Quantitative Health Impact Assessment

An exploration of methods and validity

Lennert Veerman

Painting on cover: "The Crystal Ball" (1902) by John William Waterhouse.

Quantitative health impact assessment: an exploration of methods and validity
Veerman, Jacob Lennert
Thesis Erasmus MC, University Medical Center Rotterdam,
with summary in English and Dutch.

Cover design: Anna Bosselaar | www.zoiets.com
Layout: Lennert Veerman
Printed by: PrintPartners Ipskamp, Enschede

ISBN 978-90-9022108-3

Quantitative Health Impact Assessment: an exploration of methods and validity

Kwantitatieve Gezondheidseffectschatting: exploratie van methoden en validiteit

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. Dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 12 september 2007 om 11:15 uur
door

Jacob Lennert Veerman

geboren te Amsterdam

Promotiecommissie

Promotor:	Prof. dr. J.P. Mackenbach
Overige leden:	Prof. dr. T.E.D. van der Grinten
	Prof. dr. ing. H.A.M. van Oers
	Dr. W.B.F. Brouwer
Copromotor:	Dr. J.J. Barendregt

Contents

1

General introduction

A brief introduction to Health Impact Assessment

Health Impact Assessment (HIA) has been described as "a combination of procedures, methods and tools by which a policy, program or project may be judged as to its potential effects on the health of a population, and the distribution of those effects within the population" (European Centre for Health Policy 1999). Various other definitions have been proposed, but as John Kemm concludes, the two essential characteristics of HIA are that it seeks to predict the future consequences for health of possible decisions; and that it seeks to inform decision-making (Kemm 2003). HIA typically focuses on interventions outside the health sector that do not have health improvement as their primary aim. This thesis is about making predictions for HIA.

HIA developed about 15 years ago and has its roots in two prior developments. For one part, it is the logical extension of environmental impact assessment (EIA) (Birley 1995). EIA informs decision makers of the consequences of development projects for the physical and chemical environment, and HIA adds the human health consequences of these environmental changes. The inclusion of health in EIA started in developing countries, where it became clear that large developmental projects had negative health consequences. In particular the construction of dams and irrigation schemes, that, among other things, provided breeding grounds for malaria mosquitoes, which prompted attention for health. Malaria had long been kept at bay by spraying DDT, but when this strategy started failing the interest in assessing health impacts and possible mitigation measures increased. Other important health issues are the displacement of people and the rising prevalence of schistosomiasis (Birley 2004). Interest in health and EIA was also growing in developed countries. Environmental impact assessment had been mandatory for specific types of projects since the early seventies, and in the early nineties the health consequences started to be added to the aims of these assessments in various countries (Kemm and Parry 2004). Examples of such environmental health impact assessments are the studies of the enlargement of a waste disposal facility in Lower-Saxony and of a city bypass highway project in Krefeld, Germany (Fehr 1999). In general, however, few EIA studies include impacts on human health (Alenius 2001). These early developments in EIA and HIA focused on projects. In recent years the EIA community realised that projects are often the consequence of earlier policy development at a more strategic level, and that it is important to influence policy at these earlier stages as well. This resulted in the development of Strategic Environmental Assessment. It is unclear

to what extent health concerns are currently included but it seems likely that HIA will find a place in this stream of impact assessment.

The second tradition on which HIA draws is the movement concerned with the (social) determinants of health, healthy public policy (Lalonde 1974), and the WHO Healthy Cities initiative. It focuses on social and behavioural determinants of health and emphasises the interaction with the people whose health it concerns. The model of health is a holistic one, which recognises that factors at the individual, communal and macro-political level influence the health of people and populations. In part, this development was stimulated by increased attention for health inequalities and the political will to reduce these (Acheson 1998). This required the ability to predict the health consequences of decisions. The landmark paper entitled 'Health impact assessment, an idea whose time has come' (Scott-Samuel 1996) heralded the start of much activity in this line of HIA, especially in the UK. It focused mostly, but not exclusively, on local projects. An example of this kind of HIA is the HIA of the former US Alconbury airbase in England where a large rail and road freight transport centre was planned. The HIA engaged the local community, authorities and the developer, and identified positive economic and negative environmental effects (Cambridgeshire Health Authority 2001). It has also been applied to regeneration initiatives, such as the rapid HIA that used a workshop to optimise the expected benefit from a 'Healthy Living Centre' (Barnes, Cooke et al. 2001).

A milestone in the development of HIA was the Gothenburg Consensus meeting in 1999. This meeting resulted in the formulation of the definition of HIA quoted at the start of this chapter. It also identified four values underpinning HIA: *democracy*, emphasizing the right of people to participate in a transparent process for the formulation, implementation and evaluation of policies that affect their life, both directly and through the elected political decision makers; *equity*, emphasizing that HIA is not only interested in the aggregate impact of the assessed policy on the health of a population but also on the distribution of the impact within the population, in terms of gender, age, ethnic background and socio-economic status; *sustainable development*, emphasizing that both short term and long term as well as more and less direct impacts are taken into consideration; and *ethical use of evidence*, emphasizing that the use of quantitative and qualitative evidence has to be rigorous, and based on different scientific disciplines and methodologies to get as comprehensive assessment as possible of the expected impacts (ECHP 1999). These values reflect the diversity in the background and practices of HIA.

In the Netherlands, HIA developed at the national level in the nineties. In 1996 a support desk was set up with the Netherlands School of Public Health

to assist the ministry of health. This office screened policy proposals for health impacts, and commissioned a number of health impact assessments. The topics varied and included energy tax regulation (eco-tax, 1996), tobacco policy (1997, 1998) and housing policy (1999, 2002). Reports were also published on national budgets (1998-2002), election programmes of political parties (1998) and coalition agreements (1998, 2003) (Roscam Abbing 2004). However, around 2002 the political climate changed and the interest in HIA declined markedly. The attention of the ministry of health shifted further to the (financial) management of the health care system. Health impact assessment was delegated to local authorities, along with part of the responsibility for health and social care.

The developments at the European level are more encouraging. Article 152 of the EU Amsterdam Treaty (1997) states that "A high level of human health protection shall be ensured in the definition and implementation of all Community policies and activities". This has been followed by investment in research, which resulted in a guide for European policy health impact assessment (EPHIA Working Group 2004). Further work on quantitative modelling for HIA is underway in the DYNAMO-HIA project, which is coordinated by the department of Public Health at Erasmus MC.

Given that HIA derives from two rather different precursors, that it is applied to very diverse types of policies and projects, and that there is little agreement on terminology, it is not surprising that there is not much agreement on methodology either. Some HIA practitioners see HIA mainly as a way to incorporate the concerns of the people affected into a policy decision (Lester and Temple 2004) and stress democracy as one of the central values in HIA. Others express concern over the validity of many such community consultations and emphasise the epidemiological basis of HIA (Parry and Stevens 2001), and are subsequently, in barely concealed terms, accused of a paternalistic and expertist attitude and of maintaining existing inequities (Scott-Samuel et al, rapid response to previous reference).

HIA is a kind of ‘evaluation before the fact’, and in the debates on what constitutes a good HIA study and what counts as evidence, much can be recognised from the literature on the evaluation of interventions in health care. Some of the insights gained in the evaluation are also useful for HIA. Øvretveit for example recognises that an evaluation can be done from various perspectives, each with its own criteria (Øvretveit 1998). He distinguishes four such perspectives. The experimental perspective focuses on outcomes to discover evidence of causes and effects. The economic perspective is much like the experimental, but includes inputs to look at cost-effectiveness. Third, the

developmental perspective, in contrast, is much more process orientated and primarily qualitative: it aims to help (health care) providers to improve in the short term. Many HIA-practitioners, especially those in the UK, seem to share this perspective. The last perspective is the managerial perspective, which focuses on inputs, process and outputs and in which the evaluator has an inspectorial, 'quasi-independent' role. Many of the studies in environmental health impact assessment share traits of this perspective. In Øvretveit's terms, the perspective we take on HIA in this thesis is most like the experimental and economic perspectives, though we disregard the cost side of the economic equation, and though true experimentation is not possible in HIA, which does not wait to see whether the predictions come true. With the experimental perspective we do share the 'positive' idea that there is a reality that can be measured and forecasted (albeit imperfectly), and we perform a sort of formalised thought experiments. Also, where possible, the evidence is derived from experiments. Øvretveit concludes that there is not one perspective that is better than the others; rather, the most suitable approach depends on the characteristics of the policy process the evaluation is intended to inform. Likewise, we do not claim that our perspective and methods are suitable for each and every HIA.

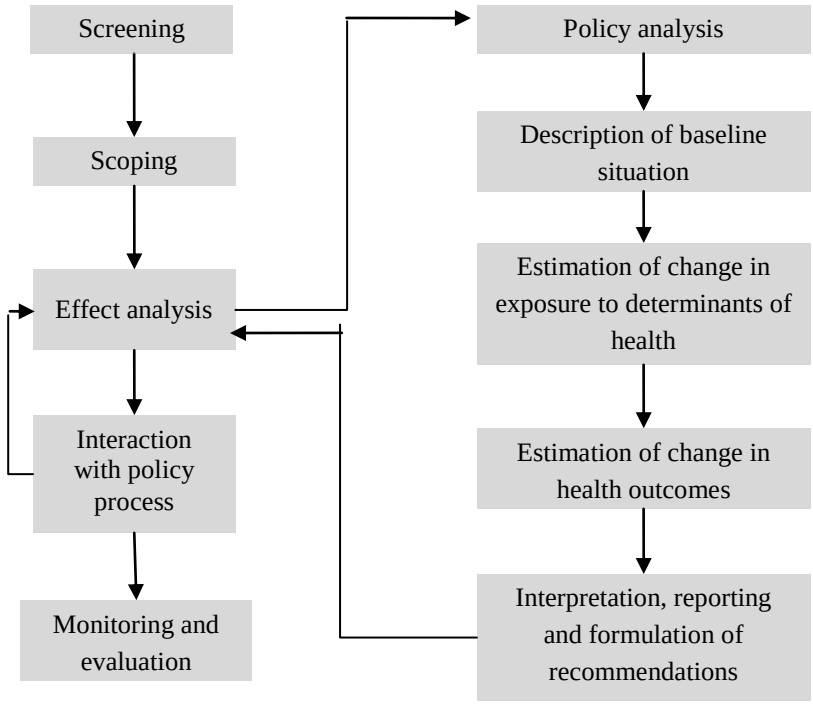
HIA is not unique in trying to inform policy processes by forecasting health consequences of possible decisions. Health economic evaluation has the same aim (Gold, Siegel et al. 1996; Drummond, O'Brien et al. 1997; Tan-Torres Edejer, Baltussen et al. 2003). Despite this similarity in aims, the current literature on HIA seldom refers to economic evaluation. This may be because health economic evaluation is a quantitative enterprise, while HIA has, at best, just started to explore how much can be quantified. Furthermore, economic evaluation mainly deals with health care interventions, such as new pharmaceuticals. It thus has a focus on interventions that specifically aim to improve health, in contrast to 'classical' HIA which focuses on interventions outside the health field that have health consequences as an unintended effect. In practice, this means that whereas the results of health economic analyses are welcomed by decision makers, HIA can often be expected to receive a less enthusiastic welcome because in many cases the decision makers do not have special knowledge of health, nor bear any formal responsibility for it. Still, they may have to be convinced to change their plans in the interest of health. This may be why, compared to health economic evaluation, HIA seems to place more emphasis on informing and influencing policy. The difference in the fields of application also results in a difference in the evidence base and the possibilities for quantification. While economic evaluation can often base its predictions on

randomised controlled trials, HIA seldom has such good quality evidence to go by. Probably more important, however, is an apparent distrust of the values in economic analysis, which is seen to give precedence to money over health, leading to rejection of its methods. The expression of health (and other) outcomes in a single metric facilitates trade-offs, but it is perceived to hide value judgements and the differential impact on different groups in society (Mindell, Hansell et al. 2001).

Note, however, that the definitions of HIA stated above do not restrict HIA to interventions outside the health sector, nor do they exclude the possibility of performing an HIA on an intervention that has health improvement as its primary objective. Conversely, health economic evaluation is not restricted to interventions within health care, or which have health improvement as the primary aim. Therefore, an HIA analysis *could* be identical to a health economic evaluation of that same intervention minus the costing-side. We will re-visit this issue in the general discussion (Chapter 8).

