and on the respective characteristics of patients receiving BPH-treatments, the potential for substituting TUMT for these treatments is estimated. Table 3 shows there is potential for performing almost 13,500 TUMTs in the year 2002.

The second scenario acknowledges that the demand for TUMT may also include individuals who are not currently seeking BPH-treatment but who might if the introduction of TUMT decreases the barrier to treatment. This scenario is labelled 'increased demand'. The prevalence of BPH in the community is estimated as the starting point in this scenario. This prevalence has been reported to be 9-27% in all men between the age of 55 years and 74 years in the Netherlands, defined according to generally accepted criteria (prostate volume > 30 cm³, maximum flow rate > 15 ml/s and IPSS > 7).¹⁶ In absolute figures, this means a prevalence of approximately 360,000 men in 1998 (assuming that the prevalence of BPH among men aged between 50 and 54 years is equal to those aged between 55 and 59 years, and that the prevalence of BPH among men older than 75 years is equal to those aged between 70 and 74 years).¹⁷

However, it is not evident whether or not these men are sufficiently bothered by their symptoms to seek treatment. Of all men with BPH, 51% report that their urinary symptoms interfere with at least one activity of daily life.¹⁸ In other research, it was found that of all men with BPH-symptoms, only about a quarter perceive their urinary symptoms to be a problem.¹⁹ In the scenario analyses, the number of men who seek treatment varies between the current situation (~ 36,000: 10% of all symptomatic men in 1998 [Table 1 excluding 'other treatments']), and the situation where all men finding their symptoms problematic present themselves for treatment (~ 90,000: 25% of all symptomatic men in 1998 [excluding 'other treatments']). It is assumed that the increased number of men seeking treatment following the introduction of TUMT have the same characteristics as the population who are currently receiving treatment for BPH and will subsequently undergo the same array of treatments.

Supply scenarios. TUMT may be provided in alternative ways as reflected in a number of scenarios. In all scenarios, it is assumed that all BPH patients who present themselves for treatment and who are eligible for TUMT are indeed treated. The first supply scenario assumes a high capacity utilisation (80%) of the applicators. For example, this can be fostered by the strong influence of the Dutch government on the allocation of TUMT-centres to hospitals. A limited number of hospitals cover the total demand for TUMT as defined in the demand scenarios. This scenario is labelled 'controlled provision'. A second scenario assumes a sub-optimal capacity utilisation of the applicators. This may resemble the situation if the Dutch government leaves the decision to provide TUMT to the individual hospitals. This may lead to a larger number of hospitals adopting TUMT and consequently to a relatively low average capacity utilisation. This scenario is labelled 'widespread provision'.

TABLE 3. Scenario-analyses on demand for BPH-treatments, 2002 figures

Treatment	Scenario		
	'baseline'	'constant demand' ^{b,c}	'increased demand' ^{b,d}
Surgery			
TUMT	0	13,453	32,730
TURP	13,625	9,182	22,372
OP	1,189	397	1,191
Medication	9,876	5,229	12,667
Watchful waiting	15,138	11,944	28,962
Total	39,828 ^a	40,205 ^a	97,922

a) Figures do not correspond as the total number of retreatments increases after the introduction of TUMT.

b) Assuming a 100% diffusion within one year.

c) Estimates are based on the current distribution of therapies across various therapies and age groups (Table 2), and that of indication criteria for TUMT.⁸

d) Estimates are based on the prevalence figures of BPH in the community, assuming that 25% demand treatment for BPH following the introduction of TUMT.^{16,19}

RESULTS

The (cost-)consequences of the various scenarios are presented as estimates for the year 2002 (Table 4). In the baseline scenario, costs of BPH-treatment equal ~ NLG 203 mln. In the 'constant demand' scenario - assuming that the demand for BPH treatment does not increase following the adoption of TUMT -, about 13,500 TUMTs will be carried out. If the provision of TUMT is controlled, to foster an optimal capacity utilisation, treatment will be provided in 22 centers, inducing total costs of NLG 187 mln. Sensitivity analyses show minimum and maximum values of NLG 172 and NLG 204 mln.

If there is widespread provision of TUMT, through, for example, 50 hospitals, total costs equal NLG 189 mln. Minimum and maximum values are NLG 174 mln and NLG 204 mln. If demand increases up to 25% following the introduction of TUMT, approximately 32,700 TUMTs will be performed. In the case of a controlled provision, treatment will take place through 52 hospitals. In this scenario, total costs of BPH treatment are NLG 457 mln. Sensitivity analyses show minimum and maximum values of NLG 419 mln and NLG 497 mln. If provision is widespread, (i.e. through 100 hospitals), total costs equal NLG 466. Minimum and maximum values are NLG 427 mln and NLG 505 mln.

TABLE 4. Scenario analysis, costs in 2002 (x million NLG), all age groups^a

Scenarios	'baseline'	'controlled provision' ^c	'wide-spread provision' ^d
'baseline'	203	-	-
'constant demand'	-	187 (172 - 204)	189 (174 - 204)
'increased demand' ^b	-	457 (419 - 497)	466 (427 - 505)

a) Figures between brackets represent minimum and maximum values stemming from sensitivity analyses.

b) Assuming that 25% of all symptomatic men demand treatment for BPH following the introduction of TUMT.

c) The scenario 'controlled provision' assumes an optimal capacity utilisation.

d) The scenario 'widespread provision' assumes a sub-optimal capacity utilisation.

Figure 1 shows the effects of increasing demand for BPH-treatment in more detail. In case the percentage of symptomatic men who receive BPH treatment increases from the current 10.2% to e.g. 12% following the adoption of TUMT, total costs in 2002 equal ~ NLG 219 mln (applying the 'controlled provision' scenario, baseline estimate). Further increases to 20% or 25% leads to increases in costs up to ~ NLG 365 mln and ~ NLG 457 mln respectively.

DISCUSSION

The analysis shows that potentially large cost savings, up to NLG 16 mln annually, may be associated with the introduction of TUMT in the Netherlands. The number of men being treated is an important determinant of total BPH treatment costs. Costs may increase by ~ NLG 263 mln if 25% of all symptomatic men seek treatment following the introduction of TUMT. The introduction of TUMT is cost-neutral if the percentage of symptomatic men seeking treatment increases from the current 10.2% up to between 11% and 12%. Such an increase may be induced by an enhanced prostate awareness among men, or by the increased attention that industry is paying to this health care problem. However, there is not sufficient information to sustain or reject the statement that the demand for BPH treatment is increasing.

Analysis also shows that the way the provision of TUMT is organised hardly affects the costs of treatment. Although large (investment) costs are associated with TUMT, its large treatment capacity causes these costs per patient to remain relatively low. There are small efficiency gains if TUMT is provided through a controlled supply, but these may be offset by higher travel costs of patients to the hospitals. Another option is to provide TUMT on a mobile basis. Although this will result in a relatively high capacity utilisation, efficiency gains are expected to be minimal because of the small impact of capacity utilisation on total costs.

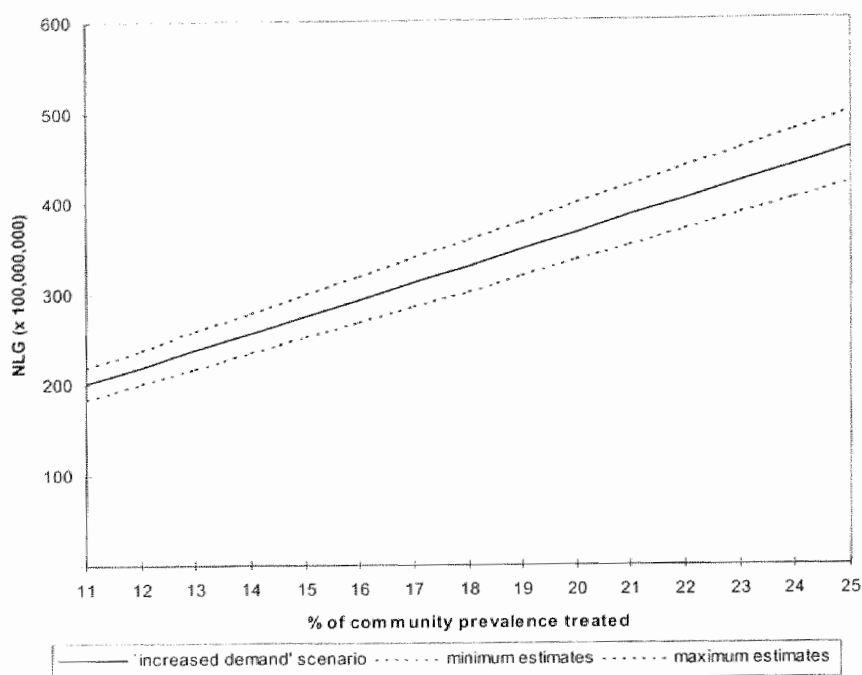


FIGURE 1. Costs related to an increase in men seeking BPH treatment ('increased demand' scenario)

The analysis focuses only on cost-consequences stemming from the introduction of TUMT, and considers the efficacy of high-energy TUMT to be equal to that of TURP. Randomised studies show that high-energy TUMT exerts effects in reducing symptoms and increasing flow rate that are greater than can now be achieved by drug therapy but somewhat less than those commonly seen following TURP. On the other hand, the less invasive character of TUMT compared to TURP is likely to cause less complications. The WHO notes that further evaluation is required before the final place of TUMT is defined.³ Currently, for example, follow-up data on retreatment rates are available only for a period of two years. If these rates increase significantly when measured over a longer time period, total costs related to TUMT may be much larger than estimated in the present analysis. The results of this study must be considered in light of these shortcomings.

There are a number of assumptions in this analysis that should be considered carefully. First, the analysis assumes that all BPH-patients, as considered in the analysis and who meet the indication criteria for TUMT, are also willing to undergo treatment with TUMT. However, some may decline treatment with TUMT because they are willing to live with their symptoms, or are satisfied with their current treatment. Second, the analysis does not distinguish between high-energy TUMT and low-energy TUMT. However, as high-energy TUMT involves higher temperatures,

its clinical course after treatment can be considered somewhat more severe. As this may lead to higher costs, the assumption may cause an overestimation of the costs related to TUMT. Third, the analysis assumes that the additional men who seek treatment following the introduction of TUMT have the same characteristics as the population who currently receive treatment for BPH. This seems plausible recognising that these additional men will present themselves in an earlier stadium of disease. However, the additional group will most likely not be comprised of acute retention patients. Finally, cost estimates in the analysis are based on charges which may bias the analysis in the sense that charges do not necessarily reflect real costs.²⁰

A number of possible future developments which are not included in this analysis may affect the total costs of BPH-treatment in the Netherlands. Further technological changes related to TUMT may cause its costs to decrease. However, the adoption of different applicators should be critically assessed before being adopted on a widespread basis as effectiveness may vary. Also, in the future, indication criteria for TUMT may expand to less severe BPH patient groups. From an economic perspective, this development should be regarded with caution as it is not certain that the benefits of applying TUMT to these groups would justify the extra costs. Furthermore, besides TUMT, various other minimally invasive techniques, such as laser therapy, are being researched.³ Depending on the indication criteria of these techniques, their diffusion may further displace traditional treatments or may affect the diffusion process of TUMT as defined in this study. As most of these developing techniques are minimally invasive, they are not expected to have a major impact on the overall future costs of BPH treatment as estimated in the present study. Finally, demographic developments lead to a dramatic increase in the number of older men in the future. In the year 2010, the number of men over 50 years of age is predicted to be 21% larger than in 2000. In 2020, this increase is predicted to be 46%.¹⁷ Persons over 50 years of age constitute the most rapidly growing segment of the Dutch population, making BPH a major health expense.

It has been shown that the introduction of TUMT seems a cost-saving activity, both on the individual (patient) level, as well as on the aggregate level. Furthermore, the present analysis has revealed that the organisation of providing TUMT hardly affects the costs at the system level, thereby indicating that there is no obvious need for regulation of TUMT by the government. The analysis has also made clear that the most important determinant of the total costs of BPH-treatment is the possible increase in the total demand for BPH-treatment that may be induced by the introduction of TUMT. These data may facilitate the decision making process regarding the introduction of TUMT.

REFERENCES

1. Baltussen R.M.P.M., Ament A.J.H.A., Leidl R.M. Making cost assessments based on RCTs more useful to decision makers. *Health Policy* 1996; 37:163-183
2. De la Rosette J.J.M.C.H., Tubaro A., Hofner K., Carter S.S.C., Transurethral microwave thermotherapy: past, present and future. *World J Urol* 1994;12:352-356
3. Smith P. Other non-medical therapies. In: Cockett A.T.K., Khoury F., Aso J.Y., Chatelain C., Denis L., Griffith K., Murphy G.(eds) *The 2nd international consultation on benign prostatic hyperplasia (BPH)*, Monaco. Scientific Communication Institute Ltd, Channel Islands, 1995: 575-600
4. d' Ancona F.C.H., Francisca E.A.E., Debruyne F.M.J.H., De la Rosette J.J.M.C.H. High-energy transurethral microwave thermotherapy in men with lower urinary tract symptoms. *J Endour* 1997;11: 285-9
5. Wielink G. Unpublished data
6. National Medical Registration 1996. SIG Zorginformatie: data tape. SIG Cooperatie UA, Utrecht, 1997
7. Stoevelaar H.J., Van de Beek C., Casparie A.F., Nijs H.G.T., McDonell J., Janknegt R.A. Variation in diagnostics and treatment of benign prostatic hyperplasia in urological practices (in Dutch). *Ned Tijdschr Geneesk* 1996; 140: 837-42
8. Stoevelaar HJ. Personal communication
9. The intramural health care in figures per January 1 1996 (in Dutch). Nederlandse Zorgfederatie Drukkerij Nzf Faciliteiten, Utrecht, 1997
10. Diagnostisch Kompas 1997 (in Dutch) Boekhoven-Bosch, Utrecht, 1997
11. Farmacotherapeutisch Kompas 1997 (in Dutch) Boekhoven-Bosch, Utrecht, 1997
12. Vademecum 1997 (in Dutch) Centraal Orgaan Tarieven Gezondheidszorg, 1997
13. Tariffs for medical specialists excluding psychiatrists 1996 (in Dutch) Centraal Orgaan Tarieven Gezondheidszorg, 1996
14. Roos N.P., Wennberg J.E., Malenka D.J. Mortality and reoperation after operation and transurethral resection of the prostate for benign prostatic hyperplasia. *New Engl J Med* 1989; 320: 1120
15. De la Rosette J.J.M.C.H. Personal communication
16. Bosch R.J.H.L., Hop W.J., Kirkels W.J., Schröder F.H. Natural history of benign prostatic hyperplasia: appropriate case definition and estimation of its prevalence in the community. *Urol* 1995; 46 (Supplement 3A): 34-40
17. Centraal Bureau voor de Statistiek (CBS) *Population month statistics*, January 1997 (in Dutch). CBS: Voorburg / Heerlen, 1997
18. Garraway W.M., Collins G.N., Lee R.J. High prevalence of benign prostatic hyperthrophy in the community. *Lancet* 1991; 338: 469-71
19. Jolleys J.V., Donovan J.L., Nanchanal K., Peters T.J., Abrams P. Urinary symptoms in the community: how bothersome are they? *Br J Urol* 1994; 74: 551-5
20. Finkler S.A. The distinction between costs and charges. *Ann Intern Med* 1982; 96: 102-9

CHAPTER 6

THE USE OF MODELING IN ECONOMIC EVALUATION: PNEUMOCOCCAL VACCINATION

Published as: Baltussen R.M.P.M.¹, Ament A.J.H.A.¹, Leidl R.M.^{1,2}, van Furth R.³ Cost-effectiveness of vaccination against pneumococcal pneumonia in The Netherlands. *European Journal of Public Health*, 1997; 7: 153-61
Baltussen R.M.P.M.¹, Ament A.J.H.A.¹, Leidl R.M.^{1,2}, van Furth R.³ Kosten-effectiviteit van pneumokokken vaccinatie van ouderen. *Nederlands Tijdschrift voor Geneeskunde*, 1997; 141: 2188-92

¹ From the Department of Health, Organisation, Policy and Economics, Maastricht University

² From the Department of Health Economics, University of Ulm, Germany

³ From the Department of Internal Medicine and Infectious Diseases, University of Leiden

SUMMARY

Because of the prevalence of pneumococcal pneumonia and the substantial morbidity and mortality associated with many pneumococcal infections, considerable efforts have been made in disease prevention, using a polyvalent polysaccheride pneumococcal vaccine. Although the WHO has endorsed a policy of universal vaccination of all elderly people, the economic aspects of such a policy have not been determined. Using a decision-tree framework, this paper examines the cost-effectiveness of various strategies of pneumococcal vaccination for the elderly in The Netherlands. For various age categories, the economic attractiveness of the vaccination of all individuals as well as the vaccination of only those individuals with a specific disease has been calculated. We conclude that, allowing for some uncertainty regarding key variables such as the vaccine efficacy and the hospital admission rate, the vaccination of all individuals above the age of 65 years is comparable in terms of cost-effectiveness to many existing health care interventions. The vaccination of individuals above the age of 55 years with chronic lung disease or chronic heart disease is similarly attractive from an economic point of view, as is the vaccination of individuals above the age of 65 years with diabetes mellitus.

INTRODUCTION

Pneumococcal infections are an important cause of hospitalization and death among elderly individuals and those with a variety of high-risk medical conditions.^{1,2} Pneumococcal pneumonia is the most frequent serious pneumococcal infection, it represents the most common bacteriologic cause of community-acquired pneumonia and it is the primary target for pneumococcal immunization programmes in the elderly. In The Netherlands alone, it is estimated that 16,000 cases of pneumonia required hospitalization in 1994,³ pneumococcal infections being responsible for ~30% of this total.⁴ Despite the availability of antibiotic therapy, the hospital mortality rate for pneumococcal pneumonia is as high as 16.5% for those above the age of 55 years.³

The 23-valent pneumococcal polysaccharide vaccine was licensed for use in The Netherlands in 1984. In 1988 the World Health Organization (WHO) recommended that all elderly people should receive pneumococcal vaccination.¹ This policy was found to be economically attractive for the USA.⁵ However, the cost-effectiveness of this vaccination strategy may be highly dependent on country-specific factors, such as the incidence of the disease, hospital admission rates and the costs of treatment of pneumococcal infections. Using a model approach, this study considers the assumptions under which the vaccination of the elderly will be cost-effective in The Netherlands.

METHODS AND MATERIALS

Model structure

To estimate the economic aspects of pneumococcal vaccination strategies, a decision tree has been applied. Figure 1 summarizes this tree as it is operationalized in this study. The decision tree describes the risks of becoming infected after vaccination and the associated probabilities of recovery and hospital admission. For various age categories, the economic attractiveness of vaccination has been determined. We distinguished subgroups of individuals without any specific underlying disease and subgroups of individuals with underlying diseases that are known to constitute a risk factor for acquiring pneumococcal pneumonia and/or which make the course of the disease more severe. From the literature, we identified 3 types of underlying diseases: diabetes mellitus (ICD-9 code: 250), chronic lung disease (ICD-9 codes: 491 - 494, 496) and chronic heart disease (ICD-9 codes: 412 - 414 and 428). A cost-effectiveness analysis requires a comparison of the costs and health effects for each of the vaccination strategies. We used DATA 2.6 (copyright TreeAge Software, 1995) to perform these calculations and then compared the strategies with the costs and health effects in case no vaccination is initiated.

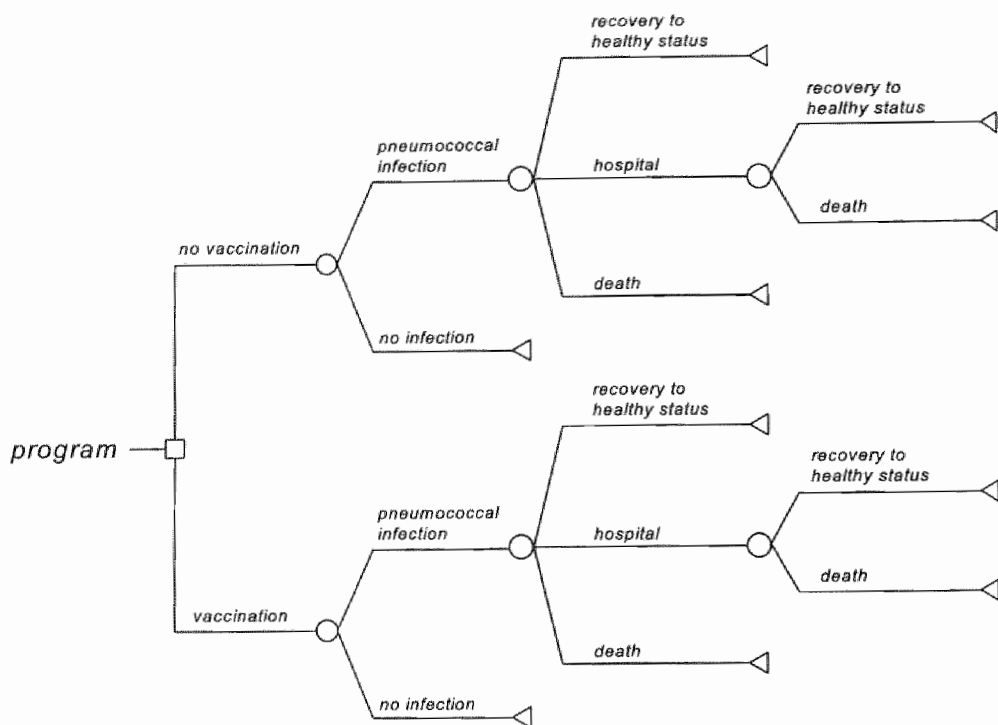


FIGURE 1. Decision tree

The analyses is carried out from a societal point of view and takes into account all direct medical costs, whether paid by patients or by third parties. These direct medical costs include expenditure for purchasing and administering the vaccine, the treatment costs of vaccine side-effects and the treatment costs saved by preventing the disease. Indirect costs, in the sense of the productivity losses of the elderly, are not taken into account: should the use of economic evaluation be restricted to allocation decisions in health care, cost assessments should be limited to (health services) resources which have no alternative other than the production of health.⁶ Costs are estimated in Dutch guilders but are presented in this paper in ECUs (exchange rate December 1995: 1 ECU = 2.07 Dutch guilders). All costs are standardized for 1995.

Health effects are expressed in life-years gained due to preventing pneumococcal pneumonia. Data has been extracted by literature review concerning factors such as vaccine efficacy or from Dutch databases concerning factors such as the incidence of disease or hospital mortality. The cost-effectiveness ratio (CER) expresses the net costs that would be expended on vaccination in order to gain 1 life-year. To reflect the time preferences of society, all future costs and life-years are discounted at a 5% discount rate as mandated by the Dutch Ministry of Finance.⁷

Vaccine

The current recommended immunizing agent is polyvalent pneumococcal vaccine containing purified capsular polysaccharides from 23 of the 83 different types of *Streptococcus pneumoniae*. The 23-type vaccine was licensed in The Netherlands in 1984. Studies of pneumococcal isolates obtained from blood or other sterile body sites indicate that this 23-type vaccine usually covers 90% of capsular polysaccharide types.⁸

Efficacy. Although randomized controlled trials can theoretically provide the best estimate of vaccine efficacy, the large sample size required to have adequate statistical power makes the detection of effects difficult.^{2,9-11} This may be one explanation for the negative results of various randomized controlled trials.¹²⁻¹⁴ A meta-analysis of randomized controlled trials failed to prove efficacy for the elderly, probably due to the broad definition of the subgroups.¹⁵ Observational, non-randomized studies of efficacy, however, have demonstrated an estimated efficacy against bacteremic pneumococcal infections ranging from 46 to 78% for persons over 55 years of age, with or without chronic underlying diseases.¹⁶⁻¹⁹ Some authors argue it is 'reasonable' to extrapolate results on the efficacy of vaccination against bacteremic pneumococcal infections (which are more convincing) to non-bacteremic pneumococcal infections.^{5,10} As a result, the view shared by most authors is that the efficacy of the vaccine ranges from 60 to 70% for immuno competent persons over age 55 years with or without chronic underlying disease,^{1,2,9,10,19,20} although others are more sceptical.^{11,21} The base case analysis assumes an efficacy rate of 60% for all subgroups of elderly people.