But HIA involves more than the forecasting of health effects. Most methods for HIA share a core of five steps: screening, scoping, effect analysis, interaction with the policy process and monitoring & evaluation. In the screening phase policy proposals are judged by their potential to influence health and, if influential, whether a HIA has any chance of influencing the decisions taken. If a HIA is deemed potentially useful, scoping determines how the study will be conducted and by whom, and which parties will be involved. The effect analysis estimates the nature and size of the health impacts, as well as their distribution over different groups in the population. Recommendations are drafted and, together with the results, communicated to the policymakers and stakeholders. Once a decision has been taken, it may be decided to monitor developments in health or in exposure to determinants of health, and evaluation of the HIA may take place. Figure 1.1 gives a schematic ‘generic’ overview of the procedures in HIA. This example is loosely based on the Merseyside Guidelines, an early approach to HIA that is widely used (Scott-Samuel 2001).

Figure 1.1 Schematic overview of the process of health impact assessment



Predicting health impacts

In this thesis we focus on predictions of health impacts in HIA, which constitutes the core of what is here called the effect analysis. Although predictions are a fundamentally important aspect of HIA, the literature on HIA largely neglects the issue of how to make valid predictions. Neither does the literature on epidemiology pay much attention to this issue. Most epidemiologists seem to shy away from moving beyond the analysis of 'hard' data, and are reluctant to make predictions using data that may not always be directly observed. This is in contrast to a field like econometrics, which focuses on predictive models. Econometrics is not without success in influencing policy decisions. For example, Dutch government economic plans are frequently adapted or even abandoned if the models of the CPB (Netherlands Bureau for Policy Analysis) predict adverse effects.

Furthermore, in HIA little attempt is made to quantify the expected health impacts. Quantification would be worthwhile for several reasons. First because knowing the size of an effect helps decision makers to distinguish between the details and the main issues that need to be addressed. It facilitates decision making by clarifying the trade offs that may be entailed (Kemm 2000; Mindell, Hansell et al. 2001). Secondly, because adding up all positive and negative health effects into a net effect permits the use of economic instruments such as cost effectiveness analysis, which further aids decision making.

To facilitate comparison of different health effects, these are best expressed not only in disease-specific terms but also in summary measures of population health (SMPH) (Murray 2002). SMPH are measures that combine information on mortality and non-fatal health outcomes, such as health-adjusted life years (DALYs or QALYs) or health-adjusted life expectancy (e.g. DALE). The use of SMPH makes the effects of interventions comparable across diseases and policies and permits the estimation of net effects if a proposal has both positive and negative health consequences.

However, in practice few HIA studies present quantitative health outcomes and summary measures of population health are virtually never used. One of the likely reasons for this paucity of quantification in HIA is that the required data and methods are lacking. This thesis aims to contribute to the development of methods that can assist in quantifying health outcomes in HIA. It advocates a two-step approach to the analytical process. The first step is to assess the effect of implementing a policy (i.e., an intervention) on exposure of a population to determinants of health. The second step translates this change in exposure to a change in health status. For this second step we explore the potential use of relatively simple life-table based simulation models. The scarcity of data for HIA is not the focus of this thesis, but will be discussed along the way.

Introduction to the methodology employed in this thesis

Health impact assessment starts with a policy relevant question: what will be the health consequences of executing plan A? To answer this question we need to hypothesise causal chains that run from the policy decision to health outcomes. That is, we need a conceptual model of what happens if the policy is executed, and we need to compare this with what would happen if it is not executed, or if an alternative decision is taken. This model must be based on (scientific) theory of how things work in reality. Epidemiological theory will deliver some of the

elements, but much will also have to be derived from other disciplines, depending on the policy field that needs to be modelled. Once we have a conceptual framework, we have to operationalise it by making quantitative (mathematical) models. A mathematical model is the operationalisation of how we think a system works in reality, reduced to its essentials.

We distinguish two broad steps in this framework: one from the implementation of a policy plan to changes in exposure of determinants of health, and a second that relates this change in exposure to health outcomes. We will deal with these two steps in the next paragraphs. The first step is discussed only briefly because little can be said about the methods in general, and because Chapter 2 explores the issue in more depth. The second step is discussed in more detailed because it explains the methods we aimed to apply in HIA.

From policy to determinants

To estimate the effect of a policy on exposure to determinants, various methods can be used. Methodology depends on the policy proposal at hand, and especially on the determinants it influences. To estimate a change in exposure to fine dust as a result of the construction of a new road, for example, requires different methods than the assessment of a change in income as a result of that same road. The methods required for this first step in the effect analysis often have to be found outside the field of epidemiology or other health research. This necessitates collaboration with other scientific disciplines: econom(etr)ists for changes in price and availability, environmental health specialists to predict the dispersion of pollutants, mathematicians to predict changes in road traffic flows, sociologists or other behavioural scientists to forecast changes in sexual or other health-relevant behaviour, and so forth. Standardisation of the analytical approach to this part of the causal framework may be possible for some specific determinants, but it is not possible to give one method that suits all HIAs.

Modelling shifts in exposure

Most policy decisions in HIA can be seen as ‘universal’ or ‘public health’ interventions, which are directed at everyone in a community (World Health Organization 1998). Exposure to many – but not all - determinants of health can be conceptualised as having a continuous distribution over the population, a distribution that can shift in time (Rose 1992). Policy decisions can cause such shifts in exposure distributions, often against a background of ‘autonomous’ trends that are caused by other factors (known or unknown). We therefore model exposure of populations as continuous distributions that can shift under the influence of policy decisions or trends in time. This leaves us with the question

how much the mean exposure would change as a result of the execution of the policy plan at hand. This model is not valid if an intervention affects different exposure groups to a different degree. In that case a similar but slightly more complex methodology can be used (Barendregt manuscript).

From determinants to health outcomes

The second step, from change in exposure to change in health outcomes, is more uniform across policies and determinants. This step lies firmly in the field of epidemiology: by definition, a determinant of health has been linked to health outcomes in previous (epidemiological or toxicological) research. To link exposure of populations to incidence of disease (or death), the Potential Impact Fraction (PIF) is used. The PIF can be applied to incidence or mortality, in multi-state life tables or in dynamic simulation models. The most suitable approach depends partly on the degree of precision that is needed and partly on the outcome measures that are chosen.

Potential Impact Fraction

In epidemiology, the effect of changes in the exposure to determinants on the incidence of disease is calculated using the concept of the Potential Impact Fraction (PIF) (Morgenstern and Bursic 1982). The PIF is defined as the proportional change in the incidence (or mortality) as a function of a change in exposure, and is calculated using relative risks from epidemiological literature and the following formula:

$$PIF = \frac{\sum_{c=1}^{n} p_c RR_c - \sum_{c=1}^{n} p_c^* RR_c}{\sum_{c=1}^{n} p_c RR_c} \qquad 1$$

where p_c is the proportion of the population in category c, RR_c is the relative risk for that category, and p_c* is the proportion in category c after the intervention.

For continuous risk factors the formula becomes:

$$PIF = \frac{\int_{l}^{h} RR(x)P(x)dx - \int_{l}^{h} RR(x)P^{*}(x)dx}{\int_{l}^{h} RR(x)P(x)dx} \qquad 2$$

where *RR(x)* is the risk function, *P(x)* is the original risk factor distribution, $P^{*}(x)$ the risk factor distribution after the intervention, and *l* and *h* are the integration boundaries.

The PIF is a relative measure; it indicates the proportion of cases that an intervention prevents. The absolute number of cases prevented can be calculated by multiplying the PIF by the incidence in the baseline population. Separate PIFs need to be calculated for each risk factor-disease combination. The method requires information about the present level of exposure, the expected change in exposure and the present aggregated incidence or mortality. Age-specific information is not required. This method is therefore easy to perform in most situations, but not very precise if applied to crude incidence or mortality data because it does not take the age-structure of the population into account. Another disadvantage is that it only gives incident cases and/or deaths prevented as outcomes, which precludes the use of summary measures of population health.

Multi-state life table

The Potential Impact Fraction can also be used in more elaborate methods, such as life table-based models, to link changes in risk factor exposure to changes in disease incidence. Though more elaborate than applying PIF calculations to crude incidence or mortality, the standard life table is a simple and much used instrument for the assessment of population health. However, it has the disadvantage that it ignores morbidity. Multi-state life tables that incorporate the PIF enable outcomes to be expressed in terms of summary measures of population health. For the work in this thesis, we used what has been called the 'proportional multi-state life table' (MSLT), a relatively uncomplicated model without time as a factor, that is implemented in a spreadsheet (Barendregt, Van Oortmarssen et al. 1998). The term 'proportional' refers to the fact that the model was designed to cope with a large number of diseases simultaneously while allowing for co-morbidity. Figure 1.2 gives a schematic overview of how the proportional MSLT works.

The MSLT, as we used it, models (at least) two populations: one that is designed to resemble the population of interest in the HIA, and another that is

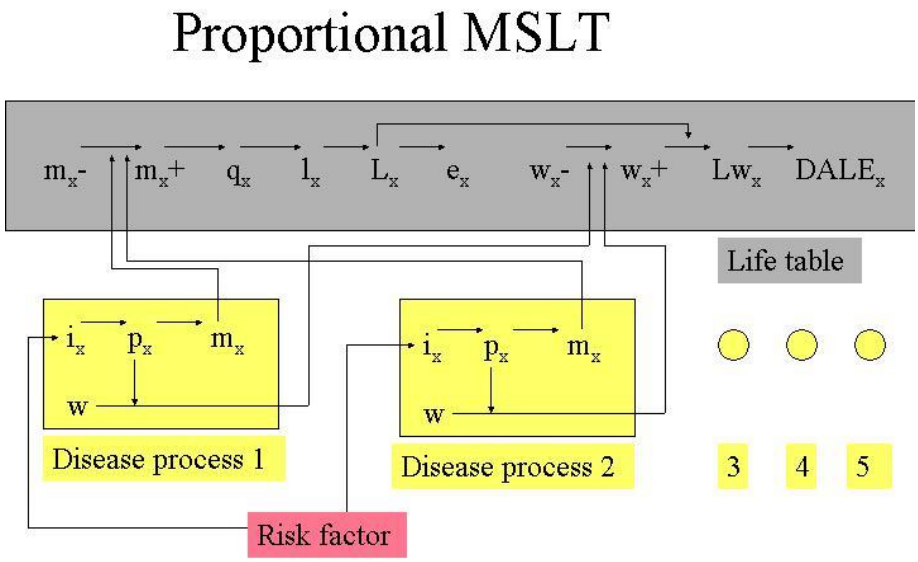
identical except that the exposure to a determinant of health can be varied. Each of these populations has a standard life table as its core, and sub-models for a number of diseases that are causally related to the determinant in question. In the proportional MSLT a change in exposure influences disease incidence. For example, an increased smoking prevalence leads to an increased incidence of chronic obstructive pulmonary disease (COPD). This will increase the number of prevalent cases at higher ages. Finally, at still higher ages mortality will follow suit. For HIA, the assumption that exposure change primarily influences disease incidence seems reasonable. Unless health care delivery is influenced, interventions outside the health care sector are mostly environmental in nature, analogous to the concept of primary prevention. Factors like diet, smoking, physical exercise and air pollution will mainly affect disease incidence, though continuing exposure may exacerbate or hasten the disease process and reduce the chances of remission.

The increased precision and greater range of outcome measures come at the price of increased data requirements compared to the simple PIF calculation. The exposure, morbidity and mortality data now need to be age-specific. The diseases have to be assigned a severity score ('disability weight') and the average health-related loss of quality of life has to be estimated for each age. The model is also more complex, which decreases transparency and increases the possibility of errors.

The MSLT allows simulation of changes in the occurrence of multiple diseases simultaneously, while accounting for substituting causes of mortality. This is necessary because if people are prevented from dying from one cause, they are at risk of dying from something else (e.g. if a cure for cancer is found, more people will die of cardiovascular disease, albeit at higher ages). The MSLT allows estimation of not only how many deaths are postponed, but also how many years of life are gained or lost by an intervention. Outcomes can be expressed at the disease-specific level, but also at the level of generic measures of population health: healthy life expectancy (DALE) and health life years saved or lost (DALYs).