Duration. Vaccine efficacy has been shown to persist for a period of 5 - 8 years but appears to wane afterwards.^{18,19} In the base case analysis, we apply a vaccine duration of 5 years.

Safety. Vaccine-associated reactions occur within 24 h of injection in small groups of elderly vaccinees and consist primarily of discomfort, erythema and induration that last a couple of days. One study reported an incidence of 8 complications requiring GP consultation per 1,000 vaccinated individuals²²; this figure has also been applied in the base case analysis.

Epidemiology and care

Incidence in the health care system. The estimates of the incidence of pneumococcal pneumonia in general practice are based on data on the annual incidence rates of community-acquired pneumonia from a large Dutch General Practitioners' survey including 340,000 patients, carried out in 1987 - 1988. This is considered representative for the Dutch population as a whole.²³ Estimates in the international literature of the proportion of all pneumonias that are pneumococcal range from 15 to 40%.^{1,5,10,22,24,25} As a base case scenario in the present study the most cited percentage

of 25% is applied. Regarding sensitivity analyses, 2 alternative scenarios are developed in which the maximum and minimum values as reported in the international literature are applied: scenario A assumes that 40% of all pneumonias are pneumococcal pneumonias, while scenario B assumes that 15% are.

Estimates of the hospital admission rate for pneumococcal pneumonia are based on a database that includes more than 99% of all hospital admissions in The Netherlands in 1994.³ Cases of pneumonia are defined as those which are admitted to hospital with pneumonia (ICD-9 480 - 486) as the first diagnosis. In the international literature, estimates of the proportion of these pneumonias that are caused by pneumococcal pneumonias vary between 20 and 40%.^{4,26-28} In the base case analysis, we apply the most cited percentage of 30%. In the sensitivity analyses, two alternative scenarios are developed in which the maximum and minimum values as reported in the international literature are applied: scenario A assumes that 40% of all in-patient pneumonias are pneumococcal pneumonias, while scenario B assumes a percentage of 20%.

These two methods are applied to calculate the incidence rates for the selected subgroups of individuals (Table 1). The number of hospital admissions divided by the incidence in the general practice can be regarded as the hospital referral rate, e.g. for individuals aged 65 - 69 years, the referral rate is 18.75% (57 out of 304).

Length of stay. Estimates of the average length of stay for the various subgroups of elderly people admitted with pneumococcal pneumonia have been based on data on hospital admissions for pneumonia as the first diagnosis, in The Netherlands in 1994. For example, the average length of stay for patients of 55 - 59 years is 14.6 days; this figure increases with the age of the individuals involved.³

Mortality rates. Estimations of general practice mortality rates of patients with pneumococcal pneumonia are based on 1994 Dutch mortality statistics for pneumonia.²⁹ Estimations of hospital mortality rates are based on data on hospital admissions for pneumonia in The Netherlands in 1994. For example, the mortality rates in the general practice and at the hospital for patients of 55 - 59 years are respectively 0.02 and 5.6%; these figures increase with the age of the individuals considered.³

Life-years gained. The (discounted) number of life-years gained as a result of a vaccination strategy is estimated as the difference between the (discounted) average number of life-years that individuals live if they are vaccinated and if they are not vaccinated. To calculate these average numbers of life-years, the life expectancy of individuals as reported in life tables for corresponding age categories is used as an input in the decision tree.³⁰

TABLE 1 The annual incidence of pneumococcal pneumonia per 100,000 individuals^a

Age category	Base case analysis	
	General practice incidence rate (25%)	Hospital admission rate (30%)
50 - 59 years	137	23
60 - 64 years	247	37
65 - 69 years	304	57
70 - 74 years	474	84
75 - 79 years	441	123
80 - 84 years	595	165
85 years and older	929	202
55 - 64 years with chronic lung disease	(606)	74
65 - 74 years with chronic lung disease	1,356	123
75 years and older with chronic lung disease	1,395	168
55 - 64 years with diabetes mellitus	--	41
65 years and older with diabetes mellitus	-248	98
55 - 64 years with chronic heart disease	-244	98
65 - 74 years with chronic heart disease	585	90
75 years and older with chronic heart disease	1,040	175

a) Based on data from the 1989 General Practitioners' survey²³ and on hospital admissions in The Netherlands in 1994.³ The assumptions, in the scenario, regarding the percentage of all pneumonias caused by pneumococcal pneumonias are presented between parentheses. Where the number of pneumonia cases is fewer than 10, the incidence figure is enclosed in parentheses. A dash -- indicates no exposure.

The life expectancy of individuals with an underlying disease is calculated using the DEALE (declining exponential approximation of life expectancy) methodology which is based on the assumption that survival follows a simple declining exponential function. Life expectancy is obtained by combining disease-specific mortalities with the age-, sex- and race-adjusted mortality rate of a healthy individual.³¹ The (excess) mortality rates for individuals with underlying diseases as applied in this model are assumed equal for all ages and are based on data from Japanese and UK insurance companies (chronic lung disease 0.026, diabetes mellitus 0.011 and chronic heart disease 0.012).³²

Costs

We distinguished vaccination costs, general practice treatment costs and hospital treatment costs. To adjust for 1995 prices, where necessary, an annual price increase of 3.6% has been assumed.³³

The 1995 vaccine retail price was ECU 12.95 per dose.³⁴ Dispensing costs are not included in the analysis; they vary with the number of purchased vaccines and may be considered low in the case of a large-scale vaccination programme. In the base case analysis it is assumed that general practice administration of the vaccine takes half of one consultation and that the treatment of vaccine side-effects requires one general practice consultation. With costs per general practice consultation estimated at ECU 14.15³⁵, the total cost of vaccination per individual is ECU 20.19. Treatment of pneumonia by the general practice is assumed to require three to four consultations (ECU 49.54) plus medication (antibiotic, ECU 24.28). If the patient is sent directly to the hospital, it is assumed that only one consultation is involved.

The calculation of costs of hospital treatment for pneumococcal pneumonia is based on separate estimates of treatment costs and hospital nursing costs. The estimation of the respective treatment costs of the various subgroups of the elderly is based on 1994 cost profiles of patients treated at the Academic Hospital of Maastricht (AZM) which include departmental personnel and material costs, but exclude nursing costs, medication costs, departmental capacity costs and hospital overhead costs. To correct for hospital overhead costs, 40% is added to these treatment costs.³⁵ The total treatment costs at general hospitals are estimated to be 55% of the costs at an academic hospital.³⁵

Hospital nursing costs are calculated by multiplying the age-specific average length of stay by the average daily hospital nursing costs. Standardizing for 1995 prices and applying the weighted average - based on the number of hospital days - of the costs of academic and general hospitals in The Netherlands (STG 1993), the daily hospital nursing costs are estimated at ECU 161.35.

From the above it can be seen that there are many variables which have identical values across all subgroups of elderly people (such as the costs of vaccination and daily hospital nursing costs), and a limited number of variables which create differences in the CERs between the various subgroups of elderly people: the general practice incidence rate, hospital admission rate, hospital length of stay, hospital treatment costs (which do not vary across age groups), hospital mortality rate, general practice mortality rate (which does not vary by type of underlying disease) and life expectancy.

RESULTS

Base case analysis

Incremental costs and incremental effects of vaccination using the base case assumptions regarding incidence rates, vaccine effectiveness and costs are shown in Figure 2. For convenience, only estimates for individuals without any specific underlying diseases are shown.

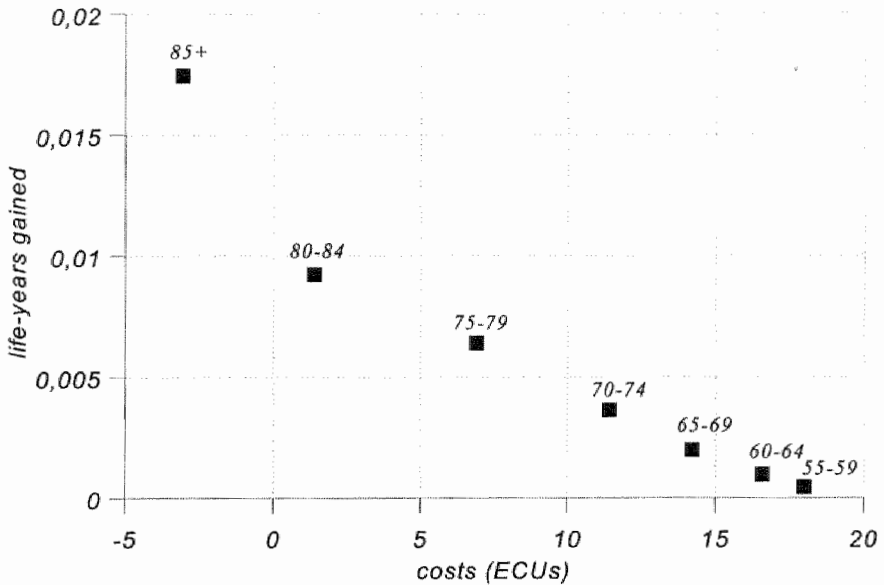


FIGURE 2. Base case costs and effects per case of vaccination against pneumococcal pneumonia for various age groups

For example, the figure shows that vaccinating an individual aged between 55 and 59 years will cost ~ ECU 18 more than not vaccinating that individual and yields *on average* very little (0.00048) in terms of additional life-years. Vaccinating an individual aged 85 years or older appears to be cost-saving and yields *on average* ~ 0.01745 life years.

The incremental CERs are shown in Table 2. Accounting for the case-mix of patients with or without specific underlying disease, the cost-effectiveness of vaccinating all individuals increases with age: vaccinating those aged 55 - 59 years yields a CER of ~ ECU 37,500 per life-year saved, whereas vaccinating individuals of age 85 years and older produces a cost-saving.

TABLE 2. Cost-effectiveness ratio (CER) of different vaccination strategies^a

Age category	No specific underlying disease	Chronic lung diseases	Diabetes mellitus	Chronic heart disease	
55 - 59 years	37,500	10,700	17,800	3,300	
60 - 64 years	16,800				
65 - 69 years	7,100	1,200	2,400	1,800	
70 - 74 years	3,100				
75 - 79 years	1,100	cost-saving		cost-saving	
80 - 84 years	150				
85 years and older	cost-saving				

a) Cost-effectiveness ratios are rounded to the nearest ECU 100.

Vaccinating all individuals above the age of 55 years yields a CER of ~ ECU 3,300 per life-year saved and vaccinating all individuals above the age of 65 years a CER of ~ ECU 1,500 per life-year saved (not shown in table). The cost-effectiveness of vaccinating individuals with a specific underlying disease appears to be moderately more cost-effective for corresponding age groups: it varies from ~ ECU 17,800 per life-year saved for those aged 55 - 64 years with diabetes mellitus to cost-savings or something approaching a cost-saving, for those in higher age groups, regardless of the underlying disease.

One-way sensitivity analyses

Table 3 shows the one-way sensitivity analyses involving the epidemiologic and economic data as applied to 3 subgroups of individuals; those of age 60 - 64 years, 65 - 69 years, and 70 - 74 years. These groups were chosen as being the most relevant groups for decision making. The impact of the variables on the cost-effectiveness of individuals aged 65 - 69 years is discussed below, in order to provide an example of the sensitivity analysis.

Assumptions about vaccine. The analysis is relatively sensitive to variations in the vaccine efficacy rate. If this rate is for example decreased from 60 to 40%, fewer cases of pneumococcal pneumonia are prevented and the incremental CER increases by more than 70% to ~ ECU 12,100 per life-year gained. A sensitivity analysis was also performed under the assumption that the vaccine remains effective for only 3 years instead of 5 years; in that case the CER increases by 79% to ~ ECU 12,700 per life-year saved. If the vaccine provides protection for 8 years, the ratio almost halves compared to the base case value. The analysis is relatively insensitive to the vaccine retail price.

TABLE 3. Results of one-way sensitivity analyses: index of cost-effectiveness ratios (base case = 100%)

	Base case analysis (=100%)	Vaccine efficacy 40% - 80%	Strings covered by vaccine 70% - 90%	Duration of immunity 3 - 8 years	Vaccine retail price ECU 10 - 17	Discounting 3 - 10%
60 - 64 years	16,800	161 - 70%	131 - 97%	176 - 61%	80 - 124%	82 - 157%
65 - 69 years	7,100	171 - 64%	137 - 97%	179 - 53%	77 - 128%	154 - 82%
70 - 74 years	3,100	189 - 56%	146 - 96%	198 - 39%	71 - 135%	84 - 149%

	Vaccine provision 1 GP consultation	Hospital admission rate scenario A, B	Hospital treatment costs 0.75 - 1.25 * base case	GP incidence rate scenario A, B	Life-years gained undiscounted
60 - 64 years	143%	74 - 145%	101 - 99%	88 - 110%	61%
65 - 69 years	150%	69 - 156%	102 - 98%	89 - 109%	63%
70 - 74 years	162%	62% - 162%	104 - 96%	82 - 116%	69%

Assumptions about general practice data. If it is assumed that vaccination requires 1 full general practice consultation instead of half a consultation, the CER increases by 50% to ~ ECU 10,700 per life-year saved. The model is insensitive to other assumptions regarding general practice data.

Assumptions about hospital data. Assumptions regarding the hospital admission rate affect the results considerably: if, for example, scenario B is applied, the CER increases by 56% to ~ ECU 11,100 per life-year saved, i.e. vaccinating becomes much less attractive from an economic point of view. The results are hardly affected by assumptions about hospital costs.

Other assumptions. Assumptions on the rate at which future costs and health outcomes are discounted have a moderate effect on the results; if, for example, the discount rate is increased from 5 to 10%, the incremental cost-effectiveness ratio increases by 54% to ~ ECU 11,000, i.e. vaccinating becomes less attractive. To provide insight into the importance of the various parts of the model, the analysis was also performed considering only hospital data, thereby ignoring the presence of General practices and their costs in the health care system (except for General

practice vaccination costs). As a result the incremental CER increased by a moderate 26% to ~ ECU 9,000 per life-year saved.

Multi way sensitivity analysis

To assess the joint influence of the various assumptions on the CERs, multi way sensitivity analyses were carried out. Multi way sensitivity analyses were carried out for those variables which were shown in the one-way sensitivity analyses to influence the results most strongly. Table 4 shows the results. For example, if we assume a low hospital admission rate for pneumococcal pneumonia (scenario B) and a low vaccine efficacy rate, the CER of vaccinating individuals between 65 and 69 years increases by 254% to ~ ECU 18,000 per life-year saved. A three-way sensitivity analysis assuming a low hospital admission rate (scenario B), a short duration of immunity (3 years) and a low vaccine efficacy rate (40%) yields a CER of ~ ECU 29,000 per life-year saved (not shown in Table 4).

TABLE 4. Results of two-way sensitivity analysis, age category 65 - 69 years: index of cost-effectiveness ratios (base case = 100%)

Variable assumptions	Hospital admission rate scenario A	Hospital admission rate scenario B	Vaccine efficacy 40%	Vaccine efficacy 80%
Vaccine efficacy 40%	124%	254%	--	--
Vaccine efficacy 80%	41%	107%	--	--
Duration of immunity 3 years	89%	188%	288%	125%
Duration of immunity 8 years	30%	94%	104%	28%

DISCUSSION

We estimated the cost-effectiveness of vaccinating the elderly population in The Netherlands against pneumococcal pneumonia. The CER is remarkably high -- some would call it cost-ineffective - for vaccinating those between 55 and 59 years old. However, vaccinating the very old can even produce cost-savings. The vaccination of individuals between 65 and 74 years old with underlying diseases such as chronic lung disease, diabetes mellitus or chronic heart disease appears to be attractive from an economic point of view, as is the vaccination of all individuals between 70 and 74 years. Vaccinating all individuals between 65 and 69 years is, in terms of cost-effectiveness, comparable with firmly established preventive medical interventions such as breast cancer screening (ECU 3,000 per life-year saved)³⁶ or cervical cancer screening (ECU 14,000 per life-year saved)³⁷ (lower bound cost-effectiveness estimates, adjusted to 1995 prices). Since the CER was found to decrease with age,

this implies that vaccination of all individuals above the age of 65 or 70 years should be considered.

The analytic result is sensitive, however, to a number of variables whose values are imprecisely known at present. The results are relatively sensitive to the hospital admission rate of pneumococcal pneumonia. Unfortunately, data concerning the hospital admission rate of pneumococcal pneumonia are inadequate for the purpose of establishing reliable estimates. However, the hospital admission rates for pneumococcal pneumonia as estimated are comparable to those found in other countries and applied in other studies.^{1,5,9,24,25,38} Moreover, by comparing the estimated number of hospital admissions for pneumococcal pneumonia with the estimated incidence rate of pneumococcal pneumonia in general practice, we can calculate hospital referral rates which are comparable to those reported in studies on referral behaviour by GPs in The Netherlands.^{23,39,40}

The results are also somewhat sensitive to vaccine characteristics such as the vaccine efficacy and duration of immunity. This is a critical point since the vaccine efficacy and duration of immunity of vaccination against pneumococcal pneumonia have only been estimated in case-control studies, while randomized controlled trial studies did not prove efficacy. However the assumptions applied in the model do reflect a contemporary opinion on these matters shared by many authors, although others have a more pessimistic view. Assumptions regarding the percentage of pneumococcal pneumonias which are caused by the types covered by the vaccine affect the analytic results moderately; unfortunately reliable values have been established for the pneumococcal bacteremia but not for pneumococcal pneumonia. Finally, the results proved to be moderately sensitive to assumptions regarding the costs of vaccine administration and retail price. The two-way sensitivity analyses showed that the CER for vaccinating individuals aged 65-69 years ranges, depending on the assumptions, between ~ ECU 2,000 per life-year saved and ECU 20,000 per life-year saved, which confirms that the model used is relatively stable to the assumptions applied.

It should be noted that the analysis as a whole is conservative. First, the analysis considered only the effect of vaccination on the incidence of community-acquired pneumococcal pneumonia. Hospital-acquired pneumococcal pneumonia, however, may constitute 30% of the total number of pneumococcal pneumonia cases treated in the hospital.^{4,41} This implies that we may have underestimated the number of prevented cases of pneumococcal pneumonia. Second, we did not include the vaccine's efficacy in preventing meningitis or other diseases; as *S. pneumoniae* ranks third among the pathogens causing meningitis in The Netherlands⁴², this may imply that we have seriously underestimated the total health effects gained. Third, it is the belief that vaccinated individuals have a less severe course of disease if they do

become infected has not been included in the analysis; its inclusion would make vaccination more economically attractive.

Fourth, as a proxy for hospital length of stay and mortality, we used statistics relating to pneumonia rather than pneumococcal pneumonia. As the latter is a more severe disease, this underestimates the effect of vaccination on mortality and on the costs of hospital treatment. Fifth, in estimating hospital treatment costs we ignored costs such as the cost of medicines; in estimating the costs in general practice we ignored the costs of diagnostic tests. However, we did not include vaccine dispensing costs in the analysis.

On the other hand, it is argued that the successful prevention of pneumococcal pneumonia may not be attended by a proportionate reduction in pneumonia deaths as other pathogens may replace *S. pneumoniae* in elderly persons.¹¹ This issue has not been considered in this analysis and we therefore may have somewhat overestimated the reduction in pneumonia deaths and, consequently, the economic attractiveness of vaccination. Moreover, if the number of life-years gained was to be adjusted for quality of life, the CER - in terms of costs per quality-adjusted life year gained -- would have increased. This is particularly true for those in the higher age categories and those with underlying diseases, since the life years they gain will not likely be spent in full health. For example, in a study concerning the perceived health of individuals of all age categories, it has been found that, on average, individuals without any specific disease perceived their health as 75 on a scale of 0 to 100, whereas individuals with diabetes mellitus or chronic lung disease scored, on average, only 60.⁴³ This highlights the need for a cautious assessment of the value of the life-years gained as reported in this analysis.

A number of future events and trends will substantially affect the economic attractiveness of vaccination. One possible future scenario for pneumococcal vaccination in The Netherlands is the joint provision of pneumococcal and influenza vaccination. As influenza predisposes to bacterial pneumonias - including pneumococcal pneumonia - additional pneumococcal pneumonias could be prevented by this joint provision.⁴⁴ The impact on the economic attractiveness of pneumococcal vaccination, however, is uncertain. Yet, important economies of scale can be expected in the sense that general practice vaccination costs decrease considerably. Furthermore, pneumonias have reportedly become increasingly resistant to antibiotics,⁴⁵ thereby offering less scope for medical treatment and increasing the need for vaccination. It thus seems safe to assume that current projections underestimate future health effects and therefore underestimate the attractiveness of vaccination.

There is, however, some discussion whether the incremental CER of specific age groups or the average CER of all age groups above a certain age should be

considered in decision making.⁴⁶ In the latter case, the CER is determined by dividing the total costs by the total health effects stemming from vaccination of all age groups above a certain age. However, average CERs are generally not regarded as good yardsticks for allocation decisions in health care since they may disguise the fact that the strategy in question includes not only economically attractive interventions but also economically unattractive interventions. Average CERs are only relevant when a discrete decision has to be made, i.e. *whether* a vaccination programme should be implemented.

In general and certainly with respect to this vaccination programme, decisions need to be taken regarding *how extensive* the vaccination program should be and which age groups should be invited to participate, so incremental age-related CERs, rather than average CERs, are required. This can be seen from this analysis, which has shown that vaccinating all individuals above the ages of 55 or 65 years yields relatively low average CERs (ECU 3,300 and ECU 1,500 per life-year saved respectively). However, the incremental CER of vaccinating individuals aged between 55 and 59 years (ECU 37,500 per life-year saved) informs decision makers that the inclusion of this age group in a nationwide vaccination programme is relatively unattractive from an economic point of view.

Can these analytical results be extrapolated to other European countries? The epidemiologic data and the impact of effects on life expectancy are unlikely to differ between countries. Hospital admission rates, however, which were an important element in the analysis, may differ significantly between countries. The admission rates in in-patient institutions vary from ~ 10% in The Netherlands and Spain, to more than 20% in countries such as France and Germany.⁴⁷ Costs may differ substantially too. For example, daily hospital costs in Spain have been found to be twice as high as those in The Netherlands or France and even 3 times as high as those in Germany⁴⁸

On the other hand, the average length of hospital stay in Spain is considerably shorter than in the countries mentioned above and less than half of the average stay in The Netherlands.⁴⁸ The overall influence of these effects on the cost-effectiveness of vaccination against pneumococcal pneumonia is difficult to predict. Yet, a number of studies on the economic aspects of pneumococcal vaccination performed in other countries^{5,49,50} yield the same conclusion, i.e. that vaccination of all elderly above the age of 65 years is cost-effective. These conclusions, however, are based on the average cost-effectiveness ratios rather than on incremental analysis and can therefore not be compared with the conclusions derived in this analysis for the case of The Netherlands.

Information on the cost-effectiveness of pneumococcal vaccination is only a first step in providing support for policy making: the next step requires an analysis at the

population level. A decision maker who also has budget responsibility and who allocates funds at the population level must also consider the total budget impact and the total achievable health gains, which implies an explicit formulation of the epidemiologic aspects of a medical intervention to assess how many patients are and will become medically eligible. As a first attempt to provide such information, Figure 3 shows the health production function of vaccination against pneumococcal pneumonia.

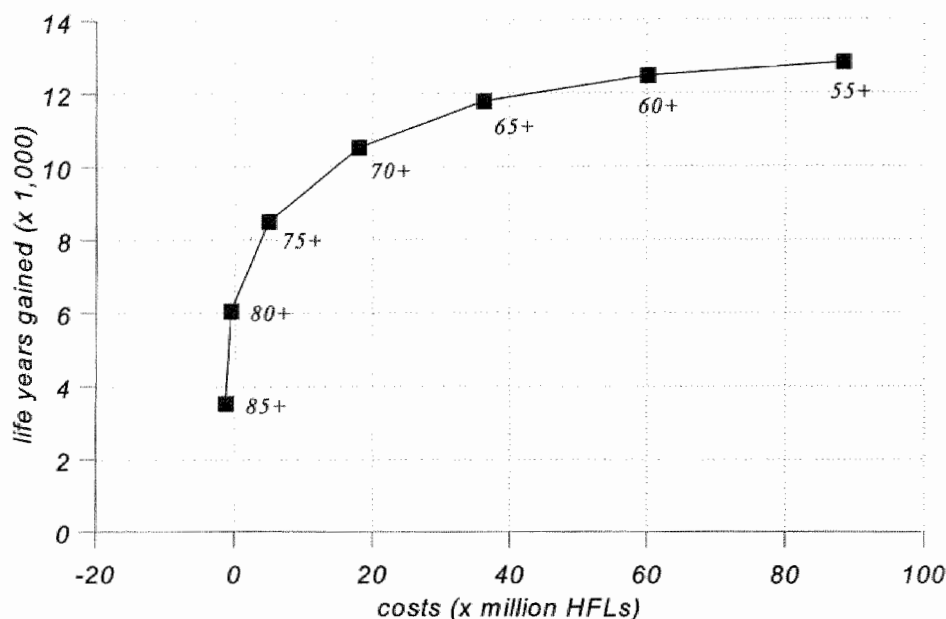


FIGURE 3. Total costs and effects of vaccination against pneumococcal pneumonia in The Netherlands.

A health production function describes the total health gains and the total costs or savings stemming from the introduction of a medical intervention at various scale levels. In the case of vaccination against pneumococcal pneumonia, it is assumed that all individuals eligible for vaccination are indeed vaccinated, so that factors such as accessibility, awareness or acceptability, which are known to limit the number of individuals undergoing a medical intervention in practice⁵¹ are disregarded here.

It should be noted, however, that the costs saved or incurred, as shown in Figure 3, may fall partially within the budget of the decision makers responsible for the vaccination programme and partially in other budgets for the health care sectors. The figure demonstrates that the total costs and total health effects from vaccinating all individuals in The Netherlands above the age of 65 years come to ~ ECU 19 million

and 12,000 life-years respectively. Vaccinating those above the age of 85 years seems to be a cost-saving intervention and it yields more than 3,000 life-years.

We conclude that, allowing for some uncertainty regarding key variables such as the vaccine efficacy and the hospital admission rate, vaccination of all individuals above the age of 65 years is in terms of cost-effectiveness comparable to many existing health care interventions. The vaccination of individuals above the age of 55 years with chronic lung disease or chronic heart disease is similarly attractive from an economic point of view, as is the vaccination of individuals above the age of 65 years with diabetes mellitus.

We would like to thank Past ur Merieux MSD for their financial support. We would also like to thank G.J. Ligthart, G.B.W.E. Vos, J.F.M. Metsemakers, and V. Coenen, for their advice and contribution to this project.

REFERENCES

1. Fedson D., Henrichsen J., M kel  P.H., Austrian R. WHO recommendations on pneumococcal vaccination. Immunization of elderly people with polyvalent pneumococcal vaccine. *Infection*, 1989; 17: 437-441.
2. Bruyn G.A.W., van Furth R. Pneumococcal polysaccharide vaccines: indications, efficacy and recommendations. *Eur J Clin Microbiol Infec Dis*, 1991; 10: 897-910.
3. SIG Zorginformatie. *Landelijke Medische Registratie 1994*. (National Medical Registration 1994) Utrecht: SIG Cooperatie UA, 1994.
4. Bohte R., van Furth R., van den Broek P.J. Aetiology of community-acquired pneumonia: a prospective study among adults requiring admission to hospital. *Thorax*, 1995; 50: 543-547.
5. Sisk J.E., Riegelman R.K. Cost-effectiveness of vaccination against pneumococcal pneumonia: an update. *Ann Int Med*, 1986; 104: 79-86.
6. Mooney G. *Key issues in health economics*. Hemel Hempstead: Harvester Wheatsheaf, 1994.
7. Ministerie van Financi n, Studiegroep Discontering. *De rol van de disconteringsvoet in het beleid: eindrapport*. (The role of the discount rate in policy making: final report). The Hague: Ministry of Finance, 1985.
8. Hager H.L., Woolley T.W., Berk S.L. Review of recent pneumococcal infections with attention to vaccine and nonvaccine serotypes. *Rev Infect Dis*, 1990; 12: 267-272.
9. Fedson D.S. Clinical practice and public policy for influenza and pneumococcal vaccination of the elderly. *Clin Geriat Med*, 1992; 8: 183-199.
10. Fiebach N., Beckett W. Prevention of respiratory infections in adults. Influenza and pneumococcal vaccines. *Arch Int Med*, 1994; 154: 2545-2557.
11. Hirschmann J.V., Lipsky B.A. The pneumococcal vaccine after 15 years of use. *Arch Int Med*. 1994; 154: 373-377.
12. Austrian R., Gold J. Pneumococcal bacteremia with especial reference to bacteremic pneumococcal pneumonia. *Arch Dis Child*, 1964; 60: 759-776.

13. Simberkoff M.S., Cross A.P., Al-Ibrahim M. et al. Efficacy of pneumococcal vaccine in high-risk patients. *New Engl J Med*, 1986; 315: 1318-1331.
14. Forrester H.L., Jahnigen D.W., LaForce F.M. Inefficacy of pneumococcal vaccine in a high-risk population. *Am J Med*, 1987; 83: 425-430.
15. Fine M.J., Smith M.A., Carson C.A. et al. Efficacy of pneumococcal vaccination in adults. A meta-analysis of randomized controlled trials. *Arch Int Med*, 1994; 154: 2666-2677.
16. Shapiro E.D., Clemens J.D. A controlled evaluation of the protective efficacy of pneumococcal vaccine for patients at high risk of serious pneumococcal infection. *Ann Int Med*, 1984; 101: 325-330.
17. Sims R.V., Steinmann W.C., McConville J.H., King L.R., Zwick W.C., Schwartz J.S. The clinical effectiveness of pneumococcal vaccine in the elderly. *Ann Int Med*, 1988; 108: 653-657.
18. Shapiro E.D., Berg A.T., Austrian R., et al. The protective efficacy of polyvalent pneumococcal polysaccharides vaccine. *New Eng J Med*, 1991; 325: 1453-1460.
19. Butler J.C., Breiman R.F., Campbell J.F., Lipman H.B., Broome C.V., Facklam R.R. Pneumococcal polysaccharide vaccine efficacy. An evaluation of current recommendations. *JAMA*, 1993; 270: 1826-1831.
20. Bentley D.W. Vaccinations. *Clin Geriat Med*, 1992; 8: 745-760.
21. Simberkoff M.S. Pneumococcal vaccine in the prevention of community-acquired pneumonia: a sceptical view of cost-effectiveness. *Sem Respirat Infect*, 1993; 8: 294-299.
22. Gable C.B., Holzer S.S., Engelhart L. et al. Pneumococcal vaccine: efficacy and associated cost savings. *JAMA*, 1990; 264: 2910-2915.
23. Van der Velden J., de Bakker D.H., Claessens A.A.M.C., Schellevis FG. *Een nationale studie naar ziekten en verrichtingen in de huisartspraktijk: Basisrapport: morbiditeit in de huisartsenpraktijk*. (A national study on diseases and procedures in general practice) Utrecht: Nederlands Instituut voor onderzoek van de eerstelijnsgezondheidszorg (NIVEL), 1991.
24. Morbidity and Mortality Weekly Report Centres for Disease Control. Recommendations of the immunization practices advisory committee pneumococcal polysaccharides vaccine. *JAMA*, 1989; 261: 1265-1267.
25. Bartlett J.G., Mundy L.M. Community-acquired pneumonia. *New Engl J Med*. 1995; 14: 1618-1624.
26. Ortqvist A., Hedlund J., Grillner L. et al. Aetiology, outcome and prognostic factors in community-acquired pneumonia requiring hospitalization. *Eur Respirat J*, 1990; 3: 1105-1113.
27. Karalus N.C., Cursons R.T., Leng R.A. et al. Community-acquired pneumonia: aetiology and prognostic index evaluation. *Thorax*, 1991; 46: 413-418.
28. Research Committee of the British Thoracic Society and the Public Health Laboratory Service. Community-acquired pneumonia in adults in British hospitals in 1982 - 1983: a survey of aetiology, mortality, prognostic factors and outcome. *Q J Med New Series*, 1987; 62: 195-220.
29. Centraal Bureau voor de Statistiek (CBS). *Overleden naar doodsoorzaak*. (Mortality statistics by cause) Voorburg: Afdeling Doodsoorzaken, 1994.
30. Centraal Bureau voor de Statistiek (CBS). *Overlevingstafels naar leeftijd en geslacht*. (Life tables by age and sex) Voorburg: Hoofdafdeling Bevolkingsstatistieken, 1992.

31. Beck J.R., Pauker S.G., Gottlieb J.E., Klein K., Kassirer J.P. A convenient approximation of life expectancy (the 'DEALE'): II. Use in medical decision making. *Am J Med*, 1982; 73: 889-897.
32. Lew E.A., Gajewski J. *Medical risks: patterns of mortality by age and time elapsed*. New York: Praeger, 1990.
33. Centraal Bureau voor de Statistiek (CBS). *Statistisch bulletin 46*. Voorburg: Centraal Bureau voor de Statistiek, 1995.
34. Farmacotherapeutisch Kompas 1995: medische farmaceutische voorlichting. (medical pharmaceutical counseling) Edited by van de Kuy A. Amstelveen: Centrale Medisch Pharmaceutische Commissie van de Ziekenfondsraad, 1982.
35. Stuurgroep Toekomstscenario's Gezondheidszorg (STG). *Kostenberekening bij gezondheidszorgonderzoek*. (Cost assessment at health care research) Rijswijk: STG, 1993.
36. de Koning H., van Ineveld B.M., van Oostmarssen G.J. et al. Breast cancer screening and cost-effectiveness: policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 1991; 49: 531-537.
37. Koopmanschap M.A., Lubbe K.T., van Oortmarssen G.J., van Agt H.M., van Ballegooijen M., Habbema J.K. Economic aspects of cervical cancer screening. *Soc Sc Med*, 1990; 30: 1081-1087.
38. Austrian R. Pneumococcal pneumonia: diagnostics, epidemiologic, therapeutic and prophylactic considerations. *Chest*, 1986; 90: 738-743.
39. Centraal Medische Registratie Nijmegen (CMR-Nijmegen). *Hospital referral rates for pneumonia 1990 - 1994*. Katholieke Universiteit Nijmegen: Department of General Practice and Social Medicine, 1995
40. Lamberts H. *In het huis van de huisarts: Verslag van het Transitieproject* (report from the Transition project). Lelystad: Meditekst, 1991.
41. Chang J.I., Mylotte J.M. Pneumococcal bacteremia. Update from an adult hospital with a high rate of nosocomial cases. *J Am Geriat Soc*, 1987; 35: 747-754.
42. Bruyn G.A.W., Kremer H.P.H., de Marie S., Padberg G.W., Hermans J., van Furth R. Clinical evaluation of pneumococcal meningitis in adults over a twelve-year period. *Eur J Clin Microbiol Infect Dis*, 1989; 8: 695-700.
43. Stewart A.L., Greenfield S., Hays R.D. et al. Functional status and well-being of patients with chronic conditions. *JAMA*, 1989; 262: 907-913.
44. Gross P.A., Hermogenes A.W., Sacks H.S., Lau J., Levandowski R.A. The efficacy of influenza vaccine in elderly persons: a meta-analysis and review of the literature. *Ann Int Med*, 1995; 123: 518-527.
45. Hofmann J., Cetron M.S., Farley M.M. et al. The prevalence of drug-resistant streptococcus pneumoniae in Atlanta. *New Engl J Med*, 1995; 333: 481-486.
46. Kamlett M.S. *The comparative benefits modeling project: a framework for cost-utility analysis of government health care programs*. Washington: US Department of Health and Human Services, Public Health Service, Office of Disease Prevention and Health Promotion, 1992.
47. Organisation for Economic Co-operation and Development (OECD). *OECD health systems: Fact and trends 1960-1991*. OECD, 1993.
48. Huber M., Kose A., Schneider M. *Wirtschaftlichkeit und Leistungsniveau deutscher Krankenhauser im internationalen Vergleich*. (The performance of German hospitals in international comparison) Augsburg: BASYS, 1993.

49. Willems J.S., Sanders C.R., Riddiough M.A., Bell J.R. Cost-effectiveness of vaccination against pneumococcal pneumonia. *New Engl J Med*, 1980; 303: 553-559.
50. Plans P., Garrido P., Salleras L. Coste-effectividad de la vacunación pneumocócica. *Rev Esp Salud Pública* 1995; 69: 409-417.
51. Tugwell P., Bennett K.J., Sackett D.L., Haynes R.B. The measurement iterative loop: a framework for the critical appraisal of need, benefits and costs of health interventions. *Journal of Chronic Diseases* 1985; 38: 339-351.

CHAPTER 7

REPORTING THE RESULTS FROM ECONOMIC EVALUATION TO POLICY MAKERS

Published as: Ament A.J.H.A., Baltussen R.M.P.M. The interpretation of results of economic evaluation: expliciting the value of health. *Health Economics* 1998; 6: 625-635

From the Department of Health, Organisation, Policy and Economics, Maastricht University

SUMMARY

This chapter examines the reporting of cost-effectiveness results to policy makers. Theoretically it can be proven that an optimal allocation of resources within a constrained budget can be reached by considering cost-effectiveness ratio's (CERs). In this chapter the complex priority setting process regarding compatible and incompatible alternatives of medical interventions is clarified. Priority setting in the context of compatible alternatives may refer to the selection of more than one, possibly all alternatives. Inherent to a set of incompatible alternatives is that only one alternative can be selected. This latter situation frequently occurs in health care. The value that society attaches to a unit of effectiveness (eg a QALY) has an important impact on the priority ranking of medical interventions. By explicating this value and by using the 'net-value' approach, a graphical framework is presented that allows decision makers a better understanding on the impact of particular levels of these values on their optimal policy choices. For illustrative purposes, it is shown that by the erroneous application of decision rules, some recent papers have provided sub-optimal recommendations for health care policy. This study shows the importance of adequately reporting results to policy makers. The main finding of the study is that researchers should present not only cost-effectiveness ratios, but should also report costs and effectiveness separately in order to allow an assessment of the impact of the value that society attaches to a unit of effectiveness.

INTRODUCTION

Economic evaluations are performed primarily to inform decision makers about the relation between the costs and benefits of new and existing technologies.^{1,2} This article clarifies the complex priority setting process regarding compatible and incompatible alternative medical interventions. Alternatives of medical interventions are said to be compatible if more than one alternative, and possibly all alternatives, may be selected. Alternatives are said to be incompatible if only one alternative can be selected. 'Incompatible' alternatives have also been called 'mutually exclusive', 'competing' or 'interdependent' alternatives.²⁻⁶ This situation frequently occurs in health care, e.g. regarding decisions with respect to the intensity or age-limit of screening or vaccination programs.

Considering the large number of alternatives possible, the priority setting process regarding compatible and incompatible alternative medical interventions requires sophisticated techniques such as algorithms and may result, due to its complexity, in erroneous decisions. In this paper, a graphical framework is presented that allows a better understanding of the complex decision making process. The framework recognises the problem that the value that society attaches to health outcomes (e.g. a QALY) is not known. The uncertainty with regard to this cut-off value between acceptable and unacceptable interventions heavily complicates optimal decision making in health care. It will be shown that alternative values may lead to differences in priority ranking.

For illustrative purposes, some recent articles are evaluated using this framework to see whether the insights presented here could have influenced the conclusions reached in these studies. The analytical framework will also be used to deal with a decision making problem known as the 'Rittenhouse Dilemma'.⁷

THE DECISION RULES FOR OPTIMAL DECISION MAKING

There is a vast body of literature that addresses various issues on the interpretation of results from economic evaluation. The traditional approach to the mathematical solution of the maximisation of output when constrained by a fixed budget is to use linear programming.⁸ Attention has also been paid at the impact of indivisibilities of incompatible programs in the presence of a constrained budget⁹, or at the importance of assuming constant marginal cost and benefits¹⁰ which would require the technique of integer programming. In all approaches it is shown that a Pareto optimal allocation of resources can be achieved if perfect information is available on the incremental costs and benefits of all interventions, and on the value that society attaches to health. If such an allocation is achieved it is technically impossible to produce more output within the available budget.^{11,12}

However, a major obstacle to the optimal allocation of health care resources is that the value that society attaches to health (or: a unit of effectiveness) is unknown. In a Pareto optimal allocation, this value determines the least cost-effective intervention which is to be implemented. In the literature, this value has also been referred to as the 'willingness to pay', 'cut-off' level of permissible cost or 'acceptable cost' per unit of effects. The term used in this article is 'maximum Cost Effectiveness Ratio' (CERmax). In the context of expressing units of effectiveness in QALYs, this implies that the CERmax becomes the monetary value of a QALY. Weinstein proposes a number of approaches to estimate the values that society attaches to eg QALYs but refrains from reporting explicit figures.² Laupacis et al.¹³ suggest that the value for a QALY probably lies between \$20,000 and \$100,000: their approach, however, is subject to major criticism.¹⁴ As will be shown, uncertainty regarding this value heavily complicates decision making and may effect the priority ranking of health care interventions.

The impact of alternative values of CERmax on decision making in case of compatible alternatives should be distinguished from that in case of incompatible alternatives. Optimal decision making in the context of compatible alternatives implies the ranking of programs from the lowest to the highest CER. Programs are selected from this rank as long as resources are available. This implies that there is a relation between the available resources and the value of CERmax. Given a fixed budget a certain level of CERmax will result. Otherwise, if the value of CERmax would dominate the allocation process this would lead to a certain budget.

Optimisation in the case of incompatible alternatives is a more complicated process. Inherent to a set of incompatible alternatives is that only one alternative can be selected. To facilitate decision making, procedures have been developed as that by Kamlet⁶ or Weinstein¹⁵, which are rather complex and consists of an iterative procedure. The rule is to rank alternatives of programs according to their costs. A less effective more costly alternative is said to be strongly dominated by a more effective less costly alternative, and should never be accepted. For the remaining alternatives, the incremental CER for each successively more costly alternative (e.g. incremental cost per incremental gain in QALYs) should be calculated. If any of these incremental ratios turns out to be less than the previous one in the sequence of increasingly costly incompatible alternatives, then the more costly is ruled out by weak dominance (also called extended dominance), and it should never be implemented irrespective of the amount of resources available.

This algorithm results in a sequence of programs with increasing incremental CERs. The optimal decision rule is to implement the alternative with the highest incremental cost-effectiveness ratio that is equal or below the value society attaches to a unit of effectiveness, CERmax. Next, the selected alternative - derived from a set of

incompatible alternatives - is to be considered as a compatible program and needs to be rank ordered, on the basis of average CERs, with other compatible programs.

Another, less well-known, approach to priority setting is the 'net-value' approach. Instead of determining incremental CERs to identify the optimal alternative within a set of incompatible alternatives, net values are calculated.^{6,16,17} In this method, the value of CERmax is used to calculate the monetary value of the benefits (number of QALY multiplied by CERmax) of a certain alternative. The net benefit can be found by subtracting the cost from the (monetary) benefits. The alternative with the highest net value is the optimal one. Next, this alternative is considered as a compatible program which is rank ordered among other compatible programs. Subsequently, the rule is to select programs from this rank list until the average CER of a program equals CERmax. Both approaches - the algorithm and the net-value - lead to the selection of the same alternative as optimal. However, the latter approach allows us to show explicitly the influence that various values of CERmax have in the priority setting process regarding compatible and incompatible (alternatives of) programs. This approach provides the basis of the graphical framework. In this framework, alternative values of CERmax are applied to show its impact on priority setting. It should be recognised, however, that these values are not based on any theoretical or empirical findings.

In the analysis, the paradigm as defined by Weinstein² is applied; this includes the assumptions that all programs are divisible and have constant returns to scale. However, in situations of indivisible projects and a binding resource constraint, the opportunity cost of the selection of an alternative should take account of both the direct opportunity costs of the resources required for this alternative and the uses of residual resources foregone by choosing that alternative. In other words, as the size of the alternative chosen increases, the optional uses of residual resources become increasingly restricted and the opportunity costs of marginal resources increase⁹. However, in the present framework, as stemming from the paradigm, the opportunity costs of resources are assumed constant. This is reflected by applying a single CERmax when considering incompatible alternatives of different costs.

THE METHODS OF ISOQUANTS

The graphical framework is explained by the following example derived from Kamlet.⁶ Consider costs and effects of seven incompatible alternatives for implementing the same hypothetical intervention (Table 1). These costs and effects are graphically presented in Figure 1, in which the numbers refer to the alternatives mentioned in the table. The question to be answered is: Which alternative represents the best choice?

TABLE 1. Incremental costs (\$) and effects (QALY) of incompatible alternatives

Health interventions	Costs in \$	Gain in QALYs
0 (status quo)	0	0.0
4	40,000	4.0
2	60,000	2.0
1	60,000	3.0
5	60,000	5.0
6	100,000	7.5
3	110,000	7.0

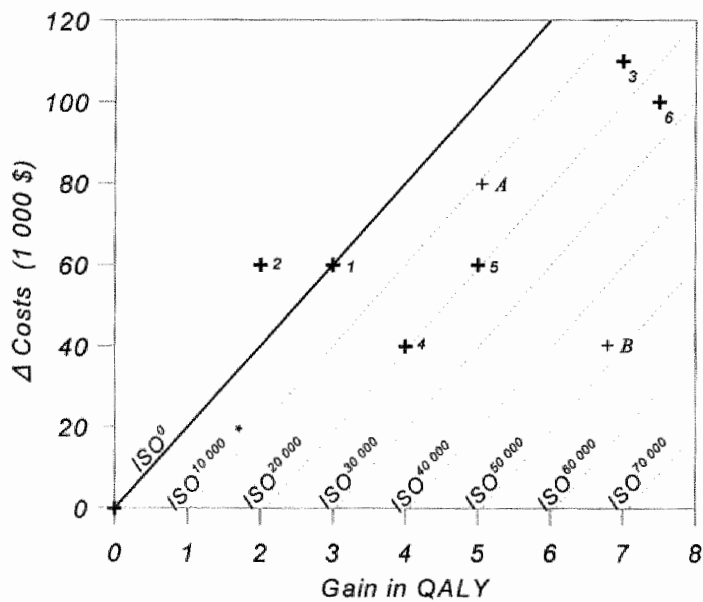


FIGURE 1. The isoquant method. Isoquants based on CERmax = \$20,000

In the figure various isoquants are drawn, representing interventions with the same net-value. The isoquant representing CERmax is starting at the origin and includes all possible combinations of costs and effects that yield a CER equal to CERmax. A flat line indicates a relative low value of CERmax, a steep line indicates a relative

high value of CERmax. The other isoquants are drawn parallel to the line that represents CERmax and connect all possible combinations of costs and effects with the same net value. Alternatives located on isoquants below the CERmax-line indicate alternatives that yield more effects for the same costs or alternatives that yield the same effects for less costs; these alternatives thus represent positive net values. Simultaneously, alternatives located on isoquants above the CERmax-line indicate negative net values. A relatively large distance between an isoquant and the CERmax-line means a relatively high net value, either positive or negative.