A MSLT model has two dimensions, age and either cohort or period. It can be interpreted as a one-year cohort that ages with time, or alternatively as an entire population in a 'steady state' of exposure to determinants of health. In this thesis we took the latter perspective, which enabled estimation of the size of health effects, but did not enable estimation of when they would occur. A model without time dimension is called a 'static' model. If it is important to know when health gains or losses will occur (for example when discounting of health

Figure 1.2 Schematic description of how the disease-specific data are linked to the life table sections in a proportional multi-state life table (MSLT) in order to calculate health expectancy (DALE: 'disability-adjusted life expectancy'). The description covers the 'counterfactual' population in the model that is derived from the 'factual' population, which itself is not shown. All of the factors are age-specific, with the x in subscript denoting age. i = incidence, p = prevalence, m = mortality rate, m- = mortality due to diseases not included in the model, m+ mortality including the diseases in the model. w = disability weight (proportion of quality of life considered to be lost due to disease or disability), w- = proportion of quality of life lost due to diseases other than those included in the model, w+ = total proportion of quality of life considered lost due to disease, including those in the model. Lw = number of disability adjusted life years lived in age interval x, DALE = healthy life expectancy. In short, PIF-calculations for each age-sex-risk factor-disease combination allow translating risk factor changes to changes in disease incidence. This in turn changes prevalence and mortality at higher ages. The disease-specific mortality changes lead to changes in the total mortality in the life table, so that effects on life expectancy can be calculated. Similarly, the changes in prevalence lead to change in the average disability-adjusted quality of life. This allows calculating healthy life expectancy.



effects is applied, or for the planning of health care), then it is better to use a dynamic model.

Dynamic simulation models add a time factor to the static multi-state life table. This makes it possible to estimate when changes in exposure will result in health gain or loss. Again, the price for these extra attributes is an increase in the data requirements and the possibility of errors and a decrease in the transparency. In addition, we also need to know how much time it takes for an exposure to lead to an increased risk of contracting a disease, and what the past exposure and disease occurrence have been.

The models described above can be termed 'macrosimulation' models, since they have groups as their unit of analysis. Microsimulation models, in contrast, are based on the analysis of individuals. These models are better suited to the analysis of the distribution of health effects within the population (e.g. by socio-economic status). Needless to say, this demands additional data and/or assumptions.

The method used in this thesis

Our aim was to develop prediction methods that can be used in a variety of settings in HIA practice. An epidemiologist should be able to use them without much additional study. With some more effort (s)he should also be able to adapt them to suit the population and the determinants of interest. In the work for this thesis we used the multi-state life table, which incorporates the PIF concept. The review of methods we performed at the start of the project (see chapter 2) showed that the PIF/MSLT had not been applied in the context of HIA, although it seemed to have suitable characteristics. In complexity and data requirements, this method is mid-way between applying the PIF to incidence or mortality data on the one hand, and dynamic modelling on the other. We judged the simple PIF calculation to be too crude for comprehensive HIA, while dynamic models require expertise of a level that is not widely available. The MSLTs are implemented in a spreadsheet, which makes them transparent and flexible. We therefore used the MSLT in our studies on the potential impact of changes to the European policy on fruits and vegetables and on a health education intervention.

Scientific and policy conditions influence the scope for methods in HIA

Two sets of circumstances impact on the ability to make predictions in HIA: the scientific environment and the policy/administrative environment. The scientific factors comprise the availability of methods and data. The policy environment

determines what resources are available in terms of manpower and time, but also what outcomes are expected and how the results are expected (or permitted) to influence the decision making process. HIA has to connect to a policy process over which the HIA practitioners usually have little influence, so this process dictates the timing of results and the amount of time available. In such circumstances, ready-for-use data and methods can provide flexibility. We developed, tested and applied models and describe what scientific and policy barriers we encountered. The emphasis in this thesis is on the quantitative methods to predict health effects. Administrative and policy aspects of HIA are dealt with in more depth in the thesis of Marleen Bekker (Bekker 2007), although we touch upon these issues in the general discussion (Chapter 8). The two research projects were conducted in close contact and resulted in a joint report on methods for HIA (in Dutch (Bekker and Veerman 2007)).

This thesis

This thesis aims to develop and test methods for the prediction of health effects in HIA. The core of the method consists of a two-step process via determinants of health, using macrosimulation modelling to reason from exposure to health outcomes. This model was applied to different policy measures: a change in the EU agricultural policy on fruits and vegetables, a hypothetical national campaign using computer-tailored health education regarding fruit- and vegetable consumption, obesity prevalence in the US, and food advertising on TV targeting children.

This thesis aims to contribute to the development of methods for the quantitative prediction of health effects in HIA. The specific questions which will be addressed are:

1. What methods have been used in previous studies to make quantified predictions of the health impact of policy outside the health sector?
2. Can a combination of potential impact fractions and multi-state life tables be used to predict changes in summary measures of population health due to policies outside the health sector?
3. How can the validity of predictions in HIA be established?

As explained above, the ability to express health changes in terms of summary measures of population health enables comparison between different policies and policy options, and it can form the basis for cost-effectiveness analyses of measures that aim to mitigate negative health effect or enhance positive ones.

The structure of this thesis

This thesis comprises three sections. The first section (Chapter 2) explores question 1 by examining what methods have been used to make quantified predictions in HIA practice to date.

The second section attempts to answer question 2 and provides forecasts of health consequences in a number of different policy issues with the use of mathematical (macro-)simulation modelling. Chapter 3 explores the potential health gain for the Dutch population of a hypothetical change in the European Common Agricultural Policy regarding fruits and vegetables. Chapter 4 estimates the potential health effects of a national campaign of computer-tailored fruit and vegetable promotion. Chapter 5 predicts the obesity prevalence the US will be facing if present trends in average body mass continue. Chapter 6 estimates how much childhood obesity can be prevented in the Netherlands by limiting exposure of children to food advertising on television.

The third and last section consists of chapter 7, which proposes a theoretical framework for the assessment of the validity of predictions in HIA.

Finally, in chapter 8 a summary of results is given, the research questions are answered and suggestions for further research and development are made.

References

Acheson, D. (1998). Independent Inquiry into Inequalities in Health. London, The Stationery Office.

Alenius, K. (2001). Consideration of health aspects in environmental impact assessments for roads. Umea (Swe), National Institute of Public Health, Sweden**:** 32.

Barendregt, J. J. (manuscript). "Categorical versus continuous risk factors, and the calculation of potential impact fractions."

Barendregt, J. J., G. J. Van Oortmarssen, et al. (1998). "Coping with multiple morbidity in a life table." Math Popul Stud **7**(1): 29-49, 109.

Barnes, R., A. Cooke, et al. (2001). Health impact assessment of regeneration programmes. London, Borough of Hammersmith and Fulham.

Bekker, M. P. M. (2007). Lost in translation. Reframing Health Impact Assessment for health policy integration. Institute of Health Policy and Management. Rotterdam, Erasmus University Rotterdam.

Bekker, M. P. M. and J. L. Veerman (2007). Gezondheidseffectschatting voor Gezond Beleid: Teamsport op projectbasis. Rotterdam, Erasmus MC.

Birley, M. (1995). The health impact assessment of development projects. London, HMSO.

Birley, M. (2004). HIA in developing countries. Health impact assessment. Concepts, theory, techniques and applications. J. Kemm, J. Parry and S. Palmer. Oxford, OUP.

Cambridgeshire Health Authority (2001). Alconbury health impact assessment Final report. Huntingdon, Cambridgeshire Health Authority,.

Drummond, M. E., B. J. O'Brien, et al. (1997). Methods for the economic evaluation of health care programmes. Oxford, Oxford University Press.

ECHP, W.-R. (1999). Health Impact Assessment - main concepts and suggested approach. Gothenburg Consensus Paper. Health Impact Assessment: From theory to practice, Gothenburg, Sweden.

EPHIA Working Group (2004). European policy health impact assessment - a guide. Brussels, European Commission.

European Centre for Health Policy (1999). Health Impact Assessment - main concepts and suggested approach. Gothenburg Consensus Paper. Health Impact Assessment: From theory to practice, Gothenburg, Sweden, ECHP/ World Health Organization Regional Office for Europe.

Fehr, R. (1999). "Environmental health impact assessment: evaluation of a ten-step model." Epidemiology **10**(5): 618-25.

Gold, M. R., J. E. Siegel, et al., Eds. (1996). Cost-effectiveness in health and medicine. New York, Oxford University Press.

Kemm, J. (2003). "Perspectives on health impact assessment." Bull World Health Organ **81**(6): 387.

Kemm, J. and J. Parry (2004). The development of HIA. Health impact assessment. Concepts, theory, techniques and applications. J. Kemm, J. Parry and S. Palmer. Oxford, OUP.

Kemm, J. R. (2000). "Can health impact assessment fulfil the expectations it raises?" Public Health **114**(6): 431-3.

Lalonde, M. (1974). A New Perspective on the Health of Canadians
(Lalonde Report). Ottawa, Ministry of Supply and Services.

Lester, C. and M. Temple (2004). "Rapid collaborative health impact assessment: a three-meeting process." Public Health **118**(3): 218-24.

Mindell, J., A. Hansell, et al. (2001). "What do we need for robust, quantitative health impact assessment?" J Public Health Med **23**(3): 173-8.

Morgenstern, H. and E. S. Bursic (1982). "A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population [abs]." J Community Health **7**(4): 292-309.

Murray, C. J., Ed. (2002). Summary Measures of Population Health: concepts, ethics, measurements and applications. Geneva, World Health Organization.

Øvretveit, J. (1998). Evaluating Health Interventions: an introduction to evaluation of health treatments, services, policies, and organizational interventions. Buckingham, Open University Press.

Parry, J. and A. Stevens (2001). "Prospective health impact assessment: pitfalls, problems, and possible ways forward." Bmj **323**(7322): 1177-82.

Roscam Abbing, E. W. (2004). HIA and national policy in the Netherlands. Health impact assessment. Concepts, theory, techniques and applications. J. Kemm, J. Parry and S. Palmer. Oxford, OUP.

Rose, G. (1992). The Strategy of Preventive Medicine. Oxford, Oxford University Press.

Scott-Samuel, A. (1996). "Health impact assessment: an idea whose time has come." Bmj **313**(7051): 183-4.

Scott-Samuel, A. B., M; Adern, K. (2001). The Merseyside Guidelines for Health Impact Assessment. Liverpool, The University of Liverpool.

Tan-Torres Edejer, T., R. Baltussen, et al., Eds. (2003). Making choices in health: WHO guide to cost-effectiveness analysis. Geneva, World Health Organization.

World Health Organization (1998). Obesity: Preventing and Managing the Epidemic. Geneva, WHO.

PART A

Quantification in health impact assessment

2

Quantitative health impact assessment: current practice and future directions

Based on: Veerman JL, Barendregt JJ, Mackenbach JP. Quantitative health impact assessment: current practice and future directions. *Journal of Epidemiology and Community Health.* 2005 May;59(5):361-70.

Summary

Objective
To assess what methods are used in quantitative health impact assessment (HIA), and to identify areas for future research and development.

Design
HIA reports were assessed for (1) methods used to quantify effects of policy on determinants of health (exposure impact assessment) and (2) methods used to quantify health outcomes resulting from changes in exposure to determinants (outcome assessment).

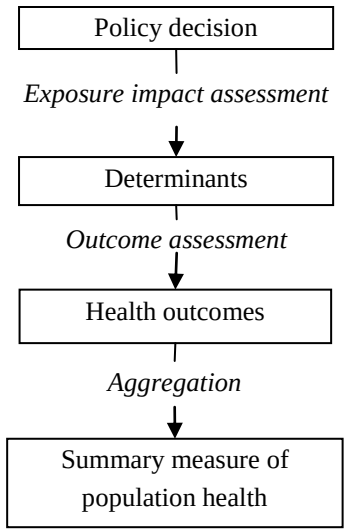
Results
Of 98 prospective HIA studies, 17 reported quantitative estimates of change in exposure to determinants, and 16 gave quantified health outcomes. Eleven (categories of) determinants were quantified up to the level of health outcomes. Methods for exposure impact assessment were: estimation on the basis of routine data and measurements, and various kinds of modelling of traffic related and environmental factors, supplemented with experts' estimates and author's assumptions. Some studies used estimates from other documents pertaining to the policy. For the calculation of health outcomes, variants of epidemiological and toxicological risk assessment were used, in some cases in mathematical models.