To identify the alternative with the highest net value, strongly and weakly dominated alternatives can be identified first. Graphically, strongly dominated alternatives are located north-west of another alternative (the latter representing lower costs and more effects). Weakly dominated alternatives are located north-west of an imaginary line between two other alternatives (dominated by a more costly alternative that has a lower incremental CER in the sequence of increasingly costly alternatives). Having eliminated the strongly or weakly dominated alternatives, it is possible to focus on the remaining alternatives. These remaining alternatives form the 'south east' boundary of the total set of incompatible alternatives and can easily be identified directly in a graph. The optimal choice among these remaining alternatives depends on the value of CERmax and is to be determined by drawing isoquants crossing the remaining alternatives. The optimal alternative is located on the isoquant representing the highest net-value. Note that also in this approach, as in existing algorithm^{6,15,18}, the alternative is selected with the highest incremental CER equal to or below CERmax. Moreover, also in this approach alternatives with incremental CER higher than CERmax will never be selected as optimal.

In the example, alternatives 1 and 2 are strongly dominated by alternative 4 because they consume more resources while yielding less benefits. The same applies to alternative 3 which is strongly dominated by alternative 6. Alternative 5 is weakly dominated by alternative 6 because it is located above the imaginary line between alternative 4 and 6: the incremental CER of alternative 6 is lower than that of alternative 5, taking alternative 4 as a base case. Consequently, alternative 4 and 6 establish the south-east-boundary of the total set of alternatives. One of the alternatives should be selected as optimal. By drawing isoquants parallel to a line through the origin and with slope CERmax (\$20,000 per QALY), alternative 6 can be identified as the optimal alternative.

The optimal alternative can also be determined merely by the last step of the above described procedure: determining which alternative is located on the highest isoquant. However, because the choice of the optimal alternative heavily depends on the value of CERmax, it is more appropriate to focus on all the alternatives along the boundary. For example, a CERmax of \$13,000 per QALY implies a relative flat CERmax-line that will lead to the selection of alternative 4 (alternative 4 has an

incremental CER of \$10,000 which is the highest CER among the remaining alternatives below the CERmax). Note that alternative 6 is only to be selected if CERmax is higher than its incremental CER of about \$17,000 per QALY. In general it can be stated that the optimal choice is the one where CERmax is higher than the incremental CER of the next best alternative.

The graphical approach, as presented in the above, may be used to determine the optimal alternative in a set of incompatible alternatives. The method can also be applied in situations where compatible and incompatible alternatives are mixed. In that case, the optimal alternative from each subset of incompatible projects should first be determined: all other alternatives from the subset may be disregarded. Consider again Figure 1 in which alternatives 1 to 6 refer to six incompatible alternatives of one intervention, and in which alternatives A and B refer to two incompatible alternatives of another intervention. As discussed above, respectively alternative 6 and alternative B are to be selected as optimal alternatives (alternative A is strongly dominated by alternative B because it consumes more resources while yielding less benefits). Next, these two alternative should be considered as compatible alternatives. The compatible alternatives are to be prioritised according to their CERs, which implies that alternative B should be rank ordered first.

The isoquant method can be easily performed and can reduce the complications involved in choosing between numerous alternatives across various interventions, thereby avoiding the need to calculate many incremental CERs. The decision making process may be eased by projecting all the alternatives in a graph and determining, visually, the alternatives on the south-east boundary. The optimal alternative is then selected by working out the effect of various values of the CERmax, which can be done using the isoquants method.

EXAMPLES FROM RECENT LITERATURE

The methodological framework developed above will be applied to some examples from recent literature. It will be shown that by applying this framework, and by recognizing the importance of CERmax, conclusions can be drawn which differ from those stated in the papers concerned. Furthermore, it will be shown that the information relevant for policy making is often not presented. The economic evaluations of tuberculin screening¹⁹, gastric cancer screening²⁰, mammography screening²¹ and the treatment of peptic ulcers²² will be discussed in some detail. It should be noted that this analysis is not intended to question the results of these studies, but to discuss issues which could cause decision makers to derive the wrong implications from these results.

Tuberculin screening

In this study, the CERs of both single tuberculin screening and more frequent tuberculin screening are considered for different age groups.¹⁹ Note that these alternatives are incompatible and that from the total set, only one alternative can be selected as the optimal one. In total 20 CERs are reported.

First, the optimal decision regarding 'single' screening and second, the optimal decision regarding more frequent screening will be identified, by considering the value of CERmax. Subsequently, the optimal alternative from these two will be selected. In Figure 2 the CERs are presented, as published in the study. For illustrative purposes only, in the figure the alternatives, belonging to a certain subset, e.g. 'single screening', are connected with each other.

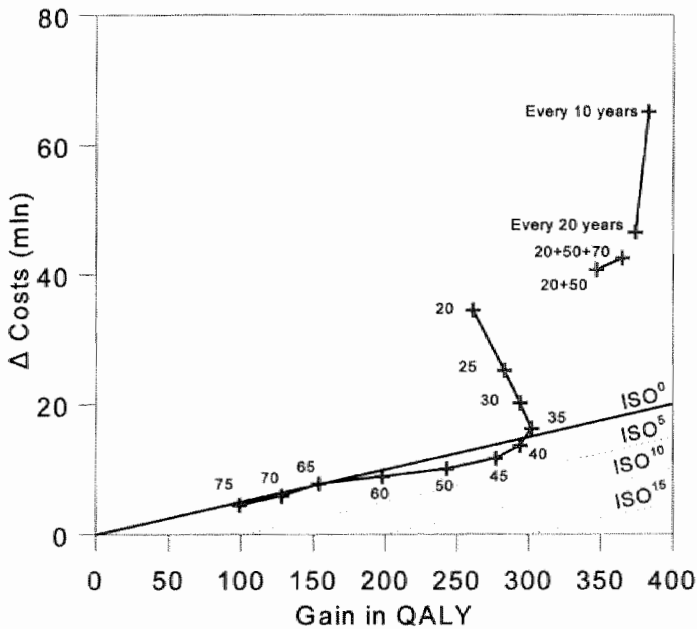


FIGURE 2. Costs and effects of tuberculin screening. Isoquants based on CERmax = \$50,000 per QALY

Figure 2 shows that all alternatives regarding single screening above the age of 40 years are on the same slope, indicating that these alternatives are more or less equally cost-effective: only the scale of costs and effects varies. The figure reveals furthermore that single screening at ages below 35 years is strongly dominated by single screening at the age of 35 years: this can be concluded from the north-west position of these alternatives relative to single screening at age 35 (these alternatives

cost more but yield less benefits). Thus this part of the analysis, regarding single screening, can be restricted to screening alternatives above the age of 35 years. Single screening programs at ages 60, 65 and 70 years are weakly dominated, as these extensions are located north-west of the line between the points indicating single screening at age 50 and at age 75.

Two other subsets of alternatives for more frequent screening are presented in the study. The figure shows that screening at both ages 20 and 50 years is weakly dominated by screening at ages 20, 50 and 70 years, taking alternative 'single screening' at age 35 as reference alternative. Screening every 20 years and every 10 year are not strongly nor weakly dominated and should be considered as possible candidate for the optimal solution. The selection among the remaining alternatives establishing the south-east boundary depends on the value of CERmax, which determines the location of the isoquants. If CERmax equals \$ 50,000, as shown in the figure, single screening at the age of 45 will be selected as optimal.

In the paper it is concluded by the authors that 'tuberculin screening about the age of 50 years is as cost-effective as many other accepted interventions and therefore should be introduced'. Based on the additional analysis above, which makes the effect of the CERmax explicit, we arrive at the same conclusion if CERmax is about \$40,000. However, we would suggest that tuberculin screening should not be accepted where the CERmax is lower than \$30,000. If CERmax is about \$50,000 then screening once in a lifetime at the age of 45 or 50 years is the best strategy. Furthermore it is concluded that 'if further research would approve certain hypotheses, more intensive forms of screening become attractive'. However, our analysis shows that screening at ages 20, 50 and 70 years will only be chosen if the CERmax equals its incremental CER which equals \$294,017 per QALY!

Screening on gastric cancer

This study determines the costs and benefits of three different strategies (indirect X-rays, direct X-rays and endoscopy) for detecting gastric cancer, using modeling techniques.²⁰ Average CERs are presented for both men and women and for interventions at four ages (40, 50, 60 and 70 years). For men aged 50 years, for example, the average CER for a program using indirect X-rays (versus normal care) is about \$10,000, for direct X-rays it is \$30,000, and for endoscopy \$27,000. The programs for gastric cancer screening for the various types of patients (age and gender) are compatible alternatives, but within each such program three strategies are to be considered as incompatible alternatives: e.g. selecting screening by indirect X-ray for example excludes screening by direct X-ray or endoscopy.

According to the method of identifying the optimal alternative discussed above, situations of strong and weak dominance should first be identified.

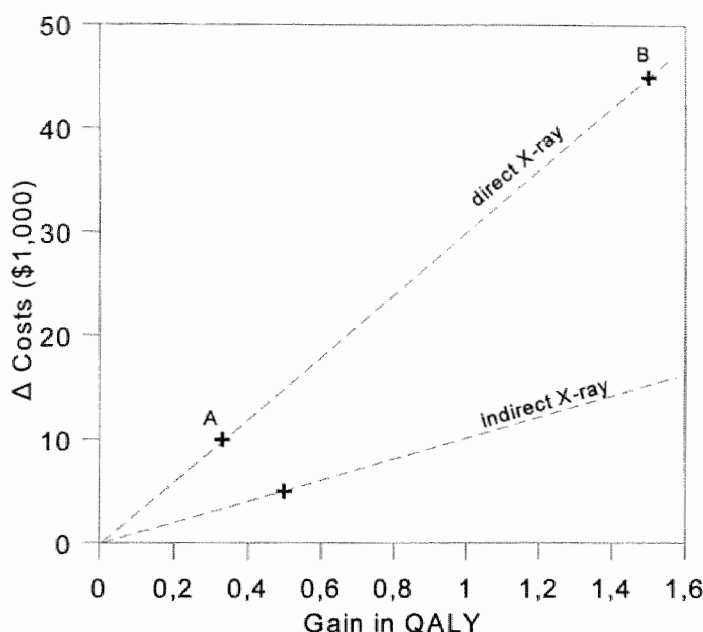


FIGURE 3. Costs and benefits of gastric cancer screening

This, however, requires separate information on the costs and benefits of the various alternatives which is not reported in the study. Suppose that screening by indirect X-ray saves 0.5 life year for \$5,000. Screening using direct X-ray is located somewhere on the line representing all interventions that have a cost-effectiveness of \$30,000 per life year saved e.g. at point A, producing 0.33 life year for \$10,000, or point B, producing 1.5 life years for \$45,000 (Figure 3). In case screening using direct X-rays is represented by point A, this alternative is strongly dominated by screening using indirect X-rays. Possibility B, however, is not strongly or weakly dominated, and could be considered as an optimal alternative, depending on the CERmax. The study concludes that indirect-X rays are a better alternative than the other two screening techniques, for every age group and for both men and women. Our hypothetical analysis shows that this conclusion, being based on average CERs rather than incremental CERs, might be inaccurate. It is evident that an optimal solution cannot be derived without explicitly considering the value of CERmax.

Mammography screening

This study calculates the costs, effects, and CERs, for different mammography screening strategies, according to age and frequency (Table 2).²¹ The alternative screening schedules considered in this study are essentially incompatible alternatives: choosing one strategy implies that all others are excluded. However, a closer look at the alternatives reveals another problem. Some alternatives consists of other schedules plus an extension. For example, schedule B is the same as schedule A (biennial screening between age 50 and 79), but includes an annual screening between age 40 and 49. The choice of which age groups should be screened has to do with compatible alternatives, where as the question how frequently the screening should be applied, has to do with incompatible alternatives. The presentation of alternatives in such a way complicates optimal decision making. The alternatives are graphically presented in Figure 4. Figure 4 shows that alternatives D, F and G are strongly dominated by alternative B. Alternative C is weakly dominated by E because of its location above the imaginary line between alternatives B and E.

TABLE 2. Cost-effectiveness of different strategies for mammography screening

Schedule	Age	Frequency	Costs of screening(\$)	Years of life saved	Cost per year of life saved ^a
A	50-79	Biennial	1,120	0.07	16,000
B	40-49 50-79	Annual Biennial	1,915	0.1	20,200
C	40-64 65-79	Annual Biennial	2,510	0.1	25,000
D	50-79	Annual	2,212	0.09	25,600
E	40-79	Annual	2,972	0.11	27,100
F	40-49 50-79	Biennial Annual	2,550	0.08	31,200
G	40-49 (high risk) 40-49 (normal risk) 50-79	Annual Biennial Annual	2,583	0.81	31,900

a) costs per life year saved are rounded to the nearest \$100

This reduces the set of possible candidates to the south-east boundary in the graph: alternatives A, B and E. Assuming a CERmax of \$30,000 per QALY implies that alternative A has a slightly higher net value than alternative B. Note that this is the alternative with the highest incremental CER (\$16,000 per QALY) which is below CERmax.

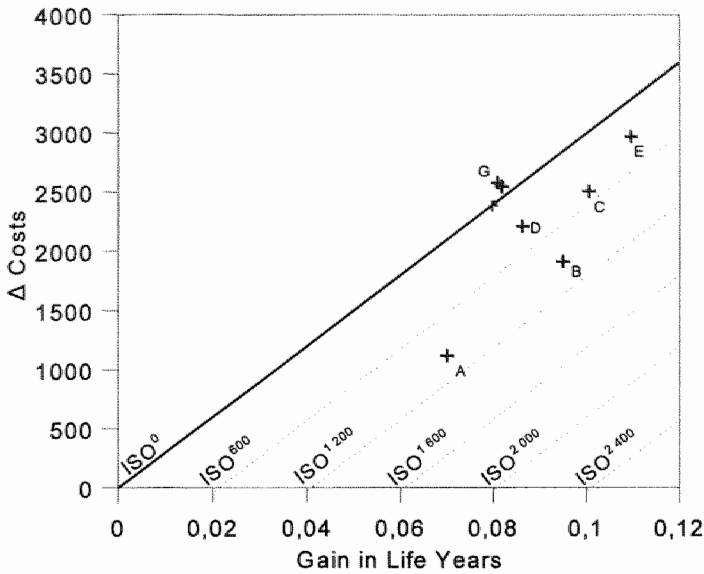


FIGURE 4. Costs and effects of mammography screening. Isoquants based on CERmax = \$30,000 per life year saved

It is suggested in the study that it is relatively cost-effective to screen women between the ages of 40 and 50 years (alternative B) and that such an intervention should be considered. Probably, this conclusion is based on the average CER rather than on an incremental analysis. Regarding its incremental CER, this intervention is only optimal if CERmax is higher than about \$ 32,000 per QALY. For optimal decision making and to inform policy makers adequately, alternatives should be presented as 'mutual exclusive'. Extensions of programs to more costly alternatives should be analysed separately, provides a better understanding of the increase (or decrease) of costs and benefits.

Treatment of peptic ulcers

This example broadens the perspective to alternatives that combine cost savings with health gains. Furthermore, in this example the effects of the different strategies are not expressed in QALYs or life years saved, but in another effect parameter: disease-free periods (in weeks). The study produces the results in terms of the costs and benefits of eight drug treatments of peptic ulcers, which can be considered as incompatible alternatives.²² All results are compared with the alternative of maintenance ranitidine. The study identifies eight incompatible strategies, of which five combine cost savings with health gains. No optimal strategy is suggested in the study. However, Figure 5 reveals that because of strong dominance the number of relevant strategies can be further reduced to three.

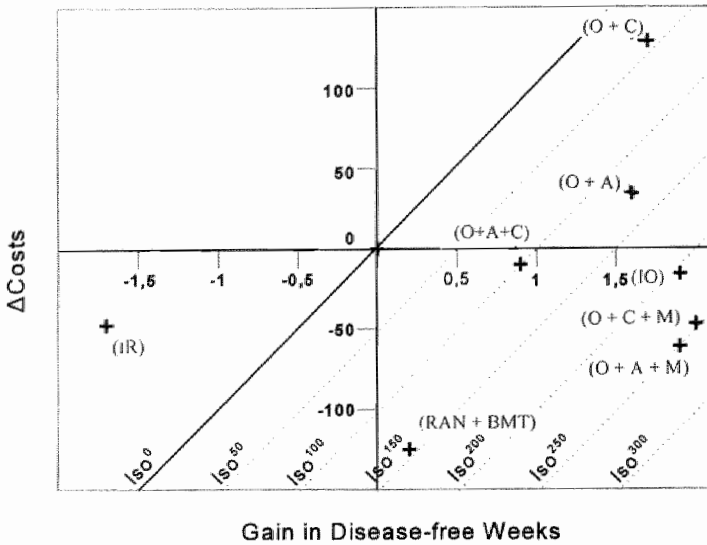


FIGURE 5. Costs and benefits of treatment of peptic ulcer. Isoquants based on CERmax = \$100 (monetary value of disease free week)

These include: 1) RAN-BMT (ranitidine-bismuth triple therapy), 2) O+A+M (omeprazole, amoxicillin and metronidazole therapy) and 3) O+C+M (omeprazole, clarithromycin and metronidazole therapy). These 3 alternatives together form the south-east boundary of alternatives. The isoquants in the figure are based on the monetary value of a disease free-period of 1 week which is assumed to be \$ 100. The figure shows that alternative B is the optimal choice under this condition. If the value of a disease-free week is lower than \$37.70, alternative A has the highest net value and is optimal. Alternative C is selected as the optimal strategy if the CERmax exceeds \$140. Although it may not be easy to put a value on a unit such as a disease free week, this should not be a reason for neglecting its importance. At the very least the impact of various values should be made clear to decision makers.

The Rittenhouse dilemma

The approach developed in this paper can also be used to deal with the 'Rittenhouse dilemma'.^{7,18} In this dilemma, it is recognized that it is sometimes difficult to determine on theoretical grounds whether a certain element should enter the denominator of the CER (as a negative benefit) or the nominator of the CER as a cost factor. This choice will influence the ratio and should therefore affect the attractiveness of the intervention to be studied. The resulting CER, and therefore possibly the conclusion could be different. 'Thus, the two methods are substitute but not equivalent methods, and neither has any particular claim for accuracy or

legitimacy. More important is the fact that other interventions may now be ranked superior to or inferior to this one, depending on how the quality adjustment is carried out.⁷ This dilemma is illustrated using the example shown in Table 3. It can be seen that the CER with regard to intervention B changes as a consequence of the procedure applied. (Intervention A, showing no change in CER, is a special case, as is discussed below).

TABLE 3. Two methods for treating negative health effects in CER (negative health effect of 0.048 QALY equals \$50⁷)

	Intervention A	Intervention B
Cost	\$ 1,000	\$ 10,000
Health Gain in QALYs	0.952	0.952
CER (treated as cost factor)	\$ 1,050 (= \$ 1,000/ 0.952)	\$ 10,500 (= \$ 10,000/0.952)
CER (treated as benefit factor)	\$ 1,050 (= \$ 1,050/1)	\$ 10,050 (= \$ 10,000 + \$ 50)/1

As noted, the distinction between compatible and incompatible alternatives is vital when choosing optimal alternatives. For incompatible alternatives, the attractiveness should be based not on average CERs but on incremental CERs - which can be considered as identical with the net value. Rittenhouse recognizes that net value is not affected by treating a benefit as a cost factor and vice versa. Therefore it can be concluded that the attractiveness or priority sequence of the (incompatible) alternatives will not be influenced by the decision to treat an element as a cost or as a benefit. In the graph, this means that the alternatives remain on the same isoquant no matter which procedure is applied. This is further illustrated for interventions A and B in Figure 6. In the example it is assumed that \$50 is equal to a health utility decrease of 0.048 (measured, for example in QALYs), which means that the value of a QALY is \$1,041.67. This amount represents the CERmax in this specific case, and enables us to draw the isoquants. If a cost factor is treated as a negative benefit, interventions A and B move along the isoquants, but remain on the same isoquant. In this specific example intervention A lies on the zero-isoquant: the net value for this specific value of a QALY is nearly zero, which is why the CER remains constant. Intervention B appears to lie on a lower isoquant and has a negative net value for this value of CERmax. This shows that for the incompatible alternatives A and B, the attractiveness in terms of net value remain the same whichever procedure is applied.

In the case of compatible alternatives, however, the attractiveness is determined by average CERs, which - as has been shown - are subject to change depending on the procedure applied. If these alternatives are depicted in a graph it can be seen that, if some costs are treated as negative benefits, the CER of the alternatives change but their net values remain on the same isoquants.

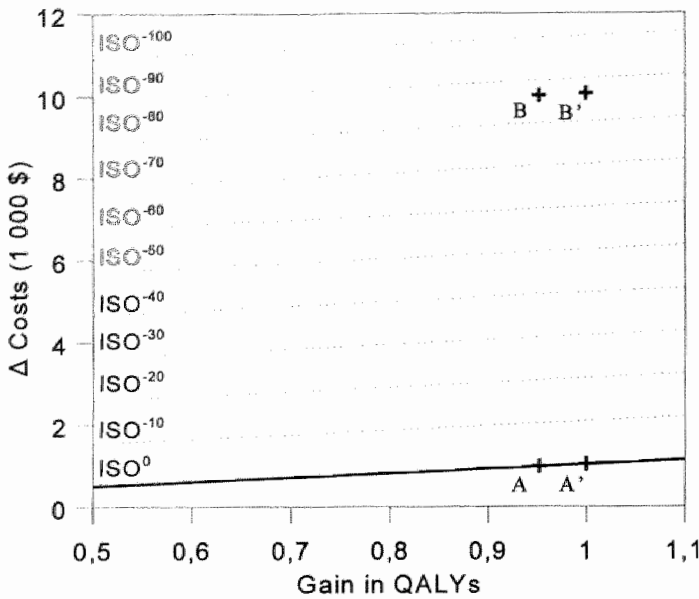


FIGURE 6. The Rittenhouse dilemma. Isoquants based on CERmax = \$1,041

Because these isoquants are located parallel to the isoquant with zero net value, this implies that an alternative can never pass this threshold. An attractive alternative, producing QALYs at a monetary cost below the CERmax, remains attractive, no matter how many costs are treated as negative benefits, and vice versa. The priority sequence of compatible alternatives might change, as Rittenhouse rightly concludes. However, the conclusion would be that all (compatible) alternatives with CERs below CERmax should be accepted, which in essence makes the sequence irrelevant. The inconsistency observed by Rittenhouse and others seems to have no impact at all in the case of incompatible alternatives. In compatible alternatives it is only the priority of the acceptable interventions which may change, and this does not influence the resulting resource allocation.

CONCLUSIONS

Economic evaluations are supposed to provide information about the relative attractiveness of (new) technologies in order to improve decision making about the allocation of scarce resources. The results of economic evaluations are often published in the form of ratios, expressing the costs involved to produce one unit of

effect. Often these results are accompanied by concluding remarks comparing these ratios with those of established technologies.

The present analysis reveals that policy decisions may be seriously impaired by misinterpreting cost and effects stemming from various alternatives. The distinction between compatible and incompatible alternatives as well as the impact of the CERmax is not always recognised. The graphic approach, as suggested in this paper, may be an aid to a more comprehensive consideration of options. The value judgments necessary for policy decisions form an integral part of the method, which enables a detailed analysis of the influence of these values on the resulting allocation of scarce resources.

The more complex algorithms and optimization techniques can be replaced by a number of rules. These rules are defined by assuming constant marginal opportunity costs. We consider this assumption as reasonable regarding the limited impact of individual medical interventions on total health care budgets. In the case of compatible alternatives, all alternatives with a CER below the CERmax should be accepted. For incompatible alternatives, the alternative located on the isoquant that represents the highest net value should be regarded as optimal. For a mix of compatible and incompatible alternatives, the optimal alternative within each subset of incompatible alternatives should first be determined, and then this optimal alternative should be considered as one of the compatible alternatives. Finally, all alternatives with CERs below CERmax should be accepted. It has been shown in this paper that this approach is also adequate in determining an optimal choice between several cost-saving alternatives.

In economic evaluations, alternatives are often compared with a 'zero' alternative, sometimes called 'doing nothing' or 'base case'. The extra costs and extra benefits of the alternatives, as compared to 'base case' alternatives, are often used to calculate so-called incremental CERs. In the case of compatible alternatives this procedure is correct and adequate, although we would prefer in such cases to use the term 'average CER' instead of 'incremental CER'. In cases of incompatible alternatives, the use of the term 'incremental' where 'average' should be applied may mislead decision makers, since it might appear that costs and benefits are being compared to the next most costly alternative.