Conclusion
Quantification is comparatively rare in HIA. Methods are available in the areas of environmental health and, to a lesser extent, traffic accidents, infectious diseases, and behavioural factors. The methods are diverse and their reliability and validity are uncertain. Research and development in the following areas could benefit quantitative HIA: methods to quantify the effect of socioeconomic and behavioural determinants; user-friendly simulation models; the use of summary measures of population health, expert opinion and scenario building; and empirical research into validity and reliability.

Introduction

As explained in Chapter 1, quantification of health effects in HIA is desirable. However, there are two difficulties in quantification: the availability of valid data, and the availability of methods to analyse the data and translate them into information on the health effect of the proposal under scrutiny. In this contribution we focus on the second problem and analyse reports of HIAs performed to date, using a framework similar to that proposed by Joffe and Mindell (Joffe and Mindell 2002) in which policy decisions influence health via its determinants (see fig. 2.1). This divides the HIA process in two steps, which for brevity we will refer to as "exposure impact assessment" and "outcome assessment" respectively. This chapter addresses two questions: Firstly, what methods are used in quantitative exposure impact assessment, and secondly, what methods are used for quantitative outcome assessment?

Figure 2.1 Conceptual model for health impact assessment



Methods

Search strategy

HIA case studies were obtained by searching (1) electronic sources (www.who.int/hia and the links it contains, accessed 20 Jul 2004), (2) the reviews analysed in Taylor and Quigley's "HIA review of reviews"(Taylor and Quigley 2002), and (3) the complete collection of Dutch national level HIA cases provided by the national coordinating agency for intersectoral health policy (Ondersteuningsfunctie Integraal Gezondheidsbeleid). (4) To include recently published cases we searched PubMed for the period 2002 until 1 August 2004. Strings used were "health AND impact AND assessment", and "impact AND (assess* OR eval*) AND (policy OR policies)". Resulting articles were first judged by title and abstract, and obtained in case of possible reference to case studies. Promising reports were requested from the authors.

Selection

We included reports of primary studies that were prospective and assessed the impact of non-health sector policy decisions. Descriptions of case studies in published articles and reviews were used if they contained sufficient information, but in most cases the original reports were obtained. We excluded reports of studies that had not been completed at the time the report was written, reports that were very restrictive in the health outcomes they presented (such as studies on the effect of bicycle helmet use on fatal injuries), and studies that only screened whether a particular proposal was health relevant. We also excluded reports in languages other than English, French, German, Spanish, or Dutch.

Analysis

We distinguish between methods used to estimate the effect of policy on determinants, and methods that calculate health impact given those changes in the exposure to determinants. The estimates of exposure to determinants and of health outcomes were assessed for quantification. Quantification was defined as the expression in numerical terms of the change in health status of a specific population that can be attributed to a specific policy decision.

Results

Figure 2.2 shows the results. The electronic sources yielded 83 cases of prospective HIA; 12 possible cases could not be obtained. This partly overlapped with the 25 studies obtained from the reviews, of which two possible cases could not be obtained. Out of the collection of Dutch HIA cases 12 studies were included. The search in PubMed added eight cases; two possible cases were not obtained. Of the total of 98 cases, 72 were from the UK, and 12 from the Netherlands (table 2.1). The Dutch HIA studies focus on national level policy, while all but one of the other studies assess local or regional level projects or programmes.

Of the 98 studies, 17 gave quantified estimates of the effect on determinants of health. Table 2.1 gives a description of the studies. Ten of the studies deal with physical infrastructure for industry or transport, the remaining seven focus on a variety of projects and policies. Sixteen studies proceed to present health outcomes.

Figure 2.2 Results search strategy. Eighty-one prospective HIA studies were found. Due to overlap, the results of the different sources seem to add up to more than that.

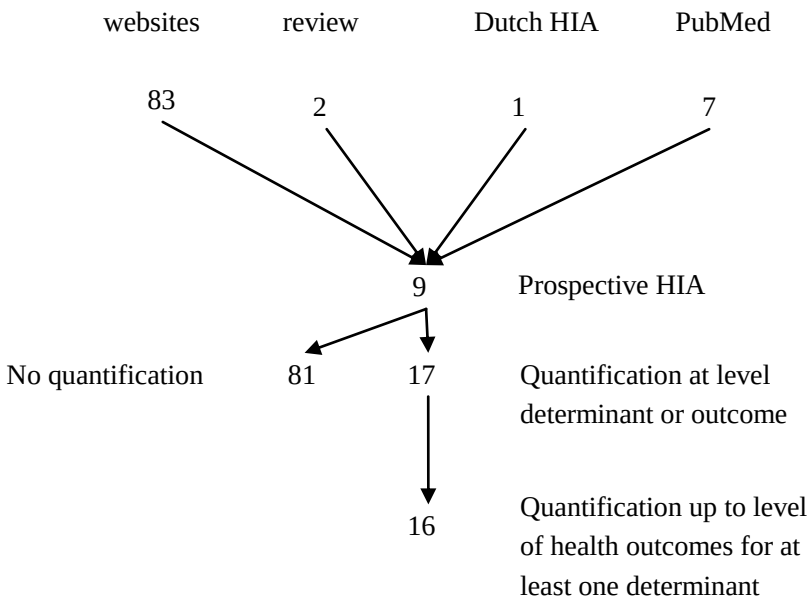


Table 2.1 Analysis of methods for quantification in Health Impact Assessment

	Exposure impact assessment	**Outcome assessment**	**Health outcomes**
Waste-to-energy facility (Hallenbeck 1995)	Pollution levels based upon worst case emission factors derived from existing similar facilities; computerised air-pollution dispersal model.	Carcinogens: maximum annual average ground level concentrations were used to compute maximum theoretical cancer risks due to lifetime inhalation, using risk factors from USEPA (i.e. quantitative risk assessment).	Lifetime cancer risks resulting from each pollutant separately (e.g. maximum theoretical cancer risk due to lifetime exposure to PCBs: 10^{-8}).
Freeway Melbourne (Dunt, Abramson et al. 1995)	Literature study; base-line data on accidents, pollution and noise levels; models predicting traffic-flows.	Literature study; model to predict traffic accidents; extrapolation on basis of data on current percentage of road traffic accidents with casualties.	Road traffic accidents (-154) and resulting injuries (-100 to 110 separate injuries).
Law on alcohol and hotel and catering industry (Zwart 1997)	A (separate) economic effect analysis gave an estimate of the change in revenue from sale of alcohol. This was interpreted as a decrease in per capita consumption.	'Ledermann-formula' to estimate the number of excessive drinkers on basis of per capita consumption.	Excessive drinkers (-11.000 persons drinking >7 E alcohol /day; -5.000 persons with >11 E/day; -3.000 persons with >15 E/day).
HIA on tobacco policy (Mooy and Gunning-Schepers 1998)	Economic literature gave an estimate of the change in amount sold as a function of price (price-elasticity) of tobacco. This was equated with a corresponding decrease in smoking prevalence.	Computer simulation model (PREVENT), a dynamic multi-state life-table-based model.	Deaths from lung cancer in absolute numbers prevented per year; life years gained (both graphically depicted as numbers per calendar year).
Extension of waste disposal site (Kobusch, Serwe et al. 1995; Fehr 1999)	Expected emissions based on extrapolation of measurements at other sites, supplemented by data from technical literature; pollution dispersal models for chemicals incl. 8 carcinogens (by external agency); noise from technical statement.	Quantitative Risk Assessment (QRA) for carcinogens using Health Risk Assessment computer programme.	Individual additional lifetime cancer risk (graphically depicted per village and pathway); estimated additional cancer burden (<0.01 additional cases in 70 years).
City Bypass road Krefeld (Serwe and Protoschill-Krebs 1997; Fehr 1999)	Changes in traffic flow (incl. resulting changes in noise) estimated in environmental impact assessment report; air-pollution dispersion models.	Quantitative Risk Assessment (QRA) for 2 carcinogens using Health Risk Assessment computer program.	Individual additional lifetime cancer risk; additional cancer burden.

	Exposure impact assessment	Outcome assessment	Health outcomes
Home Energy Efficiency Scheme (Kemm, Ballard et al. 2000)	Expert opinion (of authors) to estimate the increase in the home "comfort range" (the range of outside temperatures at which indoor temperature can be kept at or above 16° C, below which mortality starts to rise); CO2 emission using emissions inventory and estimated energy savings; expert (authors?) estimate of effect on employment on basis of financial investment.	Ecological data on the effect of (outside) temperature on mortality were combined with climate data for Wales after applying an increase in the home comfort range for the number of people benefiting.	Number of deaths per year (1 or 2, on basis of 30,000 homes treated).
Alconbury Airport (France and Lilley 2000)	Used estimate of project planner for vehicle movements and planning application estimates for increase in PM_{10} (2.5% increase; assumed base level at 50 microgram/m^3)	Application of rates of death / illness resulting from PM_{10} from WHO study; three methods / models for prediction of road traffic accident injuries on basis of vehicle movements; application of regional data on fatality of road traffic accidents.	Injury-only accidents (1 to 19 per year); RTA-induced fatality (1 per 3-57 years); "long term mortality" from air pollution (0.2 deaths/yr); respiratory hospital admissions (0.1/yr); cardiovascular hospital admissions (0.3/yr; chronic bronchitis incidents in adults over 25 (0.2/yr); bronchitis in children under 15 (0.7/yr); restricted activity days (170/yr); asthma attacks in children under 15 (0.4); asthma attacks in adults over 15 (4.3/yr).
Woodprocessing plant (Kemm 2000)	Used estimates from environmental impact report.	Application of national accident rates (per 100 million vehicle-km).	Deaths due to road traffic accidents (1 per 10 yrs.) and injury-only accidents (1 per 2 yrs).
Finningley Airport (Abdel-Aziz, Redford et al. 2000)	Data from planning application (employment) and Environmental Statement (noise, air pollution); estimate air craft risk by MRC Environment & Health (method not stated).		
National Botanic Garden of Wales (Kemm and Breeze 2000)	Business plan (employment); (uncomplicated) economic estimates to estimate income rise; estimates of numbers of visitors travelling by car, and average distance travelled.	Estimation on basis of UK data on income and mortality rates by income decile; application of national accident rates (per 100 million vehicle-km).	Reduction in overall mortality (2%) resulting from income rise; annual car user and pedestrian casualties, incl. deaths and severely injured (e.g. 0.063 car user deaths/yr).

	Exposure impact assessment	Outcome assessment	Health outcomes
Port Southampton Dibden Bay (Taylor, Soloman et al. 2001)	Used pollution data (PM_{10}) and noise levels from EIA report; Used traffic projections and estimates on amount of employment generated from Technical Statements.	COMEAP dose-response curves for PM_{10}; traffic projections are combined with present traffic accident rates and literature on effects of mitigation measures; Scott-Samuel's method of estimating effect of (un-) employment on mortality.	Annual respiratory deaths due to air pollution (0.01); hospital admissions for respiratory problems (one per 3-4 years); personal injury road traffic accidents (17.3 per year, which equals a 12% rise – assuming no mitigation measures are taken. Possibly an additional one due to more heavy traffic). Estimates of effect of speed limits (-4.7% or 1-2 reported personal injury accidents) and street lighting (-9% or 2-3 personal injury accidents/yr) are also given, as are premature deaths due to unemployment (just over 3 avoided annually).
HIA Chad-Cameroon Oil pipeline project (Jobin 2003)	Estimated truck-miles on basis of preliminary project plans; applied data from malaria study in Nigeria to estimated population exposed; used computer simulation model with estimates on HIV-seroprevalence and frequency of sexual contacts based on previous epidemiological studies, and estimated change of partner due to project.	Applied US accident data multiplied by 10 to account for road conditions; applied Nigerian malaria death rates, and assumed effects of prevention and curative services; computer simulation model included HIV transmission risk.	Road traffic deaths (2,5 per year); deaths due to malaria (3 per year); deaths due to HIV/AIDS (70 per year).
HIA Drinking water privatisation (Fehr, Mekel et al. 2003)	Difference between current levels of 6 carcinogens in drinking water and maximum allowed levels is calculated; exposure is estimated using population number, and standardized estimates and measured data on tap water intake and body weight.	"Standard methodology of quantitative risk assessment"	Additional cancer cases from a life-time of exposure (max. 10938 cases for population of 18 million)
Foot & mouth disease disposal options (UK Dept of Health 1999)	Air emissions calculated for arbitrary numbers of cattle disposed of; data on percentage of BSE-infected older cattle.	Dispersal modelling for BSE	Cases of variant Creutzfeldt-Jakob Disease as function of %-age older cattle disposed of by pyre-burning or burying.