Recently, guidelines have been formulated on how to report cost-effectiveness analyses.²³ On the basis of the present analysis, a number of additional recommendations can be formulated. These will assist researchers to report their findings in a way which allows decision makers to arrive at adequate conclusions.

(1) In addition to reporting the resulting CERs, the costs and benefits of each alternative should be presented separately. This is especially useful in case of

incompatible alternatives to allow incremental analysis, if not performed. A graphical presentation, and analysis, of the results, as proposed above, might be an attractive alternative.

(2) The impact of the CERmax should be made explicit in the analysis. Sensitivity analysis may be applied to reveal its influence on the conclusions, especially where the CERmax is difficult to anticipate (e.g. in the case of disease-free weeks).

(3) A mutually exclusive classification scheme should be used to present data on costs and effects. CERs for the composed programs can only be determined on the basis of this disaggregated information.

The essence of these recommendations is that researchers should present their results in a way that permits decision makers to derive correct conclusions regarding the use of medical interventions. A further requirement to bridge the gap between research and health policy is that more precise estimates should be acquired of society's maximum willingness to pay for unit of effectiveness (eg QALYs).

It is important to note that, by affecting the value of CERmax, the size of health care budget (implicitly) determines the optimal set of (alternatives of) medical interventions. Different societies, with different levels of wealth, could therefore have different priority rankings in health care. Choices regarding optimal alternatives of medical interventions should therefore carefully assessed before they are applied to e.g. less wealthy countries.

Finally, it is obvious that a resource allocation system based purely on cost-effectiveness does not exist in any country in the world, nor will it exist in the near future. It is increasingly recognized that decisions about resource allocation should not be based solely on efficiency criteria, but also on equity considerations. In that sense the allocation system based on CERs is to be considered as a rather theoretical framework, but of increasing importance. If the results of medical technology assessment are intended to improve the efficiency of the health care system, they should be presented in a correct way.

REFERENCES

- 1 Drummond, M.F., Stoddart, G.L., and Torrance, G.W. *Methods for the economic evaluation of health care programs*. Oxford: Oxford University Press, 1987
- 2 Weinstein, M.C., From cost-effectiveness ratios to resource allocation: where to draw the line? In Sloan, F.A, editor. *Valuing health care: costs, benefits, and effectiveness of pharmaceuticals and other medical technologies*. Cambridge: Cambridge University Press, 1995.

- 3 Thompson, M. *Decision analysis for program evaluation*. Beverly Hills: Sage Publications, 1982
- 4 Anderson, J. Policy space areas and properties of benefit-costs/utility analysis. *JAMA*, 1986; 255; 794-795
- 5 Steinbach, J.H., Hardison, W.G.M. Comparing health care options using the cost-benefit utility ratio. *JAMA*, 1986; 255; 747-748
- 6 Kamlet, M.S. *The comparative benefits modeling project: a framework for cost-utility analysis of government health care programs*. Washington: U.S. Department of Health and Human Services, 1992
- 7 Rittenhouse, B.E. Potential inconsistencies between cost-effectiveness and cost-utility analyses: an upstairs/downstairs socioeconomic distinction. *International Journal of Technology Assessment in Health Care* 1995; 11; 365-37
- 8 Baumol, W.S. *Economic theory and operations analysis*. Englewood Cliffs: Prentice-Hall, 1977
- 9 Birch S., Donaldson C. Applications of cost-benefit analysis to health care: departures from the welfare economic theory. *Journal of Health Economics* 1987; 6; 211-225
- 10 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
- 11 Mishan E.J. *Cost-Benefit Analysis: an informal introduction*. London: George Allen & Unwin, 1982
- 12 Keeler E.B., Cretin S. Uses of cost-benefit analysis. *Journal of Health Economics* 1987; 6; 275-278
- 13 Laupacis, A., Feeny, D., Detsky, A.S., Tugwell, P.X. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Canadian Medical Association Journal* 1992; 146; 473-481
- 14 Gafni, A., Birch, D. 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
- 15 Weinstein, M.C. Principles of cost-effective resource allocation in health care organizations. *International Journal of Technology Assessment in Health Care* 1990; 6; 93-107
- 16 Ament, A. *Optimal use of diagnostic tests: a decision analytical approach*. Delft: Eburon, 1989
- 17 Phelps, C.E., Mushlin, A.I. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *International Journal of Technology Assessment in Health Care* 1991; 7; 12-21.
- 18 Johannesson, M., Weinstein, M. On the decision rules of cost-effectiveness analysis. *Journal of Health Economics* 1993; 12; 459-467
- 19 Schechter, C., Rose, D. Fahs, M., Silver, A. Tuberculin Screening: cost-effectiveness analysis of various testing schedules. *American Journal of Preventive Medicine* 1990; 3; 167-175
- 20 Babazona, A., Hillman, A. Declining cost-effectiveness of screening for disease: the case of gastric cancer in Japan. *International Journal of Technology Assessment in Health Care* 1995; 11; 354-364
- 21 Lindfors, K., Rosenquist, J. The cost-effectiveness of mammographic screening strategies. *JAMA* 1995; 11; 881-884

- 22 Perras, C., Otten, N. Pharmaceutical Management of Peptic Ulcer Disease. *Technology Overview: Pharmaceuticals*, 1996, Issue 3.1 Canadian Coordinating Office for Health Technology Assessment.
- 23 Siegel, J., Weinstein, M., Russell, L., Gold, M. Recommendations for reporting cost-effectiveness analyses. *JAMA* 276; 1339-1341; 1996

CHAPTER 8

REAL WORLD DESIGNS IN ECONOMIC EVALUATION

Based on: Baltussen R.M.P.M.¹, Leidl R.M.², Ament A.J.H.A.¹ Issues in the design of economic evaluation: Bridging the gap between research and policy making (submitted)

1. From the Department of Health, Organisation, Policy and Economics, Maastricht University
2. From the Department of Health Economics, University of Ulm, Germany

SUMMARY

This chapter proposes a general framework to real world designs in economic evaluation. The chapter identifies the information that economic evaluation should provide to policy makers, so that they may make appropriate and effective decisions. First, policy makers need cost-effectiveness information that is both internally and externally valid. The latter aspect is often ignored and refers to how relevant the results are to the specific decision making context of the policy maker. Second, policy makers, like purchasers of care, may want assessments of the overall impact on budget and health following the adoption of an intervention. This requires a more aggregate approach than the traditional analyses that are typically individual-oriented. Against this background, this chapter examines the three main conceptual designs for economic evaluation: the use of RCTs, observational studies, and modeling. The chapter identifies a number of options for enhancing the informative value of economic evaluation studies. Of these options, the use of additional modeling and observational data seems to be the most promising. To address issues at the system level, disease modeling or public health modeling is suggested. The chapter concludes that there is potentially a large role for modeling in economic evaluation, but that in order to enhance its credibility, more attention should be paid to validity aspects.

INTRODUCTION

A wide array of tools is available for measuring the costs and effects of interventions in health care. However, the large variety of methods, and of their specific characteristics, make it very difficult for policy makers to assess their appropriateness for allocation decisions in health care. The 'gold standard' today is still the economic analysis carried out alongside the randomised controlled trial (RCT). However, confidence in this system is beginning to falter, as we question whether the results of experimental research provide us with adequate evidence for our economic decisions regarding health care. In particular, the limited generalisability of results from RCT-based economic analysis may seriously restrict its relevance for policymaking. In situations where RCT-based economic analysis is not practical or feasible, decision analytic models are used to synthesize the best available data. There is criticism that many of these models have a 'black box' feel about them, hampering the ability of policy makers to judge their appropriateness and usefulness for allocation decisions in health care.

In addition, purchasers of care, like budget holders and insurance companies in Europe or managed care organisations in the US, may want assessments of the full impact of new therapies being introduced in a disease or treatment area, rather than only a rank ordering of interventions by their average or incremental cost-effectiveness. Such assessments focus on the net impact of an intervention on the aggregate costs and outcomes of health care delivery for the treatment population. Such analyses can be carried out from the societal perspective for institutes representing society, but may also be more confined, e.g. in the case of insurance companies, to include only costs falling on the company.

Hence, policy makers need a critical assessment of the various methods of economic evaluation in order to judge their appropriateness for policymaking. To foster effective policymaking, the shortcomings of the methods should be addressed and links for adjustments should be incorporated in future economic evaluation research. Policy makers need economic analyses that provide a somewhat broader perspective than the individual level. They have to assess the cost-effectiveness specific to their decision context and the overall or net impact that adoption of a new intervention will have on costs and outcomes.

This paper begins with an analysis of the informative value of traditional economic evaluation at the individual level. From a conceptual basis, three different approaches to economic evaluation are distinguished: economic evaluation carried out alongside RCTs, economic evaluation based on observational studies, and the modeling approach to economic evaluation. Next, various aspects of appraising the net impact of an intervention at the system level are discussed. The paper concludes with a proposal of a framework which encompasses the various issues related to

transposing cost-effectiveness estimates from the individual level to the more aggregate system level.

This paper does not share the view assumed by many contributions to economic evaluation, where study results are supposed to support a fictitious, omnipotent decision maker who allocates health care resources in order to maximize health for the whole population. Rather, we recognize that health care is characterised by pluralism: various actors (providers of care, purchasers of care) independently decide on planning in health care. Therefore, the requirements of economic evaluations have to be defined according to the specific needs and decision contexts of the various policy makers. Analyses at the system level, as designed to support policymaking, must therefore focus on the differing perspectives of the various policy makers.

ECONOMIC EVALUATION AT THE INDIVIDUAL LEVEL

Ideally, the cost-effectiveness information of interventions intended to foster policymaking in health care should be valid internally as well as externally. Internal validity concerns the extent to which an instrument measures what it is intended to measure.¹ The issue of internal validity has received much attention in research concerning the methodology of economic evaluation. There are questions as to whether costs and outcomes are properly measured and valued.² Less attention has been focussed on the issue of external validity. External validity refers to the extent to which the results of a study can be generalized to other populations, conditions, or times.³ In this paper, external validity is also labelled as real world relevance. This concept implies that the results from economic evaluation must be adapted as closely as possible to the real world, where patients are actually treated. This real world has been described by Spilker as:

"the manner in which a particular medicine or treatment is usually used, given, or administered (which) includes an incredible diversity and complexity of factors that may interact with any medicine or treatment prescribed and influence a patient's clinical outcome".⁴

In this section, the three basic approaches to economic evaluation are discussed in terms of their internal validity and their external validity, or real world relevance.

RCT-based analyses

Because of the randomization process, an RCT-based economic analysis avoids the influence of patient and practitioner attitudes towards the definition of patient groups and allows an unbiased comparison of groups of patients undergoing the interventions under study.⁵ Regarding these characteristics, RCTs can be considered as the optimal vehicle for economic analyses as far as the issue of internal validity is concerned.

There are serious doubts about the role of RCTs in relation to the issue of external validity. Although it is argued that RCT-based analyses inform us in a very precise situation with great internal validity, they may not provide the answer to our real world question. That is, RCTs may give a precise answer to the wrong question.⁶ Four classes of biases can be distinguished which reduce the applicability of RCT-based analysis to the specific decision making context of a policy maker.³ The first relates to the limited scope that is often associated with RCTs. Economic analysts may adopt a reductionist stance which does not fully account for factors like adverse effects and interactions with other diseases, and this may bias the cost estimation involved. For example, assuming smoking is a risk factor for both coronary heart diseases (CHD) and lung cancer, a reduction in smoking will reduce the incidence of both CHD and lung cancer.⁷ An RCT that focuses solely on the costs of an intervention aimed at CHD, and that ignores any savings stemming from interactions with lung cancer, will lead to biased estimates of real world costs.⁸

The second class of biases involves the methodological aspects of RCTs and questions the real world relevance of the tools with which economic analysts estimate costs in connection with RCTs. RCTs often have an artificial design, which means that the cost figures obtained from an RCT may very well deviate from the costs that would occur in the real world. For example, more safety and efficacy tests are usually performed in an RCT than in normal medical practice⁴, which may cause cost estimates based on RCTs to be higher than might be the case in actual medical practice.⁹

The third class of biases relates to the generalizability of the results of an RCT-based economic analysis. The costs of treatment may well vary depending on which patients are treated, which physicians treat them and where the treatment takes place. For example, RCTs are typically conducted in teaching hospitals, which may yield cost estimates that are relatively high compared to the costs which would be incurred at non-teaching hospitals. Hospital learning effects may also play an important role. The first heart transplant procedure performed at a hospital has been predicted to cost \$81,297 in the US. For the twenty-fifth case, the predicted cost is only \$35,352, and for the fiftieth case the cost is \$25,458 (31 percent of the original cost).¹⁰ The implication is that, over time, costs in the real world may decrease, so that an intervention which is considered to be cost-ineffective based on the figures from an RCT may very well become cost-effective after wider adoption. This is a highly relevant issue when considering technologies in a very early phase of their life cycle.

The fourth class of bias refers to the definition of results from economic evaluation. Clinical research tends to define endpoints as the state of affairs at a fixed follow-up time, such as the n-year survival rate difference, whereas adequate policymaking requires endpoints which represent the cumulative lifetime effect. Furthermore, RCTs often report intermediate clinical endpoints like changes in blood pressure or

changes in cholesterol level, while quality-adjusted life years can be looked upon as a good indicator for final outcome in cost-effectiveness analyses.

Observational studies

Observational studies register the actual costs and outcomes of the groups being compared. Case-control and cohort study designs are types of observational studies commonly used to make direct comparisons of costs where experimental designs are not feasible. In case-control studies, the costs and outcomes of a group of individuals with the characteristics of interest (cases) are compared with the costs and outcomes of individuals without that characteristic (controls) regarding their previous exposure to some factor. In cohort studies, individuals are classified according to their exposure or non-exposure to a disease or intervention and followed forward to track the costs and outcomes.

The main advantage of the use of observational data compared with the collection of data through RCTs is that observational data typically represent ordinary medical practice rather than very selective patient populations or settings. However, there is a major drawback in that observational data sources contain biases that often render the results invalid. The validity of observational studies rests heavily on the degree to which the populations being compared are truly equivalent in all relevant aspects. For example, the costs and effects of pneumococcal vaccination have been estimated in the US by means of a cohort study of 762 vaccinated and 1,161 randomly selected unvaccinated age-sex matched persons using medical and pharmaceutical claims. The study concluded that the vaccine is efficacious and cost saving for various patient groups. However, the risk for pneumonia in the vaccine's group prior to vaccination substantially exceeded that of the comparison group, which may have confounded the results.¹¹

In the US, Patient Outcomes Research Teams (PORTs) have done extensive work on the use of observational data to address the need for information on the comparative costs and effects of alternative treatments. To test directly whether it is possible to use administrative data to make valid comparisons among technologies, the results of observational studies have been compared with the results of clinical trials. The results were inconclusive: "Sometimes non-randomised studies will tell you the right answer, sometimes the wrong answer, and there is no way to tell the difference without an RCT to determine the true answer".¹² Moreover, they found that data from existing claims and other databases are only rarely sufficiently clear and seldom show differences of a sufficiently large magnitude to be useful in drawing conclusions about relative (cost)-effectiveness.¹² These findings hamper strategies to rely solely on observational studies when the cost-effectiveness for a real world target group is to be determined.

Modeling designs

Another option for finding information on cost-effectiveness in situations where RCT-based economic analysis is not practical or feasible, is to use decision analytic models. These models are often used to synthesize the best available data. Input data for models can be derived from a variety of sources with varying degrees of scientific rigour. Effectiveness data can be RCT-based, as applied e.g. in the cost-effectiveness study of the treatment of hypertension¹³, but can also be derived from the less scientifically rigorous evidence of observational studies, as applied e.g. in the cost-effectiveness study of mammographic screening strategies¹⁴, or can even be based on expert opinion. Evidence-based medicine approaches and databases can provide a useful synthesis of knowledge about effectiveness and a good basis for economic modeling.

Costs data, as an input to models, is often a mixture of data on resource use (e.g. length of hospital stay) and data on the assigned (unit) cost or price values (e.g. costs per hospital day). Measurement of resource use can be based on RCTs, but in practice is often based on less rigorous scientific assumptions regarding supposed resource consumption, supported with observational data.¹³ Data on the values assigned to resource use can be retrieved from a variety of sources; the choice does not affect the internal validity aspects, but is merely related to the external validity of the model.

The use of models is surrounded by debate. It is argued that the inability to verify underlying assumptions (e.g. regarding resource use), and the inability to assess the strength of conclusions based on them, should lead to serious questions about the internal validity of study findings.¹⁵⁻¹⁷ However, there is common agreement among health economists that models that integrate scientifically rigorous data from different studies are fit to be used for policymaking.^{6,17} There have also been studies showing that decision model results can be validated and adapted once information from RCTs becomes available, although their primary purpose is to inform decisions before the trial data emerges.¹⁸

There are also situations where scientifically rigorous data is not available, as it may not be possible to conduct RCTs due to ethical constraints, time pressures, political considerations or costs. The use of less scientifically rigorous models has been allowed by some.¹⁹ They argue that in medical care, good prospectively derived clinical and economic data is often lacking and decisions need to be made with insufficient information. They refer to this as the choice between saying 'we do not know' and 'what is the best way of bringing together what little we know to best answer the necessary questions'. However, other authors argue that decision models are very often used when data are poor, or are based on assumptions that come from clinicians. This apparently contradicts the reason to model, i.e. to replace the implicit

judgements of clinicians.⁶ They argue that because these models lack internal validity, they are not suitable for policymaking.¹⁵⁻¹⁷

One important virtue of models relates to the issue of representativeness. As modeling approaches may apply observational data that may generalize over the site of care delivery (e.g. teaching versus non-teaching hospital), and patient-level characteristics (e.g. age, number of diseased vessels), the results can be considered more relevant to daily practice, in comparison with RCT designs.

THE NET IMPACT AT THE SYSTEM LEVEL

Valid economic analysis at the individual level may not fulfill all of the needs for adequate policymaking; this may also require information on the net system impact stemming from the adoption of an intervention. The budget impact - the total costs falling on an institution or a funding agency - constitutes the constraint by which policy makers are limited in funding interventions and may therefore have decisive value for any intervention.²⁰⁻²²

In a study on the possibilities for incorporating economic evaluation in the control of the diffusion of interventions, Rutten and Drummond identify nine different options.²³ In all of these options, the budget impact of the intervention to be adopted (either falling on the government, the purchaser or the provider) most likely affects the definition of the regulation. The government's budget is affected in the case of planning specialist facilities or specific 'big ticket' interventions, in decisions regarding in- or exclusion from public reimbursement and in decisions regarding reforming payment schemes for health care institutions. Also, policy decisions on changing payment schemes for health care professionals, on developing schemes for reviewing medical auditing and utilization, and on introducing co-payments for service users affect the governmental health care budget as well. In order to make adequate use of those options to control the diffusion of interventions, policy makers should have a clear indication of the consequences of implementing proposed regulations - consequences on resources used and on those left for other interventions. This has been acknowledged, for example, in discussion on reimbursements for a therapy for sleep disorders; considering the substantial prevalence of this disorder, introduction of the therapy '...would almost certainly over-stretch health care resources in most countries.'²⁴

Another regulation, applying cost-effectiveness information to encourage clinical budgeting within institutions (especially hospitals), mainly affects the institutional budget. Within the institution, an indication of the total costs stemming from the regulation (i.e. from the adoption of an intervention) is indispensable in the process of planning and budgeting for coming years. Finally, purchasers of care may use

economic evaluation in the context of competitive arrangements in health care. In their decisions on placing contracts for an intervention, purchasers could use cost-effectiveness information that is most likely to include estimates of the total costs to be met. These costs are part of a total cost picture that constitutes essential information in the process of setting insurance premiums. In the US, the question central to many managed care organisations is the system impact of the introduction of a new intervention on the cost and outcomes of treatment delivery.²²

Complicating the issue further is the need of policy makers to be informed on the distribution of costs over time. A patient-based economic evaluation leaves the policy maker with only the present values of costs and effects for a single patient at a given time. This raises problems for judging the budgetary impact of allocation decisions in the case of e.g. periodic budgets.²² It has been suggested that the choice to fund an intervention that features cost savings now but incurs more expensive treatments later should be combined with other treatments that make the best use of the remaining resources.²¹ Consequently, the differential timing of costs stemming from the adoption of interventions should be considered in order to foster an optimal allocation over time at the population level.

THE THREE-STEP APPROACH

A three-step approach is suggested to enhance the informative value of traditional economic evaluation (Figure 1). As a starting point, or as a basic requirement, economic evaluations should be based on an approach that renders internally valid results. As a second step, the internally valid results should be critically considered in the context of their real world relevance. As a third step, these results are to be extrapolated to assess the net impact on the system level. The figure indicates the approaches which can be used to achieve the successive steps.

Step 1: Deriving internally valid results

To successfully distinguish between the cost-effectiveness of interventions under scrutiny, the first condition economic evaluation has to meet is internal validity. As noted above, the RCT is the only available instrument in health care research that generates internally valid results, and therefore constitutes the only acceptable basis of any economic evaluation study. Effectiveness parameters should be estimated by this vehicle. Also, patient specific data on resource utilization, i.e. the physical quantities, should be obtained through an RCT. The assignment of values or (unit) cost figures to these quantities of resources can be applied by so called 'micro-costing' or 'gross-costing'.² Micro-costing refers to the valuation of each resource consumed by the patient. What is sought in using gross costs is a satisfactory estimate of the 'typical' costs of the service.

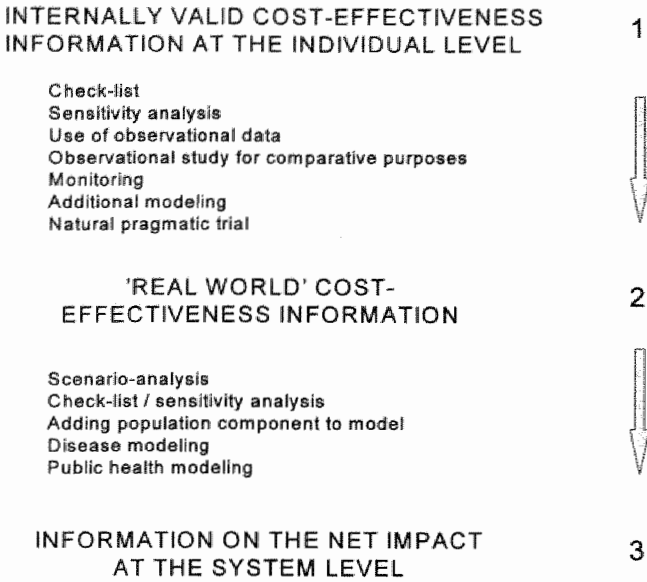


FIGURE 1. A three-step approach to enhance the informative value of economic analysis

Both approaches may render internally valid results, while the latter has the benefit that its results are more generalizable to other settings.

RCT-based economic analyses comply with the requirements for internally valid economic evaluation studies. Models that are based on RCT-evidence on effectiveness and resource use, but that have otherwise assigned costs, also meet the requirements. Economic observational studies, and economic models that are not based on RCTs for data on effectiveness and resource use, cannot be accepted as rigorous economic evaluations.

Step 2: Warranting the real world relevance of results

This second step concerns the concept of real world relevance and implies that results from economic evaluation must be adapted as closely as possible to the real world, where patients are actually treated. This step implies the consideration of context-specific factors like the various procedures considered in the economic evaluation, and specific physician, hospital and health care system characteristics. Also, indications of possible levels of non-compliance and their impact on costs are important in this regard. Furthermore, future developments that may have an impact on the cost-effectiveness of the intervention under scrutiny should be considered. An analysis at the more aggregate level may also be required to include those factors

that cause the cost-effectiveness to vary with the number and type of patients who are medically eligible for an intervention.

This second step is relevant to modeling approaches to economic evaluation, but especially for RCT-based economic analyses because of their artificial character. Similarly, many features of this step also apply when study results from other countries are to be considered in decision making. There are various approaches to increase the real world relevance of cost-effectiveness estimates.

Check-list. One option for enhancing the real world relevance of economic evaluation is to encourage researchers involved in economic evaluation to report their findings in such a way that policy makers can make an assessment of the relevancy of these results in their own context. Working through checklists is one approach to this problem.³ A checklist can highlight various issues that researchers and policy makers should take note of in realizing a high real world relevance of results from economic evaluations. In the checklist approach, analysis of the results, interpretation, and the forming of conclusions are left (almost) entirely to the policy maker. Recently, a checklist analysis was performed to assess the internal and external validity of trial results for deciding on the use of two alternative treatments for acute myocardial infarction within the context of a hospital. The analysis was found to be characterised by great uncertainty on the cost-side of the analysis in particular, stemming from the mutual influence of various interfering factors.²⁵ Accordingly, a checklist approach alone may not be sufficient to guide decisions.

Sensitivity analysis. The application of sensitivity analysis is closely related to the checklist approach. By applying sensitivity analysis to those factors that seem to most affect cost-effectiveness in the real world, and by screening the range of parameters relevant to the specific decision context, policy makers can assess the specified impact. This requires not only one-way sensitivity analysis, but also multi-way sensitivity analysis to show the mutual impact of alternative values for several factors. Although such analyses may contribute to the informative value of the economic evaluation in question, the specific context of the various policy makers may be difficult to address in detail.

Use of observational data. As noted above, observational data describe ordinary medical practice and may therefore encompass a high real world relevance. The use of observational data to assign (unit) costs on resource use has been advocated to substitute cost measurements alongside RCTs.^{2,17} The observational data should be considered carefully as to whether the characteristics of the patients involved are comparable to those enrolled in the RCT. The economic evaluation of thrombolytic therapy is an example of a study in which observational data are applied to an RCT. In this study, estimates drawn from observational databases were used instead of estimates of unit costs as observed in the trial.²⁶

Observational study for comparative purposes. One may also use observational studies to compare the results reported in a clinical trial during the investigating phase of a technology, and the results that occur when the technology is in general use. Such comparisons might illuminate differences in the efficacy of a technology under strict conditions and the outcomes borne out in widespread use. However, as noted, the issue of internal validity is a critical aspect in the use of observational studies, and should be carefully appraised in this context before strong conclusions can be drawn.¹²

Monitoring. Various factors may cause the cost-effectiveness of an intervention to vary over time. The most important determinants in this context, as indicated by the sensitivity analysis, can be closely monitored to assess their future impact. For example, in the economic evaluation of pneumococcal vaccination in the Netherlands, the average length of hospital stay for pneumococcal pneumonia appeared to be an important contributor to costs. As the average length has decreased from 16.2 days in 1988 to 15.1 days in 1993 (entire population), a further decrease can be expected, which may cause the cost-savings of pneumococcal vaccination to diminish compared to what was estimated at the time of the study.²⁷

Additional modeling. As the issue is complex, there may be a need for a more explicit approach to improve the real world relevance of results from economic evaluation. Building on the information provided by the checklist approach and the sensitivity-analysis, and considering the context-specific information, additional modeling can be used to achieve results that are more informative to policy makers. Additional modeling facilitates the simultaneous consideration of the effects of different factors that may all affect the cost-effectiveness of an intervention. Additional modeling also permits the simultaneous application of data on cost figures and effects from different sources that, in total, may be more relevant to the real world than those estimated in an RCT. For example, evidence-based medicine can provide a useful synthesis of knowledge about effectiveness and a good basis for economic modeling.²⁸ The flexibility of additional modeling also implies that (unit) cost estimates relevant for one context can, at least in theory, be replaced by cost estimates relevant to other (decision making) contexts.

Sometimes additional modeling is applied to supplement RCT-based economic analyses to express units of effects that are suitable for decision making. For example, additional modeling was found necessary in an economic evaluation concerning alternative treatments against HIV. As the trials under scrutiny used delay of progression to further stages of the disease and to death as clinical endpoints, these data had to be translated into lives saved to arrive at cost-effectiveness ratios that could be used for policymaking.²⁹

Natural pragmatic trials. The above approaches can only be applied as additional instruments to traditional economic evaluation. Some authors suggest using a completely different approach to economic evaluation: that of natural pragmatic trials. These should have both the internal validity (freedom from bias) associated with RCTs and the greater external validity (real world relevance) typically associated with models.^{6,17} The aim is to evaluate the (cost-) effectiveness of an intervention in real world circumstances that will prevail once the intervention is in routine use. Ideally, these large scale trials take place in a normal setting (thus excluding protocol driven costs), involve standard treatment as comparator, have an extended period of observation, a wide range of outcomes, include a follow-up on all patients, and have an experimental design. A drawback is that, albeit these trials do indeed include a number of real world aspects, their generalisability remains limited to the (geographical) setting in which they are organised. Moreover, due to high costs, ethical aspects and a number of practical requirements such as routine documentation systems, their application will necessarily be limited to a small number of interventions.

Step 3: Assessing the net impact at the system level

In the third step, assessing the net impact of introducing a new intervention, the net costs and outcome changes associated with introducing the new intervention must be estimated. If the intervention will replace other interventions, costs and outcomes associated with the current situation must be assessed as a starting point. This requires an overview of treatment practices for the disease area, including information on the indication criteria of these practices, patient streams, and related costs and outcomes. Subsequently, the net costs and outcome changes associated with introducing the new intervention must be estimated. The degree of substitution must be assessed: to what extent will patients switch from the current therapy to the newly introduced intervention. The indication criteria of the intervention under scrutiny are especially important; this determines the number of patients eligible for the intervention and its diffusion path.

If an intervention is aimed at patients who have not been treated previously, incidence and prevalence statistics - eventually adjusted for disease stage - seem to be most important in this context, but must be carefully assessed in light of potential complicating factors. For example, in the case of an intervention focussing on several age-cohorts, it should be considered that the treatment of a relatively young cohort of patients affects the incidence of disease in older cohorts. This affects the number of patients eligible for the intervention and consequently the budget-impact.

Complementary to the epidemiological factors, the quantitative assessment of patient streams involved is complicated further by a bewildering array of socio-economic, demographic, medical and technological developments. Tugwell notes that availability, awareness and acceptability are factors that determine the number of

people using an intervention.³⁰ The aging of the population is the most prominent example of demographic change, and can be expected to lead to an increased use of those interventions aimed at the elderly. This has been illustrated, for example, by the structure of acute inpatient care in Germany.³¹ The gradual diffusion of an intervention to a broader range of indications, following its introduction, is a known medical phenomenon. This impact on the utilization of an intervention has been described in the case of laser treatment.³² It is perhaps most difficult to assess the influence of rapid technological progress in health care. Alternative treatments may be developed that cover parts of the indication-criteria for the intervention under study, and consequently parts of the total number of patients eligible for that intervention.

In regard to this third step, there are a number of possible approaches to translate results, as derived from economic evaluation at the individual level, to the more aggregate level. Scenario analysis is a useful tool within these approaches. A scenario describes how the present situation for a part of the society may be changed in a particular way in the future by a series of possible developments. The aim of creating the scenario is to obtain greater insight into the underlying mechanisms and the ways in which these can be influenced.³³ Using 'if-then' reasoning, large event spaces for wide arrays of assumptions can be detailed to cope with multiple uncertainties. Scenario-analysis enables the researcher to distinguish various possible diffusion paths of an intervention following its introduction, including its impact on total costs at the system level. Scenario techniques can be applied when reviewing checklists, but also to more complex epidemiologic models, as discussed in the following section.

Checklist. Just as checklists can be used to facilitate the real world relevance of results, they can be applied to critically assess the budget consequences of adopting interventions. Such a checklist would focus on system level issues, like substitution patterns and patient streams. This approach does not permit the simultaneous consideration of various issues and may therefore be less suitable for a full quantitative assessment of a budget impact.

Adding a population component to models. Simple modeling techniques, like decision-tree analyses, typically focus on a single intervention in single disease area, ignoring interactions with other interventions or with other diseases. By adding a population component to these models, the results can be extrapolated to the system level. This has been applied, e.g., in the model on the cost-effectiveness of pneumococcal vaccination in the Netherlands, in which the results of the analysis at the individual level are multiplied by the number of people eligible for vaccination in the various age groups.²⁷ However, as the primary aim of these models is to describe the process at the clinical level rather than that at the aggregate level, important

factors that are related to the latter (e.g. interactions with other diseases or other interventions) cannot be accounted for systematically.

Disease modeling. In order to examine the net impact at the system level of introducing an intervention, information is required on the costs and outcomes of other interventions in the same treatment area, as well as of the intervention under study. The heart disease policy model by Weinstein et al.³⁴ was a pioneer in this field. These models may be comprised of modules on the incidence of disease, on disease development, on patterns of costs of interventions, and on the morbidity and mortality effects of interventions. Although the models require a strong epidemiological component, they may very well be clinically based, as are e.g. models on HIV / AIDS.³⁵ In those models, the patient is the modeling unit, involving in-depth assessments of the clinical disease process, including the distinguishing of several stages of the disease. In merely epidemiologically orientated analyses, patient groups are the modeling unit. This type of analysis has been applied, for example, in a study on the cost-effectiveness of voluntary counselling and testing of hospital inpatients for HIV infection.³⁶

Public health modeling. Modeling across the whole spectrum of diseases, including costs and effects of all relevant interventions, is one step beyond disease modeling. This very broad system of modeling, which can be labelled public health modeling, facilitates analyses of all interactions of the intervention under scrutiny with other interventions and other diseases (e.g. externalities). The Dutch Integrated Public Health Model (IVM),³⁷ into which disease-specific models on CHD, diabetes mellitus, HIV, influenza, lung cancer, and rubella are integrated, is one example. Although this model currently focuses only on health outcomes, future versions will also include economic modules. With such a modeling design, the full epidemiological and economic process of an intervention into a disease can be described at the system level.²¹

Analyses at the system level may take different forms as the need for a policy maker in health care insurance will be different from that of a policy maker representing society as a whole. The former requires the inclusion only of costs and outcomes relevant to the insurance company, and will disregard e.g. co-payments of insureds. Any interactions with other diseases and interventions are relevant to the insurance company only if they interfere with their budget or with the health outcomes of their insureds. Most likely, a complex modeling technique like public health modeling will not be the preferred approach in this context. On the other hand, policy makers representing society require an analysis from the societal perspective, which should ideally include all possible changes in resources and outcomes following the adoption of an intervention. An adequate analysis may require complex techniques such as public health modeling. However, for all analysis at the system level, either from the

societal perspective or from a more confined perspective, reason should be applied to determine the detail of the analysis.

EXAMPLE: APPLYING THE THREE-STEP APPROACH

In a case study, an RCT regarding benign prostatic hyperplasia (BPH) has been assessed for its informative value to decision makers.³⁸ The three-step approach has been applied to increase the real world relevance of the RCT's results and to assess the net impact at the system level. The RCT compares two therapies in terms of cost-effectiveness: the innovative cost-saving transurethral microwave therapy (TUMT) versus the standard treatment - transurethral resection of the prostate (TURP). By additional modeling, cost estimates at the individual level are corrected for the fact that the RCT is conducted in an academic hospital encompassing large overhead costs while TUMT is expected to be performed in general hospitals as well. Furthermore, capacity utilisation of the equipment in daily practice has been assumed to be higher than the ~ 25% observed in the RCT. Estimates of these real world costs turned out to be ~ 28% lower compared to those obtained in the RCT.

The budget impact stemming from the introduction of TUMT has been assessed by applying scenario analysis. Several scenarios of the future costs of the treatment of BPH are assessed using real world cost estimates and estimates of the number of patients eligible for TUMT in the future. The latter involves not only an analysis of patients who would otherwise have been treated for BPH but includes also a consideration of patients who - because of the less invasive character of TUMT - now present themselves for treatment. Various explicit assumptions regarding demographical, epidemiological and technological developments resulted in scenarios that encompass savings of up to US\$ 8 mln (1997 values).

This study has given policy makers important information that may have impact on the policymaking process. The analysis revealed that, because of relatively low annual equipment costs, the number of centres through which TUMT is provided affects total costs at the system level only slightly, thereby indicating no need for a strong governmental influence. Another important insight is that a possible increase in the number of men seeking BPH-treatment following the introduction of TUMT will be a major determinant of the total costs of BPH at the system level. This study has emphasised the feasibility and importance of additional modeling beyond the results from RCT-based economic analysis. The feasibility of system level analyses was somewhat limited due to a shortage of adequate epidemiological information.

DISCUSSION

This paper has made the following points. It has been shown that the three main conceptual approaches to economic evaluation, RCTs, observational studies and models, all have benefits and deficiencies as vehicles for providing cost-effectiveness information to policy makers. Each approach has specific drawbacks which makes it difficult to rely on any of them solely. Furthermore, the need for information on the net impact at the system level on costs and outcomes following the adoption of an intervention is identified. An analytical framework is proposed for enhancing the informative value of economic evaluation studies.

Table 1 summarises the above discussion by showing the data to be collected in an adequate economic evaluation. It has been noted that, as a starting point, any economic evaluation study should be based on RCT-evidence, at least as far as the measurements of the effect-parameters and the physical quantities of resources (resource use) are concerned. Model designs comprising unit cost assignment from observational data are acceptable frameworks for economic analysis.

TABLE 1 Data collection for economic evaluation

	from RCT	from other sources	can be either
Effectiveness	✓	-	-
Physical quantities of resources	✓	-	-
Unit costs / prices	-	-	✓
System impact	-	✓	-

There are a number of possible approaches to enhance the real world relevance of results. The provision of checklists and - closely related - the application of sensitivity analyses may fall short in quantitatively addressing the various issues. The use of additional modeling seems to be an adequate option to increase the real world relevance of results from RCT-based economic analyses. It may be especially useful to apply observational data on unit cost assignment. Unit costs or prices that are relevant to the context of the policy maker may be retrieved from outside the RCT. However, depending on the context of the policy maker, they may also be as observed in the RCT (Table 1).

Furthermore, as traditional economic analyses are typically confined to assessments at the individual level, the use of modeling - in extrapolating cost-effectiveness results, disease modeling, or public health modeling - is indispensable in addressing issues at the system level. For this level, it is clear that data must be added from outside the RCT (Table 1). Scenario analysis may be a useful tool for exploring different paths related to the adoption of an intervention. The choice of a specific

methodology depends on the level of detail needed to answer the research question. Whereas fund holders will consider only interventions and diseases that involve changes in overall costs to them, institutes representing society may want to cover the whole spectrum of interventions and diseases. Analyses at the system level will typically be characterised by a lack of (reliable) information. Estimates on the costs and outcomes of interventions in a certain treatment area will be available only incidentally as the field of economic evaluation has not yet been applied on a widespread basis. Furthermore, there will be a high degree of uncertainty with respect to epidemiological factors that may affect cost-effectiveness estimates and the impact at the system level following the adoption of an intervention.

This uncertainty contrasts sharply with the large amount of detail that is often associated with economic evaluation in general, and with RCT-based economic analysis specifically. This shortage of information may very well hinder adequate allocation decisions in health care. Research on economic appraisal should therefore also focus on providing accurate information encompassing these factors, rather than on detailed information from economic evaluation alone.

This paper has established a framework for assessing the scientific rigour of various approaches to economic evaluation. The requirements of an optimal economic evaluation are defined within this framework. However, this does not imply that other approaches to economic evaluation cannot be used for policymaking. As noted, RCTs may not always be feasible due to practical or ethical factors. The narrow focus of the *New England Journal of Medicine*,³⁹ for example, on the use of RCTs for economic evaluation bypasses this fact and would imply that cost-effectiveness information for coverage decisions on certain interventions would never be available. A more practical approach would suggest that decisions are made on the basis of the best available data.

The use of modeling in economic evaluation has been called an unavoidable fact of life.⁶ The increasing need for applying modeling has also been shown in this paper. This trend is given further support in recent work on the evaluation of cholesterol-modifying pharmacotherapy, where a need was identified for giving more weight to modeling designs, due to the low real world relevance of RCT-based economic analysis.⁴⁰ However, again in the *New England Journal of Medicine*, great concern was expressed regarding the use of (additional) modeling in economic evaluation: 'Bias can compromise even original scientific studies, but we believe that the opportunities for introducing bias into economic studies are far greater given the discretionary nature of model building and data selection in these analyses'.³⁹

The US Food and Drug Administration⁴¹ argues that claims about cost-effectiveness should be substantiated by adequate and well-controlled studies and that models to provide estimates of pharma-economic parameters should be used only when it is

impractical or impossible to gather data using adequate and well-controlled studies. The Australian Government argues that primary cost-effectiveness assessment should be based on data from the most relevant clinical trials, but that some limited use of additional modeling can be useful in transferring or generalizing from trial-based evidence.⁴² That is, they acknowledge that trial-based evidence is the most internally valid, but may be insufficient for decision making. Their pragmatic approach allows modelled economic evaluation in a number of circumstances as a matter of judgement (e.g. to link intermediate clinical outcomes, to extrapolate outcomes or to exclude protocol driven resources).

The controversy over models probably stems from the background of economic evaluation in health care. Although it originated in clinical medicine, which has a tradition of experimental research, it is also influenced by the field of economics, in which modeling has long been practised. Furthermore, the increasing role of economic evaluation, and especially of modeling, in defining health policy, has led to a more critical approach. In particular the role of pharmaceutical companies, who apply economic evaluation to influence reimbursement decisions, has recently received much attention.⁴³ Future research should therefore focus on the role of modeling in health care, and on enhancing the validity of modeling approaches to economic evaluation. This will add to the credibility of models in economic evaluation, and may therefore improve the allocation of scarce resources in health care. By identifying the need for modeling in health care, this paper has been a first step in this process.

REFERENCES

1. Portney L.G., and Watkins M.P. *Foundations of Clinical Research: Applications to Practice*. Connecticut: Appleton & Lange, 1993
2. Gold M.R., Siegel J.E., Russell L.B., Weinstein M.C. (editors) *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996
3. Baltussen R.M.P.M., Ament A.J.H.A., Leidl R.M. Making cost assessments based on RCTs more useful to decision makers. *Health Policy* 1996; 37: 163-183
4. Spilker B. *Guide to clinical trials*. New York: Raven Press, 1991
5. Fletcher R.H., Fletcher S.W., Wagner E.H. *Clinical epidemiology: the essentials*. Baltimore: Williams and Wilkins, 1988
6. Buxton M.J., Drummond M.F., van Hout B.A., Prince R.L., Sheldon T.A., Szucs T., Vray M. Modeling in economic evaluation: an unavoidable fact of life. *Health Economics* 1997; 6: 217-227
7. Hoogenveen, R.T., Verkley, H. and Jager, J.C. *Verkennde studie betreffende concurrerende en vervangende doodsoorzaken in verband met kwantitatieve gezondheidsdoelen* (Explanatory study concerning concurrent and replacing causes of death in relation to quantitative health goals), 1995. Report nr. 958606 004, RIVM, Bilthoven 1990 (in Dutch)

8. Garber, A.M. Competing risks and the cost-effectiveness of health interventions (Annual Meeting Program and Abstracts). *Medical Decision Making* 1989; 9: 319
9. Luce, B.R. and Elixhauser, A. Estimating costs in the economic evaluation of medical technologies. *International Journal of Technology Assessment in Health Care* 1990; 6: 57-76
10. Woods, J.R., Saywell, R.M., Nijhuis, A.W., Jay, S.J., Lohrman, R.G. and Halbrook, H.G. The learning curve and the costs of heart transplantation. *Health Services Research* 1992; 27: 219-238
11. Gable C.B., Holzer S.S., Engelhart L., Friedmann R.B., Schmeltz F., Schroeder D., Baum K.B. Pneumococcal vaccine: efficacy and associated cost savings. *JAMA* 1990; 264: 2910-5
12. US Congress, Office of Technology Assessment. Identifying health technologies that work: searching for evidence (OTA-H-608). Washington DC: US Government Printing Office, 1994
13. Kawachi I. and Malcolm L.A. Treating mild to moderate hypertension: cost-effectiveness and policy implications. *Journal of Cardiovascular Pharmacology* 1990; 16 (Suppl.7): S126-8
14. Lindfors K.K., Rosenquist C.J. The cost-effectiveness of mammography screening strategies. *JAMA* 1995; 274: 881-4
15. Drummond M. Cost-effectiveness guidelines for reimbursement of pharmaceuticals: is economic evaluation ready for its enhanced status? *Health Economics* 1992; 1: 85-92
16. Udvarhelyi I.S., Colditz G.A., Rai A., Epstein A.M. Cost-effectiveness and cost benefit analysis in the medical literature: are methods being used correctly. *Annals of Internal Medicine* 1992; 16: 238-244
17. Sheldon T.A. Problems of using modeling in the economic evaluation of health care. *Health Economic* 1996; 5: 1-11
18. Freedberg K.A., Hardy W.D., Holzman R.S., Tosteson A.N., Craven D.E. Validating literature-based models with direct clinical trial results: the cost-effectiveness of secondary prophylaxis for PCP in AIDS patients. *Medical Decision Making* 1996; 16: 29-35
19. OHE Briefing no 33. Towse A., Drummond M.F. (editors). Buxton M.J. in: The pros and cons of modeling in economic evaluation. London: Office of Health Economics, 1997
20. Laupacis A, Feeney D., Detsky A.S., Tugwell P.X. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Canadian Medical Association Journal* 1992; 146: 473-481
21. Leidl R. Some factors to consider when using the results of economic evaluation studies at the population level. *International Journal of Technology Assessment* 1994; 10: 467-78
22. Langley P.C., Martin R.E. Managed care guidelines for the economic evaluation of pharmaceuticals. *American Journal of Managed Care* 1997; 3: 1013-21
23. Rutten F., Drummond M.F. *Making Decisions about Health Technologies: A Cost-Effectiveness Perspective*, Institute for Medical Technology Assessment, York: University of York, Centre for Health Economics, 1994
24. Anonymous. Sleeping giant. *Lancet* 1991; 337: 1067

25. O'Brien B.J., Heyland D., Richardson S., Levine M., Drummond M.F. Users' guide to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? *JAMA* 1997; 277: 1802-6
26. Kalish S.C., Gurwitz J.H., Krumholz H.M., Avorn J. Cost-effectiveness of model of thrombolytic therapy for acute myocardial infarction. *J Gen Intern Med*. 1995; 10: 321-30
27. Baltussen R.M.P.M., Ament A.J.H.A., Leidl R.M., van Furth R. Cost-effectiveness of vaccination against pneumococcal pneumonia in the Netherlands. *European Journal of Public Health* 1997; 7: 153-61
28. Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices. *Lancet* 1997; 349: 126-28
29. Schulman K.A., Lynn L.A., Glick H.A., Eisenberg J.M. Cost-effectiveness of low-dose zidovudine therapy for asymptomatic patients with human immunodeficiency virus (HIV) infection. *Ann Intern Med*. 1991; 114: 798-802
30. Tugwell T., Bennett K.J., Sackett D.L., Haynes R.B. The measurement iterative loop: a framework for the critical appraisal of need, benefits and costs of health interventions. *J. Chron. Dis* 1985; 38: 339-351
31. Leidl R. Effects of population aging on health care, expenditure, and financing: some illustrations. In: Callahan D, ter Meulen RHJ, Topinkova (eds). *A world growing old: the coming health care challenges*. Hastings Center Studies in Ethics. Washington DC: Georgetown University Press, 1995; 49-61
32. Vondeling H. Diffusion of medical lasers in the Netherlands. In: Inzelt A, Coenen R (Eds.). *Knowledge, Technology Transfer and Foresight*. NATO ASI Series, 4, Science and Technology Policy. Vol. 8, Kluwer Academic Publishers, Dordrecht, Boston, London, 1996
33. S.T.G. (Stuurgroep Toekomstscenario's Gezondheidszorg). Scenario's in de volksgezondheid: inleiding in de methodiek van de STG (Scenarios in public health: introduction in the methodology of STG). Utrecht: Jan van Arkel, 1991
34. Weinstein M.C., Coxson P.G., Williams L.W., Pass T.M., Stason W.B., Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: the Coronary Heart Disease Policy Model. *Am J Public Health* 1987; 77: 1417-26
35. Postma M.J., Leidl R., Downs A.M., Rovira J., Tolley K., Gyldmark M., Jager J.C. Economic impact of the AIDS epidemic in the European Community: towards multinational scenarios on hospital care and costs. *AIDS* 1993; 7: 541-53
36. Lurie P., Avins A.L., Philips K.A., Kahn J.G., Lowe R.A., Ciccarone D. The cost-effectiveness of voluntary counseling and testing of hospital inpatients for HIV infection. *JAMA* 1994; 272: 1832-8
37. C.W.M. (Centrum voor Wiskundige Methoden) Documentatie Intergraal Volksgezondheid Model (Documentation Integrale Public Health Model). Internal report 005/94. Bilthoven: RIVM, 1994
38. Baltussen R.M.P.M., Wielink G., Stoevelaar H., van de Wilt G., Severens J., Ament A.J.H.A. (forthcoming). The economic impact of the introduction of TUMT in the treatment of BPH: a scenario analysis *World Journal of Urology* 1998; 16: 142-7
39. Kassirer J.P. and Angell M. The journal's policy on cost-effectiveness analyses (editorial). *New England Journal of Medicine* 1994; 331: 669-670

40. Morris S. A comparison of economic modeling and clinical trials in the economic evaluation of cholesterol-modifying pharmacotherapy. *Health Economics* 1997; 6: 589-601
41. F.D.A. Division of Drug Marketing, Advertising and Promotion *Principles for the review of pharmaceutical promotion*. Washington DC: F.D.A., 1995
42. Commonwealth of Australia, Department of Health, Housing and Community Services *Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee*. Canberra: Australian Government Publishing services, 1995
43. Hilman A.L., Eisenberg J.M., Pauly M.V. et al. Avoiding bias in the conduct and reporting of cost-effectiveness research sponsored by pharmaceutical companies. *New England Journal of Medicine* 1991; 324: 1362-5

CHAPTER 9

CONCLUSIONS

INTRODUCTION

The starting point in this thesis is the recognition of shortcomings in the traditional approaches to cost-effectiveness analysis. The RCT, often considered to be the gold standard for economic evaluation, falls short in addressing real world situations, and fails to address issues at the more aggregate level. Other approaches to economic evaluation, e.g. modeling and observational studies, can often not be used reliably for adequate policy making.