	Exposure impact assessment	Outcome assessment	Health outcomes
Foresight vehicle initiative (Abrahams 2002; Abrahams 2004)	Used consensus panel, email discussion and other methods with expert groups to estimate changes in determinants due to new technologies.	An epidemiological model (ARMADA)(McCarthy, Biddulph et al. 2002) was used to estimate consequences of road traffic accidents and air pollution.	Over period 2000-2029: number of deaths (±20) and serious injuries (±2000) (traffic accidents); first hospital admissions for respiratory and cardiovascular disease (±23,000) (air pollution).
Regional Planning Guidance Transport Chapter & West Midlands local plan (Pitches and Kemm 2003)	Uses scenarios to estimate effect on physical activity, vehicle miles and air pollutant emissions (not exposure); draft assessment report of Highways Agency for traffic flows.	Estimates of health consequences of physical activity increases; model for traffic accident injuries as function of vehicle miles.	All cause mortality, cardiovascular & colon carcinoma deaths, cardiovascular events (physical activity); slight, serious and fatal injuries (traffic accidents)

Table 2.2 Determinants of which quantified estimates were given, the evidence base and methods or sources used, whether a quantified estimate of health outcomes was provided and if so, what method was used to obtain this estimate.

	Method 1: from policy to determinants (exposure impact assessment)							**Quantified health outcomes**	**Method 2: from determinants to health outcomes (outcome assessment)**				
Determinant	Literature review	Data, own measurements	Existing models	Expert opinion	Own assumptions	Other reports pertaining to plan			Literature review	Data, analogy	Modelling	Expert opinion	Own assumptions
						From planner	In-dep.						
Environmental pollutants													
Air (10 carcinogens, 13 non-carcinogenic substances) (Hallenbeck 1995)		X	X					Partly: carcinogens	X (tox[1])				
Air (25 gasses) (Kobusch, Serwe et al. 1995)		X	X					Partly: carcinogens	X (tox)				
Air (6 substances) (Serwe and Protoschill-Krebs 1997)		X	X					Partly: carcinogens	X (tox)				
Air (14 substances + odour) (Abdel-Aziz, Redford et al. 2000)							X	No					
Air (PM10) (France and Lilley 2000)						X		Yes	X (epi[2])				
Air (PM10) (Taylor, Soloman et al. 2001)							X	Yes	X (epi)				
Air (SO_2, PM) (UK Dept of Health 1999)			X					No					
Air (PM10) (Abrahams 2002; Abrahams 2004)		X		X				Yes			X		
Range of pathways (carcinogens) (Fehr 1999)		X	X					Yes	X (tox)				

[1] Tox = toxicological literature
[2] Epi = epidemiological literature

	Method 1: from policy to determinants (exposure impact assessment)							Quantified health outcomes	Method 2: from determinants to health outcomes (outcome assessment)				
Determinant	Literature review	Data, own measurements	Existing models	Expert opinion	Own assumptions	Other reports pertaining to plan			Literature review	Data, analogy	Modelling	Expert opinion	Own assumptions
						From planner	In-dep.						
Water (6 carcinogens) (Fehr, Mekel et al. 2003)		X			X			Yes	X (tox)				
CO2 emission (greenhouse effect) (Kemm, Ballard et al. 2000)		X			X			No					
Noise (Kobusch, Serwe et al. 1995)		X				X		No					
Noise (Serwe and Protoschill-Krebs 1997)							X	No					
Noise (Abdel-Aziz, Redford et al. 2000)						X?		No					
Noise (Taylor, Soloman et al. 2001)							X	No					
Traffic accidents													
Vehicle kilometres (Dunt, Abramson et al. 1995)		X	X					Yes		X Local data	X		X
Vehicle kilometres (France and Lilley 2000)						X		Yes		X Local data	X		X
Vehicle kilometres (Kemm 2000)							X	Yes		X Nat. data[3]			
Vehicle kilometres (Kemm and Breeze 2000)					X	X		Yes		X Nat. data			

[3] Nat = national

	Method 1: from policy to determinants (exposure impact assessment)							Quantified health outcomes	Method 2: from determinants to health outcomes (outcome assessment)				
Determinant	Literature review	Data, own measurements	Existing models	Expert opinion	Own assumptions	Other reports pertaining to plan			Literature review	Data, analogy	Modelling	Expert opinion	Own assumptions
						From planner	In-dep.						
Vehicle kilometres (Taylor, Soloman et al. 2001)						X		Yes		X Local data			
Vehicle miles (Pitches and Kemm 2003)					X			Yes		X Nat. data			
Truck miles (Jobin 2003)					X	X		Yes		X Int'l. data[4]			X
Vehicle safety (Abrahams 2002; Abrahams 2004)		X		X				Yes		X Nat. data	X		
Air craft crash risk (Abdel-Aziz, Redford et al. 2000)							X	No					
Socioeconomic													
Employment (Kemm, Ballard et al. 2000)					X			No					
Employment (Kemm 2000)							X	No					
Employment (Kemm and Breeze 2000)						X		No					
Employment (Abdel-Aziz, Redford et al. 2000)						X		No					
Employment (Taylor, Soloman et al. 2001)						X		Yes	X				
Income (Kemm and Breeze 2000)					X			Yes		X Nat. data			X

[4] Int'l = international

	Method 1: from policy to determinants (exposure impact assessment)							**Quantified health outcomes**	**Method 2: from determinants to health outcomes (outcome assessment)**				
Determinant	Literature review	Data, own measurements	Existing models	Expert opinion	Own assumptions	Other reports pertaining to plan			Literature review	Data, analogy	Modelling	Expert opinion	Own assumptions
						From planner	In-dep.						
Behaviour / addictions													
Alcohol consumption (Zwart 1997)					X		X	Partly	X	X			
Smoking (Mooy and Gunning-Schepers 1998)	X							Yes (mortality only)	X		X		
Physical activity (Pitches and Kemm 2003)		X			X			Yes	X		X		
Physical environment													
Housing: in-house temperature range (Kemm, Ballard et al. 2000)					X			Yes	X	X			X
Infectious disease													
Malaria infection (Jobin 2003)					X	X		Yes					X
Sexual contact with possibility for HIV transfer (Jobin 2003)			X					Yes		X	X		
BSE/variant Creutzfeldt-Jakob Disease (UK Dept of Health 1999)			X					Yes			X		
TOTAL	***1***	***11***	***8***	***2***	***10***	***11***	***8***	***Yes: 21 (57%) Partly: 4 (11%) No: 12 (32%)***	***12***	***12***	***8***	***0***	***6***
						19							

Table 2.2 provides an overview of the determinants that were quantified, and on what sources of information the estimates were based. We will now discuss the methods used for the assessment of exposure impact and outcome for the 10 (categories of) determinants for which quantified health outcomes were presented.

1. Carcinogens

For carcinogens and other environmental factors computerised dispersal models were used to predict spread and subsequently the exposure of human populations was estimated (Hallenbeck 1995; Fehr 1999; Fehr, Mekel et al. 2003). In the outcome assessment phase, existing computer programs and toxicological data were used (Hallenbeck 1995; Fehr 1999).

2. Particulate matter

Particulate matter with a diameter less than 10 μm in size (PM10) was used as proxy for air pollution (UK Dept of Health 1999; France and Lilley 2000; Taylor, Soloman et al. 2001; Abrahams 2002). Models similar to those for carcinogens were used to estimate future exposure. For the outcome assessment authors made use of epidemiological data (Künzli, Kaiser et al. 2000).

3. Road transport: vehicle kilometres

The common determinant used in most predictions of traffic injuries and fatalities is vehicle kilometres–that is, the total number of kilometres driven by vehicles. In four out of nine cases no independent exposure impact assessment was done: the estimates are obtained from the project planner without mention of the methods used. The methods to estimate health effects given the number of vehicle kilometres differ. In three studies the expected increase in vehicle-kilometres is simply multiplied by the local or national accidents rates per kilometres (France and Lilley 2000; Kemm 2000; Kemm and Breeze 2000) while others use more complex methods and take into account road type and mode of transport (Pitches and Kemm 2003) or traffic flows and types of intersections (Dunt, Abramson et al. 1995; Taylor, Soloman et al. 2001). Jobin uses data from the US and multiplies the accident rate by ten in order to estimate the number of truck related casualties in Cameroon (Jobin 2003).

4. Employment

Five studies provide estimates of the employment effects of plans. In four cases no independent assessment was done and the estimates were obtained from documents pertaining to the project plan (Abdel-Aziz, Redford et al. 2000;

Kemm 2000; Kemm and Breeze 2000). In one case it was estimated on the basis of financial investments (Kemm, Ballard et al. 2000). Only one author provides an estimate of the effect on health (in the form of premature mortality) based on extrapolation from a longitudinal study (Scott-Samuel 1984; Taylor, Soloman et al. 2001). Others cite this study but conclude that the size of the effect is difficult to estimate.

5. Income

Kemm estimates the effect of the cash injection into the local economy represented by the creation of a botanical garden and divides this by the number of residents in the population. In the outcome assessment phase data on the relationship between income deciles in the UK and mortality are used to estimate the reduction in mortality (Kemm and Breeze 2000).

6. Alcohol

Zwart equates the alcohol sales decrease predicted in an independent economic analysis with alcohol consumption decrease on a population level (Zwart 1997). In the outcome assessment phase he applies the Ledermann formula to estimate the number of excessive drinkers. This formula supposes a log normal distribution of alcohol consumption in a population.

7. Smoking

Mooy and Gunning-Schepers use economic literature on the price elasticity of cigarettes (that is, the relationship between price and sales) to estimate the change in the number of smokers resulting from increased tobacco taxing. For the outcome assessment the Prevent model is applied, a macro-simulation model with a dynamic population that incorporates epidemiological data (Mooy and Gunning-Schepers 1998).

8. Physical activity

Pitches and Kemm estimate the effect of changes in transport infrastructure on the annual distance cycled and walked. This is translated into the number of people that would move from being sedentary to physically active. For the outcome assessment phase they constructed a simple model using survey data on physical activity, mortality and morbidity statistics and relative risk estimates from meta-analyses and estimate the number of lives (or cases of disease) saved if 1000 persons would increase their activity level (Pitches and Kemm 2003).

9. Housing

Kemm et al. estimate the effect of home insulation on the "comfort range", the lowest outside temperatures at which the in-house temperatures can be kept at or above 16°C. In the outcome assessment this estimate is linked to ecological data on the effect of outside temperature on overall mortality to arrive at the number of deaths avoided (Kemm, Ballard et al. 2000).

10. Infectious disease

To estimate the burden of malaria attributable to the construction of an oil pipeline in Cameroon, Jobin used Nigerian data in combination with assumptions on the effectiveness of preventive and curative measures. For HIV a computerised transmission model and epidemiological data on prevalence were used (Jobin 2003). Finally, chances of causing variant Creutzfeldt-Jakob disease by burning cattle carcasses were estimated using a dispersal model (UK Dept of Health 1999).

Type of evidence

Table 2.2 shows that in exposure impact assessment most assessors used the results of other reports pertaining to the policy under scrutiny, such as environmental impact assessments for exposure to chemical substances or project plans for the expected amount of employment generated. Other sources of evidence are (routine) data, measurements by the researchers, and pre-existing models (which contain data and assumptions). In some cases the evidence is supplemented with the author's own assumptions, and in two studies expert opinion was sought. A review of the literature was used explicitly in only one study. In contrast, outcome assessment was commonly based on literature reviews, as well as on routine data and pre-existing models. Expert opinion (other than that of the authors) was not used in outcome assessment.

Analysis of methods

Table 2.3 gives a further analysis of the 16 studies that presented health outcomes. This shows that the types of health outcomes used differ greatly between studies. Furthermore, 14 of the assessments are limited to the effects of proximal determinants, while two studies also include effects of determinants that may be considered distal (defined as exerting their influence via intermediate factors): increased income and employment. The methods used for

Table 2.3 Analysis of quantitative HIA studies: types of determinants, health outcomes and modelling, source of risk measures, and uncertainty.