This thesis has discussed the informative value of the various approaches to economic evaluation for policy makers, and has proposed a methodological framework to enhance this value. By means of a number of studies, different aspects of the various approaches to economic evaluation were examined. This chapter examines the main findings of the studies and places them in the context of the thesis. However, the general methodological achievements of the thesis, presented in a methodological framework, are put forward in detail in Chapter 8. In the present chapter, these findings are assessed in view of the evolution of the methodology of economic evaluation in the past decades, and directions for further research are formulated.

METHODOLOGICAL ACHIEVEMENTS

A discussion of the appropriateness of results from RCT-based economic evaluations for allocation decisions in health care revealed a number of shortcomings (Chapter 2). The study focussed especially on determining the real world relevance of results from economic evaluation as an additional step towards making results more useful to policy makers. The identified biases related to either the scope, the applied methodology or the representativeness of an analysis. Although the study focussed on the RCT, many of the observations are relevant to other approaches of economic evaluation as well - to modeling, for example. The chapter discussed options for limiting the potential confounding influences of these biases, and proposed a checklist which should be applied by policy makers when using results from economic analyses in their own decision context and by economic evaluators when constructing and describing economic analyses. The proposed checklist is probably only a partial answer to all the problems that decision makers face when interpreting results from economic evaluation studies. Although it will certainly help in identifying individual factors that may confound decision making, it does not allow an assessment of the mutual impact of several factors.

A detailed analysis of the impact of age in economic analysis was carried out to provide more insight into the potential role of biases (Chapter 3). It was noted that disregarding age as an important variable in cost-effectiveness analysis may seriously

bias allocation decisions regarding health care resources. One reason is that the age structure in the RCT may be much different from the age structure of the population in which the intervention will be applied. Another reason is that the age structure of the target population may alter over time, due to demographic and epidemiological developments. The study showed that the allocation of health care resources at the macro-level is seriously impaired when age is ignored as a variable in cost-effectiveness analysis, and indicated the need to explicitly consider the impact of age in economic analyses.

Economic evaluators should describe the impact of age, which should then be considered by policy makers in a control for age effects. It is important to note that the study focussed only on age as one potential bias in economic evaluation. However, the concept of the study can also be applied to other potential confounding factors that are related to patient characteristics. For example, gender or race characteristics of treated patients may very well have an effect on the cost-effectiveness of the medical intervention under scrutiny. Ideally, economic evaluation studies should distinguish various subgroups of patients according to their most important characteristics and report cost-effectiveness ratios accordingly. The additional benefits of the analyses must be weighed against the costs of the larger sample sizes required.

Chapter 4, concerning influenza-related hospitalisations, illustrated another potential bias in economic evaluation, stemming from ignoring interactions with other diseases. This potential bias was assessed in the context of the cost-effectiveness of influenza vaccination. To examine the full impact of influenza on hospitalisation in the Netherlands, not only the number of hospitalisations formally attributed to influenza were considered, but also those that are otherwise related to influenza. The study showed that influenza was diagnosed as the main cause for hospitalisation in only a fraction of these hospitalisations (12%). Ignoring this issue underestimates the economic attractiveness of influenza vaccination.

This study leads to the conclusion that the confined perspective of a traditional economic analysis, focussing on a single intervention within a single disease, may very well bias the results. A somewhat wider perspective, which includes the consideration of interactions with other interventions and other diseases, may be necessary to estimate the full impact of an intervention. This requires a framework that goes beyond that of the traditional clinical trial. The use of regression analysis, as applied in the influenza study, may only partly solve the problem, as this considers only a single interaction. For a comprehensive approach, the use of modeling, either disease modeling or public health modeling, may be more adequate (see Chapter 8). Of course, such an approach is necessary only when the expected impact of the interactions is relatively large.

In a study on the economic impact of the introduction of transurethral microwave therapy (TUMT) in the treatment of benign prostate hyperplasia (BPH), an approach is developed to enhance the informative value of RCT-based economic analyses (Chapter 5). The study combined results from an RCT-based economic analysis on TUMT with observational data to get a more realistic view of the economics aspects following the introduction of TUMT. Scenarios were constructed to explore the possible paths of the diffusion of TUMT in the treatment of BPH. The analysis lead to important conclusions on the overall impact of TUMT. It revealed that the organisation of providing TUMT hardly affects costs at the system level, and also made clear that the most important determinant of the total costs of BPH-treatment is an increased demand for BPH-treatment following the introduction of TUMT.

The approach in the study under scrutiny can also be applied to RCT-based economic analyses of other medical interventions. The study shows that modeling techniques, when applied in addition to RCT- based economic evaluation, can be a useful tool to enhance the real world relevance of results from RCT-based economic analyses. The availability of unit cost prices that are representative for the decision making context is a necessary condition. Furthermore, additional modeling has proven to be effective in transferring results from the individual level to the aggregate level. Depending on the analysis' required level of detail, this may either imply simple extrapolation of data, or more complex public health modeling (see Chapter 8). In all cases, this transposition requires insight into the patterns of use of medical interventions before the introduction of a new treatment, and the substitution of these interventions following the new treatment's introduction. In the assessment of the substitution effect, the definition of indication criteria for the new treatment - which is difficult to assess *a priori* its introduction - plays a pivotal role. The performance of modeling in addition to an RCT-based economic analysis may seem a labourious activity, but the gains in terms of better informed decision making - as also shown in the present BPH-study - may very well be worth the effort.

The use of models as a substitute for RCT-based economic analysis is illustrated in the study on the cost-effectiveness of pneumococcal vaccination in the Netherlands (Chapter 6). A model design was chosen because of the limited feasibility of carrying out an RCT-based economic analysis on pneumococcal vaccination, considering the large sample size required for effect detection. Using a decision-tree framework, the study suggested that vaccination of all individuals above the age of 65 years, in terms of cost-effectiveness, is comparable to many existing health care interventions.

However, the model does not provide strong evidence of cost-effectiveness due to the uncertainty of a number of model assumptions. Most important, assumptions on the effectiveness of pneumococcal vaccination are based on observational studies alone, while RCTs on pneumococcal vaccination have so far not shown proof of

effectiveness, possibly due to the small sample sizes involved. Because of this lack of strong evidence on effectiveness, one should be very careful when interpreting the results of this study. Further experimental research on the effectiveness of pneumococcal vaccination may be necessary before policy on pneumococcal vaccination can be defined.¹⁻³ Such experimental research is very costly because of the large sample sizes required, and it is not likely to be carried out in the future.

Therefore, we have chosen another approach to prove the cost-effectiveness of economic evaluation of pneumococcal vaccination in current research. The approach recognises that although RCTs of pneumococcal vaccination have not (yet) proven its efficacy against all pneumococcal pneumonias, they do show its efficacy against a small subgroup (~10%) of all pneumococcal pneumonias: the invasive infections. A model on the cost-effectiveness of pneumococcal vaccination against invasive infections alone will receive more credibility than the existing model given this proof of efficacy. We reason that if the economic attractiveness of pneumococcal vaccination has been proven against the invasive infections alone, pneumococcal vaccination will certainly be economically attractive when all infections are considered.

Some findings from this study on the feasibility of modeling in economic evaluation can be generalised to research in other intervention and disease categories as well. An important observation is that models whose core assumptions are not based on scientifically rigorous (RCT) evidence, are not capable of demonstrating cost-effectiveness of the medical interventions under scrutiny. As several national guidelines on economic evaluation suggest, these studies are not sufficiently grounded to be used as the basis for decision making.^{4,5} Exceptions should be made for those interventions for which RCT-based evidence is not available due to ethical or practical reasons.

Chapter 7 clarified the complex process of setting priority regarding compatible and incompatible alternatives of medical interventions. It is noted that setting priority in the context of compatible alternatives may refer to the selection of more than one, possibly all alternatives. In a set of incompatible alternatives, only one alternative can be selected. The value that society attaches to a unit of effectiveness (e.g. a QALY) has an important impact on the priority ranking of medical interventions. A framework is presented that allows policy makers a better understanding of the impact of particular levels of these values on their optimal policy choices. This study has shown the importance of adequately reporting results to policy makers. The main finding of the study is that researchers should present not only cost-effectiveness ratios, but should also report costs and effectiveness separately in order to allow an assessment of the impact of the value that society attaches to a unit of effectiveness.

An overview of the methodological achievements of this thesis is presented in Chapter 8. This chapter elaborates on the research question as stated in the Introduction: 'How can economic evaluation studies be prepared to optimally support policy making?'. The chapter is organised around the sub-questions as identified in the introduction chapter, and treats them consecutively. It is observed that adequate policy making requires internally and externally valid cost-effectiveness information. Also, policy makers such as health care insurers may want assessments of the overall impact on budget and health following the adoption of an intervention. A methodological framework is proposed to meet these requirements. It is observed that, as a starting point, RCT-based cost analysis or RCT-based full scale modeling techniques like decision-tree analysis are the acceptable methods in the design of economic evaluation studies. To enhance the real world relevance of these instruments, approaches like the use of sensitivity analysis, checklists, observational data or additional modeling can be applied. The latter option in particular seems methodologically attractive. To address issues at the system level, the results at the individual level are transferred by simple modeling techniques like extrapolation or more complex techniques like disease modeling or public health modeling. The paper concludes that there is potentially a large role for modeling in economic evaluation, but that in order to enhance its credibility, more attention should be paid to validity aspects.

What is the impact of the methodological achievements of this thesis for the various actors in the health care field? The suggested approach on reporting age-specific, but also possibly gender- and race specific cost-effectiveness ratios may provide more adequate information to decision makers, but may also raise concerns regarding equity. For example, as noted in Chapter 3, the economic attractiveness of medical interventions tends to decrease with increasing age. The direct application of this information into the setting of priority for medical interventions would favour the young compared to the elderly in terms of the allocation of health care resources. In this context, the decision maker clearly has an important role in judging this trade-off between efficiency and equity. It is the role of the economic evaluator, however, to deliver the necessary information to allow these decisions.

The approach suggested for arriving at cost-effectiveness estimates that are specific to the decision making context implies a recognition of cost variation in health care delivery facilities. For example, it may be noted that costs may depend on physician experience: at the lower end of the learning curve relatively young physicians may incur higher costs than their older counterparts. These can be regarded as efficiency differences. It depends on the type of decision context whether these inefficiencies in health care provision should be considered when extrapolating results from RCTs: in an allocation decision context, these inefficiencies should be included to avoid underestimating the total costs stemming from the use of a medical intervention. However, in a reimbursement context, to furnish incentives for an efficient

provisions of medical interventions, these inefficiencies should not be included when determining the fee levels.

The focus in this thesis on the role of modeling in health care may add to the credibility of modeling studies financed by the pharmaceutical industry. It has been stressed that more attention should be paid to the question of how the validity of models can be enhanced. The pivotal role of RCTs in particular in the application of modeling in economic evaluation has been amplified. Recognition of the standards for economic evaluation, with particular attention to modeling, by the pharmaceutical industry is a necessary condition for influencing reimbursement decisions.

The analyses in this thesis have focussed primarily on the question how to make results from economic evaluation more useful to the policy making process by enhancing its methodological tools. However, it should be recognised that enhancing the real world relevance of results is only one step forward in the process of making the concept of economic evaluation a more valid and credible approach to the allocation of health care resources. There are many remaining factors that may limit an unrestricted application of economic evaluation results into decision making. Whereas this thesis has focussed merely on the cost-side of economic evaluation, other work has considered the effect-side of economic evaluation. For example, Bleichrodt ⁶ has elaborated on the characteristics and limitations of utilities in the use of QALY-type measures in medical decision making, in particular on the aggregation of individual-based utilities to the population level. Furthermore, there are many factors in the political decision making process that prohibit the direct use of results of economic evaluation. Many of these factors are reflected in the efficiency-equity discussion⁷, in which the issue of the use of age as a discriminating variable in cost-effectiveness analyses and consequently in the allocation of scarce health care resources is most prominent.⁸

AREAS FOR FUTURE RESEARCH

This thesis has identified a number of areas for future research. First, further research should focus on the role of the RCT as the primary vehicle for economic analysis. The specific needs of economic evaluation in the design of RCTs need to be identified, especially in the context of the external validity of results. For example, research should examine the feasibility of subgroup analysis within RCTs to increase the appropriateness of extrapolating results to populations outside the trial. Do the benefits of such an approach offset the burden of including larger sample sizes? Closely related is research focussing on the potential of using natural pragmatic trials. So far, only a small number of such trials have been performed. Additional studies will provide more information on the possibilities and limitations of this promising approach.

Second, this thesis has identified a large role for modeling in future research in economic evaluation. Currently, modeling in health care is receiving relatively little credibility because of its suggested lack of internal validity. This 'black box' image heavily limits its role in policy making. Future research should address the scientific rigour of modeling, and should develop approaches to enhance the internal validity. As a starting point, this may be done by defining requirements that appropriate models should meet in order to be accepted in health care policy making, e.g. on transparency of design or peer review by experts.

Third, further research should be directed at the role of policy makers. The different actors using cost-analysis information should be identified, and their specific needs should be examined. The issue of system level analyses in particular, as observed in this thesis, has not yet been thoroughly discussed in research on the methodology of economic evaluation. However, the growing importance of fund holders in health care will give rise to an increased demand for impact assessments in the near future. More attention needs to be focussed on how the specific needs of various policy makers can be addressed. Currently, economic analysis are typically addressing the needs of a single specific policy maker. Future research should focus on designs that allow the results of economic evaluation to be useful for various kinds of policy makers.

DEVELOPMENT OF ECONOMIC EVALUATION

How do the conclusions of this thesis relate to the overall development of the methodology of economic evaluation in health care in the past decades? Figure 1 shows the evolution as described by Rittenhouse.⁹

Acquisition costs were the first consideration of policy makers regarding health care costs. Later on, influenced by the tradition of experimental research in the field of medicine, the RCT was adopted as the gold standard for economic evaluation. Soon, biases to RCT-based economic analyses were identified. More recently, the US PORT studies have focussed attention on observational studies. Currently, the benefits of combining evidence from multiple study designs to reduce bias are being recognised. The achievements of this thesis contribute to this last step by clarifying and detailing a number of issues. As a basis for further exploration, the thesis provides a methodological framework for identifying the information required for optimal policy making and also for identifying existing biases in current approaches to economic evaluation.

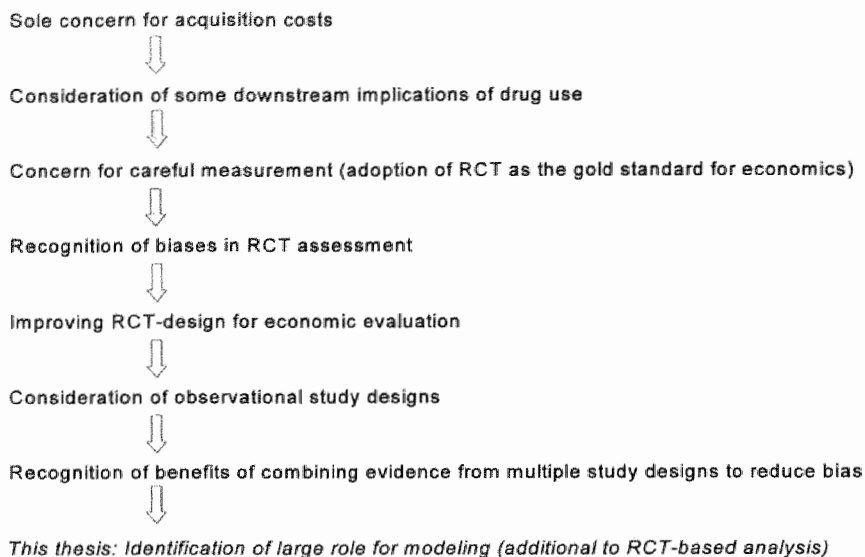


FIGURE 1. Evolution of the methodology of economic evaluation

Furthermore, the thesis has elicited the sources of evidence to be used in a comprehensive approach to economic evaluation that will address the needs of adequate policy making. The potentially large role of additional modeling that has been identified in this thesis is perhaps the next important step forward in the evolution of the methodology in economic evaluation.

REFERENCES

1. Hak E., Grobbee D.E., Verheij Th.J.M. Kosten-effectiviteit van pneumokokken-vaccinatie van ouderen. Letter to the editor. *Nederlands Tijdschrift voor Geneeskunde* 1998; 142: 201-202
2. Van den Berg J.W.K. Kosten-effectiviteit van pneumokokkenvaccinatie van ouderen. Letter to the editor. *Nederlands Tijdschrift voor Geneeskunde* 1998; 142: 202
3. Baltussen R.M.P.M., Ament A.J.H.A., Leidl R.M., van Furth R. Kosten-effectiviteit van pneumokokkenvaccinatie van ouderen. Letter to the editor. *Nederlands Tijdschrift voor Geneeskunde* 1998; 142: 202-203
4. Commonwealth of Australia, Department of Health, Housing and Community Services. *Guidelines for the pharmaceutical industry on preparation of submissions to the pharmaceutical benefits advisory committee*. Canberra: Australian Government Publishing Services, 1995

5. F.D.A. Division of Drug Marketing, Advertising and Promotion. *Principles for the review of pharmaceutical promotion*. Washington DC: F.D.A., 1995
6. Bleichrodt, H. *Applications of utility theory in the economic evaluation of health care*. Thesis. Ridderkerk: Ridderprint, 1996
7. Ubel P.A., DeKay M.L., Baron J., Asch D.A. Cost-effectiveness analysis in a setting of budget constraints - is it equitable? *N Engl J Med*. 1996; 334: 1174-7
8. Williams A. Evans J.G. The rationing debate. Rationing health care by age. *BMJ*. 1997; 15: 820-5
9. Rittenhouse B. *Uses of models in economic evaluations of medicines and other health technologies*. London: BSC Print Ltd, 1996

SUMMARY

This thesis has studied the question how economic evaluation can optimally support policy making in health care. The starting point of the thesis was the idea that traditional approaches to cost-effectiveness analysis are often inadequate for use by policy makers. The RCT, often considered to be the gold standard for economic evaluation, falls short in addressing real world situations, and fails to address issues at the more aggregate level. Other approaches to economic evaluation, e.g. modeling and observational studies, can often not be used reliably for adequate policy making. By means of a number of studies, different aspects of the various approaches to economic evaluation were examined. The requirements for adequate economic evaluation studies have been defined, and the various approaches have been tested against these requirements. This led to the recognition of a number of potential biases. In a series of individual studies, some methodological approaches to cope with these biases have been developed. This resulted in the definition of a general methodological framework on the real world design of economic evaluation.

Chapter 2 consists of a literature review on the potential biases in economic evaluation. The appropriateness of results from economic evaluation for allocation-decisions in health care is a point of major concern for decision makers. The chapter concentrated on the real world relevance of results from economic evaluation as an additional step towards making results more useful to decision makers. Three classes of biases were examined. The first relates to the limited scope that economic analysts sometimes choose in RCTs. The second class involves the methodological aspects of RCTs and questions the real world relevance of the tools with which economic analysts estimate costs on the basis of RCTs. The third class concerns the representativeness of RCT results, i.e. the generalizability of these results and their usefulness in other treatment contexts. The chapter discussed options for limiting the potential confounding influences of these biases and proposed a checklist which should be applied by policy makers when using results from economic analyses in their own decision context and by economic evaluators when constructing and describing economic analyses. This will enhance the relevance of the results of economic evaluation in decision making and improve the information basis for actual allocation decisions in health care.

Chapter 3 examined the impact of one specific bias in policy making based on economic evaluation: the impact of age. First, at the clinical level, the relation between age and the cost-effectiveness of medical interventions was analysed. Second, at the population level, a framework was presented which allows researchers and decision makers to assess the impact of these effects on the decision making process. It is shown that the allocation of health care resources at the macro-level is seriously impaired when age is ignored as a variable in cost-effectiveness

analysis. Because clinical trials typically employ 'young' populations, when the data is extrapolated to the whole population the attractiveness of medical interventions in terms of cost-effectiveness may be considerably overestimated. Furthermore, the cost-effectiveness ratio may vary across countries or over time as a result of demographic or epidemiologic variation. Economic evaluators should describe the impact of age, which should then be considered by decision makers to control for age effects.

Chapter 4 analysed the consequences of disregarding a potential confounder in economic evaluation, stemming from ignoring interactions with other diseases. This potential bias has been assessed in a study on influenza-related hospital admissions in the context of the cost-effectiveness of influenza vaccination. Influenza infection is associated with significant morbidity and mortality. Previous research has shown that large number of deaths in the Netherlands can be attributed to influenza. The purpose of this study was to examine the impact of influenza on hospitalisation in the Netherlands. Two methods were applied to estimate this effect: a) regression analysis and b) comparison of hospitalisation in epidemic years with non-epidemic years. Hospital discharge rates in the period 1984-1993 have been considered. The study shows that, during the period studied, on average, almost 2,700 people were hospitalized for influenza per annum and that influenza was diagnosed as the main cause for hospitalisation in only a fraction of these hospitalisations (326: 12%). From an economic perspective, these results imply that the cost-effectiveness of vaccination against influenza may be severely underestimated when looking only at the changes achieved in the number of hospitalisations attributed to influenza.