	Study	Proximal/distal determinants[5]	Type of health outcomes	Type of modelling in outcome assessment	Source of risk measures linking determinant to health	Time horizon	Uncertainty
1	Waste facility (Hallenbeck 1995)	Proximal	Life time cancer risks	Quantitative Risk Assessment (QRA)	Toxicological risk assessment based upon animal experiments	70 years	Point estimates of maximum exposure only
2	Highway (Dunt, Abramson et al. 1995)	Proximal	Road traffic injuries	Traffic accident model	Local historical data on accidents by type of crossing	10 years	Report includes sensitivity analysis
3	Alcohol (Zwart 1997)	Proximal	Number of excessive drinkers	Simple algorithm (Ledermann formula)	Cross-country comparison	Unclear	Point estimate only
4	Tobacco (Mooy and Gunning-Schepers 1998)	Proximal	Disease-specific death rate, life years gained	Macro-simulation model with dynamic population (PREVENT)	Epidemiological evidence	50 years	Point estimates only
5	Waste facility (Fehr 1999)	Proximal	Additional lifetime cancer risk, additional pop. cancer burden	Quantitative Risk Assessment (QRA) computer programme	Toxicological risk assessment based upon animal experiments	70 years	Point estimates only
6	Highway (Fehr 1999)	Proximal	Additional lifetime cancer risk, additional pop. cancer burden	Quantitative Risk Assessment (QRA) computer programme	Toxicological risk assessment based upon animal experiments	70 years	Point estimates only
7	Home insulation (Kemm, Ballard et al. 2000)	Proximal	Deaths per year	Innovative method	Ecological studies	Unclear	Point estimate dubbed "no more than a possible fig."
8	Freight distribution centre (France and Lilley 2000)	Proximal	Mortality, morbidity, restricted activity days, hospital admissions.	Epidemiological risk assessment; three formulas for prediction of traffic accidents	Epidemiological studies; regional data.	Unclear	Air pollution: average and max. rates. Road traffic accidents: 3 point estimates.
9	Woodprocessing plant (Kemm 2000)	Proximal	Road traffic accident deaths and injury-only accidents	Analogy	National historical data on average number of accidents per distance travelled.	Unclear	Point estimate, uncertainty expressed by prefix 'about'.

[5] Distal determinants affect health through intermediary factors. An example is income, which affects health via material circumstances, access to care, self-esteem, etc.

	Study	Proximal/distal determinants[5]	Type of health outcomes	Type of modelling in outcome assessment	Source of risk measures linking determinant to health	Time horizon	Uncertainty
10	Botanic garden (Kemm and Breeze 2000)	Proximal (accidents) and distal (income)	Road traffic accident injuries and deaths and %-age change in overall mortality (income)	Analogy (accidents); extrapolation from national data	National historical data on average number of accidents per distance travelled.	Unclear	Point estimates based upon 'very uncertain assumptions'.
11	Port extension (Taylor, Soloman et al. 2001)	Proximal (air pollution, accidents) and distal (employment)	Respiratory deaths and hospital admissions; road traffic accident injuries; premature deaths (due to unemployment)	Epidemiological risk assessment (air pollution, employment); analogy (accidents).	Epidemiological studies (air pollution); local data (traffic accidents); single epidemiological (longitudinal) study (employment)	Unclear	Point estimates only (air pollution); terms like '3-4' and '<1' (traffic casualties); 'just over three deaths' (employment).
12	Chad pipeline (Jobin 2003)	Proximal	Deaths per year	Analogy (accidents, malaria), infectious disease model (HIV)	USA historical data (accidents); epidemiological study Nigeria (malaria), ? epidemiological data (HIV transmission rates)	Unclear	Point estimates of 'likely' effects, intended to rank issues.
13	Drinking water (Fehr, Mekel et al. 2003)	Proximal	Additional lifetime cancer risk, additional cancer cases in pop.	Quantitative Risk Assessment (QRA)	Toxicological risk assessment based upon animal experiments (?)	70 years	Point estimates and 5th and 95th percentiles per increase of exposure (uncertainty in degree increase not quantified)
14	UK Dept. of Health (2001): Foot & mouth disposal	Proximal	Additional variant Creutzfeldt-Jakob Disease infections as function of %-age of older cattle destructed.	Unclear	Unclear	Unclear	Point estimates with 95% confidence range.
15	***Foresight vehicle initiative (Abrahams 2004)***	Proximal	Annual deaths and serious injuries (traffic accidents); first hospital admissions (air pollution)	Macro-simulation model with dynamic population (ARMADA) (McCarthy, Biddulph et al. 2002)	Epidemiological evidence	2000 - 2029	Point estimates, uncertainty expressed by prefix 'approximately'
16	Regional Planning Guidance Transport Chapter (Pitches and Kemm 2003)	Proximal	All-cause, cardiovascular and colon carcinoma deaths per year(s), acute myocardial infarction cases per year(s) (physical activity); slight, serious & fatal injuries (traffic accidents)	Epidemiological risk assessment (physical exercise); analogy (traffic accidents)	Epidemiological evidence (physical activity); national historical data on average number of serious personal accidents per distance travelled and mode of transport	Unclear	Mainly estimates in terms like 'one death per 2 or 3 years', and general word of caution: 'estimates have very wide margin of error'

outcome assessment vary, but are generally similar for comparable determinants. In most studies, the risk measures were the result of epidemiological research, while in three studies toxicological risk measures (derived from animal experiments) were used. The time horizon is unclear in most assessments. In three of the studies in which it is clear, it is determined by the risk measure used: toxicological risk assessment assumes lifetime exposure (70 years by convention), while in one study the properties of the simulation model limit the time horizon to 50 years. Uncertainty in outcomes is seldom made explicit by more than qualitative terms like "about".

Discussion

From policy plan to determinants of health: exposure impact assessment

The methods used to estimate effects on determinants of health are quite diverse, which is not surprising considering the diversity in factors that influence health. For physical and chemical factors methods are well developed, and also for traffic flows and accident rates models are available.

As a consequence of a narrow evidence base no such models are available for many other determinants. In the cases we reviewed, estimates were commonly made on the basis of (unpublished) data or information provided by project developers. The latter source may introduce systematic bias. Author's assumptions and expert opinion are options of last resort. In the absence of standardised, validated methods and readily applicable data, some authors display substantial creativity in quantifying socioeconomic determinants. These efforts should be critically evaluated so that they may contribute to the development of a more uniform and robust approach.

From determinants to health outcomes: outcome assessment

From the 17 studies that quantify the effects of the policy decision on determinants of health, 16 proceed to give estimates of the effect on health outcomes, although only five do so for all the determinants they identified as relevant. This compares favourably with the findings of a study on the inclusion of health in environmental impact assessments, which concluded that most studies quantified up to the level of determinants (or pollutants) and compared these with limit values, thus not extending the analysis to health aspects (Alenius 2001). In the case of non-carcinogenic pollutants that do not reach the limit value, this is justified by the generally accepted assumption that for these substances there is a threshold, below which there is no health effect (Snary

2002). However, in case this threshold is exceeded an estimate of the health effects would be desirable.

Few socioeconomic and behavioural determinants are quantified up to the level of health outcomes. One of the problems may be that a stable evidence base is lacking. Unlike physical and chemical substances, socioeconomic and behavioural determinants are context dependent. For example, being unemployed in Russia is not the same as being unemployed in Germany. This means that the evidence is only to a limited extent generalisable across time and space, and that the degree of standardisation achieved in environmental HIA will be hard to match in HIA that focuses on other policy areas.

Outcome assessment is often done for different risk factors separately as in Hallenbeck and Fehr's assessments of waste facilities, and in other models such as Mindell and Joffe's instrument for predicting the health consequences of air pollution (Hallenbeck 1995; Fehr 1999; Mindell and Joffe 2004). However, the separate health effects resulting from a policy cannot always simply be added up as this may result in double counting (Mindell, Hansell et al. 2001). A possible method for the integration of different effects is used in the HIA on tobacco policy (the PREVENT model) (Mooy and Gunning-Schepers 1998) and in McCarthy's ARMADA model for environmental HIA (McCarthy, Biddulph et al. 2002). Both use simulation models that combine epidemiology and demography to assess various effects of a proposal on the health of a population. Differences by age and sex and competing risks can be accounted for. Further development of such models that can be used "off the shelf" could do much to improve quantitative HIA (Joffe and Mindell 2002).

Indicators for health outcomes

The measures of health outcome used in the different studies are quite diverse, ranging from numbers of deaths in a specified population, to hospitalisations for asthma and injury only accidents. This diversity is justified by differences in the research questions that need to be answered, but it hinders comparison of effects. It would be useful to additionally express health outcomes in a summary measure of population health such as the disability-adjusted life year (DALY) (Murray and Lopez 1997; McCarthy, Biddulph et al. 2002; Murray 2002; Kjellstrom, Van Kerkhoff et al. 2003; Murray, Ezzati et al. 2003). DALYs combine life years lost (or gained) and time spent with disease, adjusted for the severity of that particular disease, into a single indicator. The concept has been criticised (Murray and Acharya 1997; Cohen 2000) and should not replace more conventional health outcomes, but for decision making on population level it can be a useful tool. A limitation for use in HIA is that disability weights are only

available for diseases as distinguished in the international classification of diseases (ICD), so that, for example, annoyance effects due to noise or odour cannot at present be expressed in DALY. Ideally, the aggregate health impact is subsequently differentiated for (vulnerable) subgroups: health inequalities impact assessment.

Data requirements

Whatever shape it takes, quantification in HIA will be limited by the availability of relevant and reliable data. The more detailed the techniques, the higher the information requirements. For example, demographic computer simulation models can cope with differences by age and sex, but the model has to be filled with data that specify these differences. Taking into account health inequalities also increases information needs. Our review confirms Joffe and Mindell's finding that the evidence base is especially narrow when it comes to linking policy options to health determinants. The creation of databases containing evidence for both the exposure assessment and outcome assessment phases of quantitative HIA would greatly facilitate HIA practice (Joffe and Mindell 2002).

What to do when the data do not permit quantification?

Data problems commonly hamper attempts at quantification. If that is the case, robust qualitative work may be the best option. However, before concluding that quantification is not possible it may be worthwhile to bear in mind that the perspective of a decision maker differs from that of an epidemiologist. Not taking any decision is not an option for a decision maker, while the cautious epidemiologist may conclude that further study is needed. An expert's guess may still be better than no guess at all. The use of a structured process to obtain expert opinion can improve its validity and credibility. A suitable method for obtaining the collective opinion of experts is the Delphi process. Characteristics of this method are anonymity, iteration, controlled feedback, and statistical summarisation of the group responses. It has been used in "future studies" (Jones and Hunter 1995). The outcome of a Delphi study may serve as input for outcome assessment, together with the epidemiological evidence and local data.

In HIA of policies that are broadly formulated, or where there is much uncertainty over trends and future developments, the analysis of a number of scenarios might be more informative than a single estimate of the most probable impact. This permits various assumptions to be made without losing scientific credibility, and may convey to decision makers an understanding of the dynamics of the mechanics described by the model.

Validity and reliability

Little is known about the validity and reliability in HIA. For some of the methods used in HIA, validity and reliability have been assessed (Brønnum-Hansen 1999; Nieuwenhuijsen 2003), but for many methods no such research has been done. Likewise, methods to assess the validity and reliability of complete HIA studies are yet to be developed, and even agreed upon definitions suitable for HIA are lacking. We would tentatively define the validity of HIA studies as the degree to which the predicted health effects are confirmed by empirical research. This implies a need for outcome evaluation, notwithstanding the difficulties this will entail.

Standardisation

Once methods for quantitative HIA have been developed and their validity is becoming clear, the need for standardisation will arise. Similar to developments in the field of health economics (for example cost-effectiveness analysis), guidelines will be needed to determine what effects to include, what time horizons are appropriate, how to deal with uncertainty, and what are suitable indicators of health outcomes. Standardisation will increase comparability among studies and promote HIA as a reliable and credible instrument for intersectoral health policy making.

Limitations of this study

The HIA reports included in this review do not give more than an indication of what is done in the field. Reports of HIA studies are difficult to obtain as they seldom appear in peer-reviewed literature and are not always made public. In Canada for example, HIAs are performed by proponents of projects and incorporated into environmental impact assessment (EIA) reports. Several other countries are probably underrepresented in our sample for similar reasons. In contrast, in the UK there is a tradition of local level HIA separate from EIA and of making studies available via the internet. This results in the inclusion of many small scale studies with little emphasis on quantification. As we did not exclude cases on the basis of thoroughness or amount of resources invested, this partly explains the preponderance of studies from the UK in our sample, and the low proportion of HIA reports with quantified effect estimates.

Conclusion

We conclude that quantification in HIA is useful but not often achieved and that validity is often uncertain. Quantitative HIA would benefit from research and

development of (1) methods to quantify the effect of socioeconomic and behavioural determinants; (2) the development of user friendly simulation models for outcome assessment in HIA; (3) the additional use of summary measures of population health; (4) the use of expert opinion and scenario building in HIA; (5) empirical research into the validity and reliability of methods for HIA, and of complete HIA studies.

References

Abdel-Aziz, M. I., J. Redford, et al. (2000). Health Impact Assessment, Finningley Airport. Doncaster, Doncaster Health Authority**:** 1-47.

Abrahams, D. (2002). Foresight vehicle initiative comprehensive health impact assessment - Executive summary. Liverpool, IMPACT.

Abrahams, D. (2004). Foresight vehicle initiative comprehensive health impact assessment. Liverpool, IMPACT.

Alenius, K. (2001). Consideration of health aspects in environmental impact assessments for roads. Umea (Swe), National Institute of Public Health, Sweden**:** 32.

Brønnum-Hansen, H. (1999). "How good is the Prevent model for estimating the health benefits of prevention?" J Epidemiol Community Health **53**(5): 300-5.

Cohen, J. (2000). "The Global Burden of Disease Study: a useful projection of future global health?" J Public Health Med **22**(4): 518-24.

Dunt, D. R., M. J. Abramson, et al. (1995). "Assessment of the future impact on health of a proposed freeway development." Aust J Public Health **19**(4): 347-56.

Fehr, R. (1999). "Environmental health impact assessment: evaluation of a ten-step model." Epidemiology **10**(5): 618-25.

Fehr, R., O. Mekel, et al. (2003). "Towards health impact assessment of drinking-water privatization: the example of waterborne carcinogens in North Rhine-Westphalia (Germany)." Bull World Health Organ **81**(6): 408-14.

France, C. and M. E. Lilley (2000). Alconbury Health Impact Assessment report. Huntingdon, Cambridgeshire Health Authority**:** 1-52.

Hallenbeck, W. H. (1995). "Health impact of a proposed waste-to-energy facility in Illinois." Bull Environ Contam Toxicol **54**(3): 342-8.

Jobin, W. (2003). "Health and equity impacts of a large oil project in Africa." Bull World Health Organ **81**(6): 420-6.

Joffe, M. and J. Mindell (2002). "A framework for the evidence base to support Health Impact Assessment." J Epidemiol Community Health **56**(2): 132-8.

Jones, J. and D. Hunter (1995). "Consensus methods for medical and health services research." Bmj **311**(7001): 376-80.

Kemm, J. (2000). Health Impact Assessment on the Proposed Integrated Wood Processing Plant at Newbridge on Wye. Cardiff, Health Impact Assessment Unit, Welsh Combined Centres for Public Health**:** 1-43.

Kemm, J., S. Ballard, et al. (2000). Health Impact Assessment of the New Home Energy Efficiency Scheme. Cardiff, National Assembly for Wales**:** 1-27.

Kemm, J. and C. Breeze (2000). Health Impact Assessment Report - National Botanic Garden Wales. Cardiff, Health Impact Assessment Unit, Welsh Combined Centres for Public Health**:** 1-33.

Kjellstrom, T., L. Van Kerkhoff, et al. (2003). "Comparative assessment of transport risks: how it can contribute to health impact assessment of transport policies." Bull World Health Organ **81**(6): 451-7.

Kobusch, A.-B., H.-J. Serwe, et al. (1995). Gesundheitsverträglichkeitsuntersuchung zur Erweiterung der Centraldeponie Heinde - Endbericht. LÖGD. Bielefeld, LÖGD.

Künzli, N., R. Kaiser, et al. (2000). "Public-health impact of outdoor and traffic-related air pollution: a European assessment." Lancet **356**(9232): 795-801.

McCarthy, M., J. P. Biddulph, et al. (2002). "A health impact assessment model for environmental changes attributable to development projects." J Epidemiol Community Health **56**(8): 611-6.

Mindell, J., A. Hansell, et al. (2001). "What do we need for robust, quantitative health impact assessment?" J Public Health Med **23**(3): 173-8.

Mindell, J. and M. Joffe (2004). "Predicted health impacts of urban air quality management." J Epidemiol Community Health **58**(2): 103-13.

Mooy, J. M. and L. J. Gunning-Schepers (1998). Wat levert anti-roken beleid op? Berekeningen van een computersimulatie model. Tabaksontmoedigingsbeleid gezondheidseffectrapportage. NSPH. Utrecht, Netherlands School for Public Health.

Murray, C. J., Ed. (2002). Summary Measures of Population Health: concepts, ethics, measurements and applications. Geneva, World Health Organization.

Murray, C. J. and A. K. Acharya (1997). "Understanding DALYs (disability-adjusted life years)." J Health Econ **16**(6): 703-30.

Murray, C. J., M. Ezzati, et al. (2003). "Comparative quantification of health risks conceptual framework and methodological issues." Popul Health Metr **1**(1): 1.

Murray, C. J. and A. D. Lopez (1997). "Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study." Lancet **349**(9063): 1436-42.

Nieuwenhuijsen, M. J., Ed. (2003). Exposure assessment in occupational and environmental epidemiology. Oxford Medical Publications. New York, Oxford University Press.

Pitches, D. and J. Kemm (2003). Health Impact Assessment of RPG (regional planning guidance) Transport Chapter. Birmingham, University of Birmingham.

Scott-Samuel, A. (1984). "Unemployment and health (letter)." Lancet: 1464-1465.

Serwe, H.-J. and G. Protoschill-Krebs (1997). Gesundheitverträglichkeitsuntersuchung der Umgehungsstrasse B 9n / Krefeld - Ergebnisbericht. LÖGD. Bielefeld, LÖGD.

Snary, C. (2002). "Health risk assessment for planned waste incinerators: getting the right science and the science right." Risk Anal **22**(6): 1095-1105.

Taylor, A., C. Soloman, et al. (2001). Health Impact Assessment - Proposed extension to the port of Southampton at Dibden Bay, Southampton and South West Hampshire Health Authority**:** 1-180.

Taylor, L. and R. Quigley (2002). Health Impact Assessment: a review of reviews. London, NHS Health Development Agency**:** 21.

UK Dept of Health, Ed. (1999). Health Impact Assessment - Report of a methodological seminar. London.

Zwart, W. M. d. (1997). GES Drank- en horeca Wet. Utrecht, Trimbos instituut**:** 15.

PART B

Application of the potential impact fraction and multi-state life table models

3

The European Common Agricultural Policy on fruits and vegetables: exploring potential health gain from reform

Veerman JL, Barendregt JJ, Mackenbach JP. The European Common Agricultural Policy on fruits and vegetables: exploring potential health gain of reform. *European Journal of Public Health* 2006 Feb; 16 (1): 31-5

Abstract

Background

Consumption of fruits and vegetables is associated with a reduced risk of cardiovascular disease and cancer. The European Union Common Agricultural Policy keeps prices high by limiting the availability of fruits and vegetables. This policy is at odds with public health interests. We assess the potential health gain for the Dutch population of discontinuing EU withdrawal support for fruits and vegetables.

Methods

The maximum effect of the reform was estimated by assuming that a quantity equivalent to the amount of produce withdrawn in recent years would be brought onto the market. For the calculation of the effect of consumption change on health we constructed a multi-state life table model in which consumption of fruits and vegetables is linked to ischaemic heart disease, stroke, and cancer of the oesophagus, stomach, colorectum, lung and breast. Uncertainty is quantified using Monte Carlo simulation.

Results

The reform would maximally increase the average consumption of fruits and vegetables by 1.80% (95% uncertainty interval 1.12–2.73), with an ensuing increase in life expectancy of 3.8 (2.2–5.9) days for men and 2.6 (1.5–4.2) days for women. The reform is also likely to decrease socio-economic inequalities in health.

Conclusion

Ending EU withdrawal support for fruits and vegetables could result in a modest health gain for the Dutch population, though uncertainty in the estimates is high. A more comprehensive examination of the health effects of the EU agricultural policy could help to ensure health is duly considered in decision-making.

Background

Consumption of fruits and vegetables is associated with a reduced risk of cardiovascular disease and several types of cancer (Vianio and Bianchini 2003; Jansen and Van deVijver 2004). The World Health Organization recommends a daily intake of at least 400 g per person (WHO 2003).

However, many EU citizens do not reach this level of consumption, especially in northern and central European countries (Joffe and Robertson 2001). According to the 2002 World Health Report this low intake accounts for 4.3% of the burden of disease in men and 3.4% in women in the EU (WHO 2002). Fruit and vegetable consumption is disproportionately low in groups with low income and education levels and thus contributes to socio-economic inequalities in health (James, Nelson et al. 1997; Roos, Johansson et al. 2001). In the Netherlands the trend over time is towards lower consumption levels with young people consuming less than older generations (Voedingscentrum 1998).

The European Union's Common Agricultural Policy (CAP) was introduced after the Second World War to ensure food security in Europe. For fruits and vegetables it uses two mechanisms. First, it imposes import tariffs. Second, it supports producers by guaranteeing a minimum price for their produce. When the price drops below a specified intervention level, the EU finances the withdrawal of fruits and vegetables from the market ('withdrawal support'). Most of this surplus is destroyed. In other words, the EU policy keeps prices high by limiting availability. This has been termed a threat to public health which should be terminated (Schäfer Elinder 2003). Reforms in 1996 roughly halved the amount of fruits and vegetables withdrawn, but recent years have shown no further downward trend.

In this chapter, we attempt to quantify the health gain for the Dutch population that would result from ending EU withdrawal support for fruits and vegetables. We use a simple policy model that allows estimation only of a maximum effect. The true effect is likely to lie between this maximum and zero effect.

Methods

The research question can be divided into two sub-questions. First, what is the effect of the suggested policy change on the consumption of fruits and vegetables? And second, what does this consumption change mean for the health of the population? Accordingly, the methods are subdivided into those that

examine the effect of the policy on exposure to determinants of health and those that translate this altered exposure into health outcomes.

From policy to determinant

Over the period 1997–2001 the total annual availability of fruits and vegetables in the EU was estimated at 70 million tonnes, and on average 1.27 million tonnes per year was withdrawn. Over these years the amount withdrawn fluctuated without a clear trend. The maximum effect that ending withdrawal could have is equal to the amount withdrawn, which we based on these historical data. In this scenario, all produce is brought onto the market and prices would drop until demand equals supply and the market clears. Not all that is sold is consumed; we assumed that the percentage lost remains unchanged.

From determinant to health outcomes

To answer the second question and estimate the effect of changes in fruit and vegetable consumption on population health we created a mathematical model. This proportional multi-state life table model compares two populations: one that is modelled after the Dutch population and an identical population for which the fruit and vegetable intake can be manipulated (Barendregt, Van Oortmarssen et al. 1998).

Consumption data from the most recent national nutrition survey were fitted to a Weibull distribution. Assuming that the shape-parameter of the distribution over the population remains stable, the model allows manipulation of the distribution by changing the mean consumption level. Consumption levels influence the incidence of ischaemic heart disease, stroke, and cancer of the oesophagus, stomach, colorectum, lung and female breast. This effect is quantified via the potential impact fraction (PIF): the proportional change in expected incidence as a consequence of a specified change in exposure level (Kleinbaum, Kupper et al. 1982). PIF is calculated on the basis of age- and gender-specific exposure data and the relative risks of disease incidence at the corresponding levels of exposure. The relative risks used are shown in table 3.1.

A change in incidence causes changes in prevalence and mortality. Finally, the different disease-specific data are integrated in a life table from which summary measures of population health such as life expectancy, years lost to disease and disability-adjusted life years (DALYs) can be derived.