Chapter 5 reports on a study on the economic impact of the introduction of transurethral microwave therapy (TUMT) in the treatment of benign prostate hyperplasia (BPH). In the study, an modeling approach is developed to enhance the informative value of RCT-based economic analyses to policy makers. The approach combined results from an RCT-based economic analysis on TUMT with observational data to get a more realistic real world view of the economics aspects following the introduction of TUMT. Furthermore, analyses at the system level were performed to assess its budget impact. Several scenario's were constructed, including a baseline scenario, demand scenarios reflecting the number of men that will be treated by TUMT following its introduction, and supply scenarios reflecting the number of hospitals that will provide TUMT following its introduction. The analysis lead to important conclusions on the overall impact of TUMT. It revealed that the organisation of providing TUMT hardly affects costs at the system level, and also made clear that the most important determinant of the total costs of BPH-treatment is an increased demand for BPH-treatment following the introduction of TUMT. The study showed that modeling techniques, when applied in addition to RCT-based economic evaluation, can be a useful tool to enhance the real world relevance of results from RCT-based economic analyses. Furthermore, additional

modeling has proven to be effective in transferring results from the individual level to the aggregate level.

Chapter 6 reports on the use of models as a substitute to RCT-based economic analysis of pneumococcal vaccination. The latter is not considered feasible due to the large sample sizes required. For various age categories, the economic attractiveness of the vaccination of all individuals as well as the vaccination of only those individuals with a specific disease has been calculated. It was concluded that, allowing for some uncertainty regarding key variables such as the vaccine efficacy and the hospital admission rate, the vaccination of all individuals above the age of 65 years seems comparable in terms of cost-effectiveness to many existing health care interventions. However, because of the uncertainties, the model does not provide strong evidence of cost-effectiveness. The chapter concludes with the observation that, in general, models whose core assumptions are not based on scientifically rigorous (RCT) evidence, are not capable of demonstrating cost-effectiveness of the medical interventions under scrutiny.

Chapter 7 examined the reporting of cost-effectiveness results to policy makers. Theoretically it can be proven that an optimal allocation of resources within a constrained budget can be reached by considering cost-effectiveness ratio's (CERs). In this chapter the complex priority setting process regarding compatible and incompatible alternatives of medical interventions has been clarified. Priority setting in the context of compatible alternatives may refer to the selection of more than one, possibly all alternatives. Inherent to a set of incompatible alternatives is that only one alternative can be selected. This latter situation frequently occurs in health care. The value that society attaches to a unit of effectiveness (eg a QALY) has an important impact on the priority ranking of medical interventions. By explicating this value and by using the 'net-value' approach, a graphical framework has been presented that allows decision makers a better understanding on the impact of particular levels of these values on their optimal policy choices. For illustrative purposes, it has been shown that by the erroneous application of decision rules, some recent papers have provided sub-optimal recommendations for health care policy. This study has shown the importance of adequately reporting results to policy makers. The main finding of the study is that researchers should present not only cost-effectiveness ratios, but should also report costs and effectiveness separately in order to allow an assessment of the impact of the value that society attaches to a unit of effectiveness.

Chapter 8 proposed a general framework to real world designs in economic evaluation. The chapter has identified the information that economic evaluation should provide to policy makers, so that they may make appropriate and effective decisions. First, policy makers need cost-effectiveness information that is both internally and externally valid. The latter aspect is often ignored and refers to how relevant the results are to the specific decision making context of the policy maker.

Second, policy makers, like purchasers of care, may want assessments of the overall impact on budget and health following the adoption of an intervention. This requires a more aggregate approach than the traditional analyses that are typically individual-oriented. Against this background, this chapter examined the three main conceptual designs for economic evaluation: the use of RCTs, observational studies, and modeling. The chapter identified a number of options for enhancing the informative value of economic evaluation studies. Of these options, the use of additional modeling and observational data seems to be the most promising. To address issues at the system level, disease modeling or public health modeling is suggested. The chapter concludes that there is potentially a large role for modeling in economic evaluation, but that in order to enhance its credibility, more attention should be paid to validity aspects.

In Chapter 9, some conclusions are drawn. Most important is the observation of shortcomings in the traditional approaches to economic evaluation, and the suggestion to pay more attention to (additional) modeling in health care as potential useful approach to control for biases. However, before modeling is ready for its enhanced status, its credibility among policy makers should be carefully addressed. As a starting point, this may be done by defining requirements that appropriate models should meet in order to be accepted in health care policy making, e.g. on transparency of design or peer review by experts. The thesis concludes with the observation that the potentially large role of additional modeling that has been identified in this thesis is perhaps the next important step forward in the evolution of the methodology in economic evaluation.

SAMENVATTING

Binnen de gezondheidszorg dienen zich voortdurend veelbelovende nieuwe medische interventies aan. In het verleden werden veel van deze interventies alleen op basis van evaluaties van aspecten als werkzaamheid en effectiviteit toegepast. Wanneer deze interventies ook geëvalueerd worden op basis van hun economische aspecten, is het mogelijk te achterhalen welke interventie maatschappelijk gezien de meest aantrekkelijke is. Economische evaluatie van interventies in de gezondheidszorg is een vrij recent fenomeen, en het belang ervan groeit. Door middel van economische evaluatie worden kosten en effecten van medische interventies met elkaar vergeleken, om zodoende te bepalen welke interventies de beste resultaten geven in verhouding tot de kosten.

In dit proefschrift staat de vraag centraal hoe economische evaluatie het beleid in de gezondheidszorg optimaal kan ondersteunen, en zo kan bijdragen aan een verbeterde efficiëntie in de gezondheidszorg (hoofdstuk 1). Uitgangspunt is de onderkenning van tekortkomingen in de huidige vormen van economische evaluatie, waardoor deze neit helemaal geschikt zijn voor beleidsmakers. De gerandomiseerde klinische trial (RCT), die vaak gezien wordt als de gouden standaard van kosten-effectiviteitsanalyse, levert veelal resultaten op die niet representatief zijn voor de dagelijkse praktijk, en die moeilijk generaliseerbaar zijn. In dit proefschrift wordt deze problematiek aangeduid met een lage *real world* relevantie van de resultaten. Daarnaast levert een op RCT-gebaseerde economische evaluatie vaak geen informatie over de budget-impact als gevolg van de invoering van een medische interventie. Andere benaderingen van economische evaluatie, zoals het gebruik van modelering of observationele studies, worden vaak bekritiseerd vanwege hun lage validiteit. In dit proefschrift werden aan de hand van een aantal studies de verschillende benaderingen van economische evaluatie nader onderzocht. De vereisten voor optimale economische evaluatie werden gedefinieerd, en de verschillende benaderingen werden aan de hand hiervan geëvalueerd. Dit heeft geleid tot de identificatie van een aantal mogelijke *biases* (vertekeningen) binnen de uitvoering en het gebruik van economische evaluatie. In een aantal studies zijn mogelijke methodologische benaderingen voor de interpretatie en/of het voorkomen van deze vertekeningen ontwikkeld. Dit heeft geresulteerd in een algemeen methodisch raamwerk voor economische evaluatie.

Hoofdstuk 2 bestaat uit een literatuurstudie over de mogelijke vertekeningen in economische evaluatie. Dit hoofdstuk richt zich met name op de *real world* relevantie van resultaten van economische evaluatie. Er worden in dit hoofdstuk drie verschillende soorten vertekeningen behandeld. De eerste is gerelateerd aan het beperkte blikveld van economische evaluatie: economische evaluaties richten zich vaak op één bepaalde interventie binnen één bepaald ziektegebied, zodat interacties met andere interventies of ziektes worden genegeerd. De tweede heeft te maken met

de methodische aspecten van economische evaluatie. De derde soort vertekening heeft betrekking op de generaliseerbaarheid van resultaten. In dit hoofdstuk wordt besproken hoe de invloed van deze mogelijk vertekeningen te beperken is, en er wordt een *checklist* gegeven die door beleidsmakers gebruikt kan worden bij de interpretatie van resultaten van economische evaluatie en door onderzoekers voor het ontwerpen van een economische evaluatie.

In hoofdstuk 3 wordt de invloed van leeftijd als een mogelijke vertekende factor in economische evaluatie besproken. Op klinisch niveau is de relatie tussen leeftijd en kosten-effectiviteit van medische interventies onderzocht. Daarnaast is op populatieniveau een raamwerk ontwikkeld waarbinnen beleidsmakers de invloed van de factor leeftijd kunnen interpreteren. Omdat klinische trials vaak relatief jonge individuen includeren zal de aantrekkelijkheid van de interventies in termen van kosten-effectiviteit overschat worden wanneer de data geëxtrapoleerd worden naar de gehele populatie. Daarnaast kan de kosten-effectiviteit variëren tussen landen en over de tijd heen, als een gevolg van epidemiologische verschillen. Onderzoekers zouden daarom de invloed van leeftijd moeten rapporteren, zodat deze geïnterpreteerd kan worden door beleidsmakers.

In hoofdstuk 4 worden de gevolgen van een andere mogelijke vertekening in economische evaluatie geanalyseerd, die voortkomt uit het negeren van interacties met andere ziektes. Deze vertekening is onderzocht in een studie naar influenza gerelateerde ziekenhuisopnames in de context van de kosten-effectiviteit van influenzavaccinatie. Influenza-infecties gaan gepaard met veel mortaliteit en morbiditeit. Eerder onderzoek heeft aangetoond dat een groot aantal doden in Nederland - welke niet gediagnosticeerd zijn als veroorzaakt door influenza - wel degelijk gerelateerd zijn aan influenza. Het doel van de huidige studie was het aantonen van de invloed van influenza op het aantal ziekenhuisopnames in Nederland. Twee methoden werden hiervoor gebruikt: a) regressie analyse, en b) vergelijkingen van het aantal ziekenhuisopnames in epidemische jaren met niet-epidemische jaren. Voor de studie werden gegevens over ziekenhuisontslagen in de periode 1984-1993 gebruikt. De studie toont aan dat jaarlijks gemiddeld bijna 2.700 personen zijn opgenomen in het ziekenhuis als gevolg van influenza, maar dat deze ziekte in slechts 12% van de gevallen als hoofddiagnose genoemd wordt. In economisch perspectief betekent dit dat de kosten-effectiviteit van influenzavaccinatie onderschat wordt indien alleen gekeken wordt naar de ziekenhuisopnames die ook daadwerkelijk gediagnosticeerd zijn als influenza.

In hoofdstuk 5 wordt een studie besproken over de economische gevolgen van de invoering van transurethrale microgolf therapie (TUMT) in de behandeling van benigne prostaat hyperplasie (BPH). Deze studie is uitgevoerd aan de hand van een RCT-gebaseerde economische evaluatie trial waarin TUMT vergeleken wordt met transurethrale resectie van de prostaat (TURP). Door middel van modellering

werden in de studie de resultaten van de trial gecombineerd met gegevens uit observationele studies om een *real world* beeld te krijgen van de gevolgen van de invoering van TUMT. Daarnaast werd een analyse op het meer geaggregeerde niveau uitgevoerd om de budget-impact te schatten. Diverse scenario's werden geconstrueerd, waaronder een baseline scenario, verschillende vraag-scenario's betreffende het aantal mannen dat behandeld zal worden met TUMT na de invoering, en verschillende aanbod-scenario's betreffende het aantal centra's dat TUMT zal aanbieden na de invoering. De analyses leidden tot opvallende conclusies. Het bleek dat de belangrijkste determinant van de totale kosten van TUMT het aantal mannen is dat zich aanbiedt voor behandeling. Daarentegen heeft de organisatie van het aanbod van TUMT nauwelijks invloed op de totale kosten. Vanuit methodisch gezichtspunt is gebleken dat modelleren technieken, additioneel aan een RCT-gebaseerde economische evaluatie, nuttig kunnen zijn om de *real world* relevantie van de resultaten te bevorderen. Daarnaast is additioneel modelleren nuttig gebleken om resultaten van het individuele naar het geaggregeerde niveau te extrapoleren.

Hoofdstuk 6 handelt over het gebruik van modellen als substituuut voor een RCT-gebaseerde economische evaluatie van pneumokokkenvaccinatie. Gezien de grote benodigde onderzoekspopulatie is het uitvoeren van deze laatste vorm van evaluatie erg lastig. Met behulp van een beslisboom zijn voor verschillende vaccinatie-strategieën de verwachte veranderingen in medische kosten en gezondheidseffecten geschat. Op basis hiervan zijn kosten-effectiviteitsratio's berekend. Met inachtnahme van onzekerheid ten opzichte van verschillende veronderstellingen in het model, werd geconcludeerd dat op basis van de economische aspecten, pneumokokken vaccinatie van ouderen boven 65 jaar, en van patiënten met chronische long- of hartziekten boven 55 jaar, aantrekkelijk is. Gezien de onzekerheden geeft het model geen hard bewijs van kosten-effectiviteit. In het hoofdstuk wordt geconcludeerd dat modellen die belangrijke veronderstellingen gebruiken die niet gebaseerd zijn op RCT-bewijs, in het algemeen niet geschikt zijn om kosten-effectiviteit aan te tonen.

In hoofdstuk 7 wordt de rapportage van kosten-effectiviteitsratio's aan beleidsmakers besproken. Theoretisch kan bewezen worden dat een optimale allocatie van middelen binnen een beperkt budget bereikt kan worden door gebruik te maken van kosten-effectiviteitsratio's. In dit hoofdstuk worden de complexe beslisregels aangaande compatibele en incompatibele alternatieven van medische interventies verhelderd. Het stellen van prioriteiten in de context van compatible alternatieven omvat de selectie van één of meerdere alternatieven en zo mogelijk alle alternatieven. Inherent aan een set van incompatibele alternatieven is dat slechts één alternatief gekozen kan worden. Dit laatste is vaak aan de orde in de gezondheidszorg, bijvoorbeeld bij screeningprogramma's. De waarde die de gemeenschap toekent aan een eenheid van effectiviteit (b.v. een QALY) heeft een belangrijke invloed op de prioriteitsstelling van medische interventies. Door het expliciteren van deze waarde, en het gebruik van de 'netto-waarde' methode, is een grafisch raamwerk ontwikkeld dat voor

beleidsmakers een hulpmiddel kan zijn bij het alloceren van middelen. Ter illustratie zijn enkele recente studies besproken waarin, door het foutief toepassen van de beslisregels, sub-optimale conclusies zijn getrokken. De studie laat het belang zien van een correcte rapportage van resultaten van economische evaluatie. De belangrijkste uitkomst is dat onderzoekers niet alleen kosten-effectiviteitsratio's moeten presenteren, maar ook afzonderlijk kosten en effecten moeten rapporteren om zo de invloed te kunnen interpreteren van de waarde die de gemeenschap aan een eenheid effectiviteit toekent.

In hoofdstuk 8 wordt een algemeen raamwerk voorgesteld voor de economische evaluatie van gezondheidszorginterventies. De informatie die economische evaluaties in het ideale geval zouden moeten leveren aan besluitvormers wordt omschreven. Ten eerste hebben besluitvormers behoefte aan informatie die zowel intern als extern valide is. Dit laatste aspect wordt vaak genegeerd, en heeft betrekking op de toepasbaarheid van de resultaten op de specifieke besluitvormingscontext van de beleidsmaker. Ten tweede willen beleidsmakers, zoals inkopers van zorg, informatie over de impact van de invoering van een interventie op kosten en gezondheid op het geaggregeerde niveau. Ten aanzien van deze eisen worden in dit hoofdstuk de drie belangrijkste benaderingen in economische evaluatie geëvalueerd: het gebruik van RCTs, modellen, en observationele studies. Het hoofdstuk bespreekt opties om de informatieve waarde van economische evaluatie studies te verhogen. Van deze opties zijn het gebruik van additioneel modelleren en van observationele gegevens de meest veelbelovende. Voor het extrapoleren van gegevens naar het geaggregeerde niveau is *disease modeling* of *public health modeling* mogelijk. De conclusie van dit hoofdstuk luidt dat er een potentieel grote rol voor modelleren is weggelegd binnen economische evaluatie. Daarvoor is het noodzakelijk dat er aandacht besteed wordt aan validiteitsaspecten.

In hoofdstuk 9 worden de belangrijkste conclusies van dit proefschrift besproken. Dit zijn de erkenning van tekortkomingen in economische evaluatie, en de identificatie van de rol die (additioneel) modelleren zou kunnen spelen om de invloed van die tekortkomingen te beperken. Voordat modelleren op grote schaal toegepast kan worden binnen economische evaluatie moet met name de geloofwaardigheid hiervan onder beleidsmakers verhoogd worden. Om te beginnen kan dit gedaan worden door richtlijnen op te stellen waaraan geschikte modellen moeten voldoen, bijvoorbeeld op het gebied van transparantie van ontwerp of *peer review* door experts. De belangrijkste conclusie luidt dat de potentieel grote rol van modelleren, zoals geïdentificeerd in dit proefschrift, mogelijk een belangrijke stap verder is in de evolutie van de methodologie van economisch evaluatie.

DANKWOORD

'Je kan beter spijt hebben van dingen die je wel gedaan hebt, dan van dingen die je niet gedaan hebt'. Met die gedachte in mijn achterhoofd begon ik vier jaar geleden aan mijn promotie-onderzoek. Achteraf gezien waren het vier erg goede jaren, en dat komt vooral door de samenwerking met veel mensen die betrokken zijn geweest bij mijn onderzoek. Hoewel promoveren een nogal solistische bezigheid heet te zijn, heb ik dat zeker niet zo ervaren. Een aantal mensen die me met raad en daad hebben bijgestaan wil ik graag noemen.

Allereerst wil ik graag Prof. Reiner Leidl bedanken voor zijn inzet tijdens dit promotie-onderzoek. Reiner, je grote deskundigheid op dit terrein bleek telkens weer uit je duidelijke en bondige commentaar op mijn manuscripten. Niet alleen tijdens je verblijf in Maastricht, maar ook na je vertrek naar Ulm hebben we prettig samengewerkt. Reiner, bedankt!

Co-promotor en dagelijks begeleider André Ament ben ik minstens evenveel dank verschuldigd. André, je hebt voor een groot gedeelte mijn plezier in het onderzoek bepaalt door je relativiseringsvermogen en je humor, en door de vrijheid die je me gegeven hebt. Onze gezamenlijke buitenlandse bezoeken - met name naar Lyon en Barcelona - zijn zeker hoogtepunten uit mijn AIO-schap. Ik weet nu ook hoe je Spaanse *puta*'s van het lijf houdt. Ik zal met veel plezier aan onze samenwerking terugdenken als ik straks 'in die zandbak' werk.

Graag wil ik ook de leden van de beoordelingscommissie - Prof. Wim Groot (voorzitter), Dr. Hans Jager, Prof. Richard Janssen, Prof. Hans Maarse en Dr. Frits van Merode - bedanken voor hun bereidheid om mijn proefschrift te beoordelen.

Hans Jager, Arnold Reiners en Maarten Postma van het RIVM, jullie hebben je ingespannen om mij een onderzoek te laten verrichten bij de afdeling VTV. Maarten, jouw vaak cynische commentaar op wetenschap, en op mijn studie in het bijzonder, heb ik altijd op zijn waarde weten te schatten... Verder wil ik bij deze ook alle andere medewerkers van het RIVM bedanken. De Pavlov-reactie die ik bij jullie ontwikkeld heb zorgt er nog steeds voor dat ik zin in koffie krijg als ik een hoorn hoor.

GertJan van der Wilt, Hans Severens, en Gina Wielink van de vakgroep MIES in Nijmegen hebben het voor mij mogelijk gemaakt om het TUMT-TURP project uit te voeren. GertJan, het lijkt misschien lang, maar 3 jaar na ons eerste contact tijdens de ISTACH-conferentie in Zweden ligt er dan wel een publicatie van het project. Gina, we hebben altijd erg soepel samengewerkt; dit heeft volgens mij alles te maken gehad met onze gedeelde visie dat wetenschap niet moeilijker gemaakt moet worden dan het feitelijk is. Herman Stoevelaar van de Erasmus Universiteit te Rotterdam

Bedankt!

heeft me er in deze studie steeds op moeten wijzen dat wetenschap ook niet makkelijker voorgesteld moet worden dan het feitelijk is. Herman, jouw belangeloze inzet voor dit project heb ik altijd fantastisch gevonden, bedankt!

Prof. van Furth, bedankt voor uw stimulerende rol binnen het pneumokokken-project. Graag wil ik ook Sen McGlinn, Barbara Goldberg, en Liesbeth Preng bedanken voor het nodige correctiewerk van mijn manuscripten.

Zelden is een AIO een van de meest constante factoren in een vakgroep. Tijdens mijn aanstelling op de vakgroep EGZ, en later BEOZ, heb ik echter vele collega's zien komen en gaan. Graag wil ik Grant bedanken die mij reeds op mijn eerste werkdag informeerde over de nutteloosheid van economisch evaluatie-onderzoek. We zullen het niet hebben over de nutteloosheid van onze gezamenlijke verkenningen van Internet (Doom?). Anne-Peter, jou wil ik bedanken voor de twee erg leuke jaren waarin ik een kamer met je gedeeld heb. Daarnaast wil ik ook Sybilla en Mariëlle (ik begrijp niet waarom ze jou ooit het vakgroeps-mormel genoemd hebben) bedanken voor hun collegialiteit. Graag wil ik ook Maureen noemen; je hebt mij met het pneumokokken-project goed op weg geholpen.

En dan de huidige collega's. Waar te beginnen? Misschien met het kind van de vakgroep, Brigitte. Hoe ging dat ook alweer over een zieke geest in een gezond lichaam? Paranimf Marike, *biken & boarden* is wat anders dan SPSS en Word, en gelukkig maar. Onze gesprekken gingen en gaan gelukkig over niets anders. Bert, Eddy en Aggie, bedankt voor alle adviezen over hoe te promoveren. Jan, die drie frikadellen zijn zeker de investering waard geweest, dankzij jou raakte ik niet verstrikt in het World Wide Web;). Daarnaast zorgde je en zorg je - nu samen met Hilde, Isel en Astrid - nog steeds voor een erg soepele ondersteuning van de vakgroep. Bedankt! Mijn collega's uit AIO-land, de medebestuurders van AVUM, en met name paranimf Patrick, wil ik bedanken voor alle gedeelde vreugde omtrent het AIO-schap.

Graag wil ik ook familie en vrienden bedanken voor hun interesse. Hoewel velen de indruk hadden dat ik als Ambtenaar in Opleiding bezig ben geweest, kan ik toch zeggen dat ik vaak met het zweet op de rug gewerkt heb, al was het maar om jullie E-mails te beantwoorden. Restez Zen! Daarnaast wil ik graag mijn broer Hans bedanken zonder wiens belastinggeld dit allemaal niet mogelijk zou zijn geweest.

Anneke, lang heb ik gedacht dat ik mijn proefschrift zou afmaken *ondanks* jou; nu weet ik dat het mede dankzij jou is. Met je verhalen over *just another day in St Jans Hospital* in Weert heb je me veel geleerd over hoe de gezondheidszorg in de praktijk functioneert (of niet). Maar veel belangrijker was je vaak nuchtere relativerende commentaar. *Guappa*, bedankt!

CURRICULUM VITAE

Rob Baltussen werd geboren op 24 april 1968 te Rijkevoort. Na het behalen van het VWO-diploma in 1987 aan het Elzendaal College in Boxmeer ging hij Algemene Economie studeren aan de Universiteit Maastricht. Tijdens zijn studie heeft hij via een uitwisselingsprogramma *Development Economics* gestudeerd in Hull, Engeland. Zijn afstudeerscriptie betrof een economische evaluatie van een ontwikkelingsproject in Andra Pradesh, India. In 1994 is hij als Assistent-In-Opleiding begonnen bij de vakgroep Economie van de Gezondheidszorg aan de Universiteit Maastricht, die later is opgegaan in de vakgroep Beleid, Economie en Organisatie van de Zorg. Momenteel is hij part-time werkzaam als toegevoegd onderzoeker aan dezelfde vakgroep. Tevens heeft hij een part-time aanstelling aan de afdeling *Tropical Hygiene and Public Health* aan de Universiteit van Heidelberg. In die functie verricht hij onderzoek naar gezondheidszorgsystemen in West-Afrika.

**This thesis is financially
supported by:
Pasteur Mérieux MSD
Wyeth-Lederle
Solvay Pharmaceuticals**