Table 3.1 Relative risks of disease incidence for fruits and vegetable consumption

	Relative risk (95% confidence interval)	Unit
Ischaemic Heart Disease	0.8 (0.65 to 0.90)*	400+ versus 0-99 g/day
Cerebrovascular Accident	0.8 (0.60 to 0.95)*	400+ versus 0-99 g/day
Oesophagus cancer	0.81 (0.72 to 0.90)	per 100 g/day increase
Stomach cancer	0.78 (0.72 to 0.84)	per 100 g/day increase
Colorectal cancer	0.93 (0.88 to 0.98)	per 100 g/day increase
Lung cancer	0.87 (0.80 to 0.93)	per 100 g/day increase
Breast cancer	0.98 (0.96 to 0.99)	per 100 g/day increase

** For ischaemic heart disease and CVA no confidence intervals were available but the values stated were interpreted as such.*

The difference in health outcomes of the two populations is attributed to the intervention, i.e. the change in consumption of fruits and vegetables. Inputs for the model were the most recent estimates of disease frequency provided in the National Public Health Compass (RIVM 2004), population data and mortality rates from the Central Bureau of Statistics, consumption data from the National Nutrition Survey 1997–1998, estimates of relative risk of coronary heart disease and stroke from a review by Jansen *et al*. (Jansen and Van deVijver 2004) and of cancer from a recent meta-analysis (Riboli and Norat 2003), and the Dutch disability weights (Stouthard, Essink-Bot et al. 1997). The model was implemented in Excel.

Uncertainty in the estimates was obtained by Monte Carlo simulation using the computer programme @risk version 4 (Palisade, London), including the uncertainty in the relative risk estimates and in the amount of extra fruits and vegetables consumed.

Results

If EU price support for fruits and vegetables were abolished and all previously withdrawn produce were sold, this would mean an average increase in consumption of 1.80% (95% uncertainty interval 1.12–2.73), or ~5–6 g per person per day on average. The resulting shift in the consumption pattern is illustrated in figure 3.1.

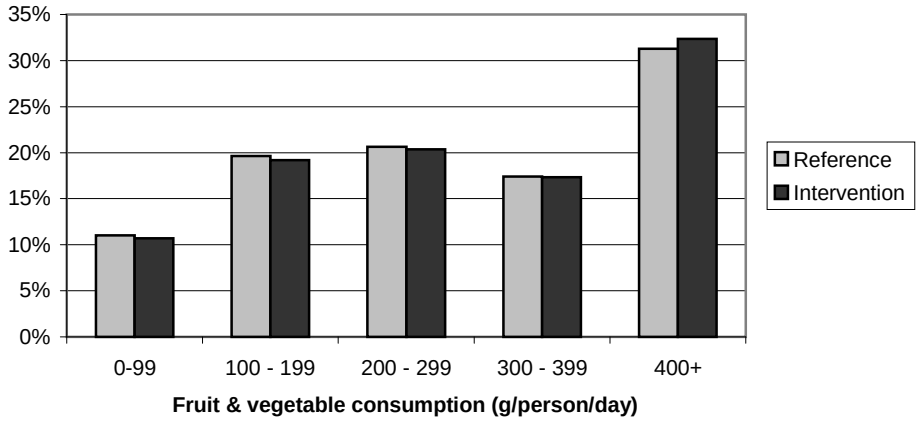


Figure 3.1 Projected consumption change among men aged 65 and over.

This change in consumption of fruits and vegetables would result in declines in the incidence of cardiovascular disease and cancer and improved population health, as shown in table 3.2.

Discussion

Ending EU subsidies for the withdrawal of fruits and vegetables and allowing the produce onto the market would result in modest health gains for the Dutch population: an estimated annual gain of 1930 DALYs, or an increase in life expectancy by 3.8 days for men and 2.6 days for women, at maximum.

This health gain may look disappointing for an intervention that affects such a large population. The reason for the modest effect lies mainly in the small shift in consumption patterns that would result from the suggested reform of EU policy. The protective effect of fruits and vegetables is also quite moderate (see table 3.1), but nevertheless the total burden of disease due to inadequate intake of fruits and vegetables is considerable: if the whole population were to consume at least 400 g per day, life expectancy for men would rise by more than half a year, and women would gain about 4 months. The reform would fill the gap between the ideal and reality to only a small degree. To put the health gain into perspective: breast cancer screening is estimated to increase life expectancy by about 35 days compared with 2.6 for the intervention evaluated here, but it is also more resource demanding (Bonneux 2003).

Table 3.2 Estimated decline in incidence and projected aggregate annual health gain in the Dutch population (just over 16 million people) that would result from ending EU withdrawal support for fruits and vegetables, with in brackets the 95% certainty interval. DALY = disability-adjusted life year; 'DA-days' = disability-adjusted days.

	Men	***Women***
Annual number of incident cases prevented per 100,000 person-years		
Ischaemic heart disease	1.5 (0.4 to 2.8)	1.0 (0.3 to 2.0)
CVA	0.5 (0.0 to 1.1)	0.5 (0.0 to 1.2)
Oesophagus cancer	0.1 (0.0 to 0.2)	0.0 (0.0 to 0.1)
Stomach cancer	0.2 (0.1 to 0.3)	0.1 (0.1 to 0.2)
Colorectal cancer	0.2 (0.1 to 0.4)	0.2 (0.1 to 0.3)
Lung cancer	0.5 (0.3 to 0.9)	0.2 (0.1 to 0.3)
Breast cancer		0.1 (0.1 to 0.3)
Aggregate annual health gain		
Deaths postponed (total population)	66 (39 to 103)	44 (25 to 71)
Life-years gained (total population)	867 (508 to 1358)	611 (357 to 963)
Life-years gained per 100,000 person-years (aged 25+)	15.8 (9.2 to 24.7)	10.6 (6.2 to 16.7)
DALYs gained (total population)	1080 (609 to 1732)	850 (467 to 1387)
DALYs gained per 100,000 person-years (aged 25+)	19.6 (11.1 to 31.5)	14.8 (8.1 to 24.1)
Increase in life expectancy (days)	3.8 (2.2 to 5.9)	2.6 (1.5 to 4.2)
Increase in disability-adjusted life expectancy ('DA-days')	4.5 (2.5 to 7.3)	3.6 (1.9 to 5.9)

Modest as they may be, these estimates represent a maximum health gain. The underlying assumption is that all produce now taken off the market would be sold as a result of the reform, which can only be an overestimation. In reality, lower prices are likely to discourage production of fruits and vegetables, which would lead to rising prices. This would lower demand until supply and demand were again in balance. If withdrawal support were replaced by other forms of support, this might affect the price level at which the market balanced. However, this longer-term effect is highly uncertain and in this study we limit ourselves to an estimate of the initial maximum effect of the policy change. Our market model is uncomplicated and ignores the role of the retail sector and other

elements in the chain between producer and consumer. However, we cannot conceive of realistic scenarios that result in higher estimates of consumption and health gain, and we would argue that the likely effect of ending withdrawal support is between zero and the results of our study.

Uncertainty

Overall uncertainty in the estimates is high. Three important sources are the relative risks of disease, the amount of the increase in availability of fruits and vegetables, and its distribution over the population. The uncertainty resulting from the first two sources is included in the uncertainty intervals presented.

First, there are wide confidence intervals around the relative risk estimates linking the consumption of fruits and vegetables to disease. Epidemiological research in this area is difficult owing to difficulty in measuring intake. There is also a paucity of good meta-analyses. This in turn is due to the way results are reported in original studies. Most often the highest consumption quantile is compared with the lowest, which is appropriate for determining whether there is an effect but not suitable for estimating its magnitude as it disregards consumption levels and spread in intake, thus hampering comparability with other populations. For cancers we were able to use a meta-analysis, but for coronary heart disease and stroke we used a review which presents a 'best guess' and 'conservative' and 'optimistic' estimates (Riboli and Norat 2003; Jansen and Van deVijver 2004). In consultation with one of the authors we tentatively interpreted the latter estimates as confidence intervals, after fitting them to a lognormal distribution. Finally, there is controversy over which type of studies are more suitable: case–control or cohort. The latter give lower estimates of effect. We used estimates for all studies combined as presented in the meta-analysis (Riboli and Norat 2003).

A second source of uncertainty is the amount of extra fruits and vegetables that would become available as a result of ending withdrawal subsidies. This depends on, among other factors, weather conditions, which vary by year.

Third, there is the question of who would consume the extra fruits and vegetables. In our model, most of the extra consumption falls to the groups that already consume relatively much. As an alternative, we created a model in which the lowest consumption group (i.e. those consuming <100 g per day) received all extra fruits and vegetables. This extreme assumption resulted in ~75% more health gain compared with the results above (table 3.3). Though the individuals concerned gain more, the effect on the total population remains modest.

Table 3.3 Estimated projected aggregate annual health gain in the Dutch population that would result from ending EU withdrawal support for fruits and vegetables under the assumption that all extra fruits and vegetables are consumed by those currently consuming less than 100 grams per day, with the 95% certainty interval in brackets. DALY = disability-adjusted life year; 'DA-days' = disability-adjusted days.

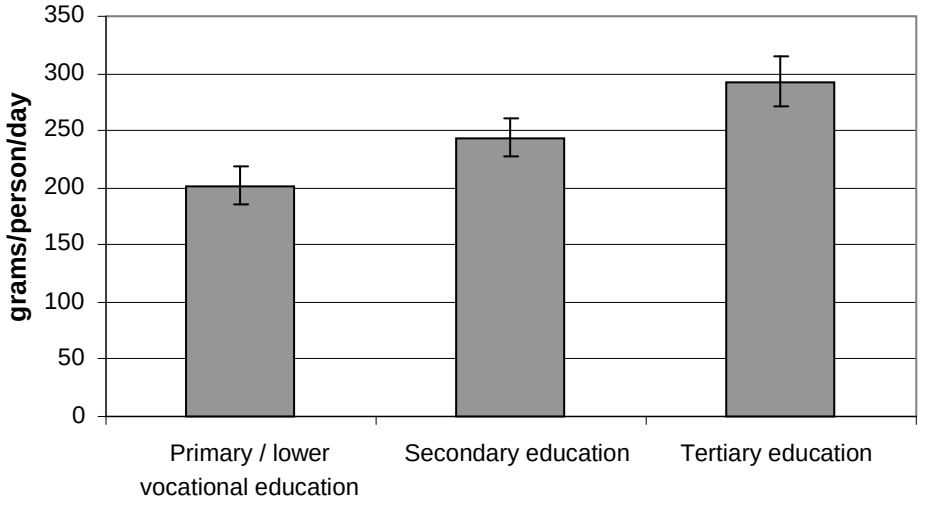
	Men	*Women*
Life-years gained per 100,000 person-years (aged 25+)	27.7 (21.0 to 35.1)	19.6 (11.3 to 31.6)
DALYs gained per 100,000 person-years (aged 25+)	33.4 (23.9 to 44.0)	26.0 (14.4 to 43.6)
Increase in life expectancy (days)	6.5 (4.9 to 8.3)	4.6 (2.7 to 7.5)
Increase in disability-adjusted life expectancy ('DA-days')	7.6 (5.3 to 10.1)	6.0 (3.2 to 10.1)

Finally, the limited number of diseases included in the model might lead to underestimation of the health effects.

Effects on socio-economic inequality

A related question is how the suggested reform would influence socio-economic inequalities in health. This depends on the difference in health gain of groups with different socio-economic status (SES), which in turn depends on three factors that may vary with SES: the current level of fruit and vegetable consumption, the corresponding burden of disease, and price-sensitivity. In the Netherlands, groups with lower SES consume less fruits and vegetables than those with higher SES (fig. 3.2), and their burden of disease owing to cardiovascular disease and cancer is higher (Rossum, Mheen et al. 1999; Bos, Kunst et al. 2000; Lenthe, Gevers et al. 2002). We found no data on differences in price-sensitivity by SES, but there are indications that price is a barrier to a healthier diet, especially in low-income groups (Ritson and Petrovici 2001). All three factors therefore indicate that the groups with lower SES stand to gain more than those with higher SES, so it seems likely that ending withdrawal subsidies for fruits and vegetables would reduce health inequalities. Quantification of this effect is possible but requires SES-specific data.

Figure 3.2 Average fruit and vegetable consumption of different socio-economic groups in the Netherlands in 1997/8 (with 95% confidence interval).



Distributing the withdrawn fruits and vegetables to charity, which is intended but rarely enforced under current policy, would also reduce health inequalities.

Immediate maximum effect

The life table model does not have a factor 'time' in the sense that it presents all effects as immediate. In reality, it takes time for an increase in consumption of fruits and vegetables to result in a decline in incidence of disease, with prevalence and mortality lagging even more. Thus, the annual health gains presented here would materialize over the years. Their magnitude is not affected by the timelessness of the model.

Policy recommendations

A recent study concluded that the EU policy of price support for fruits and vegetables has adverse health effects and should be abandoned (Schäfer Elinder 2003). We attempted to quantify the health gain that reform might yield. The modest results presented here do not imply that CAP as a whole has a small influence on health in Europe. Rather, they can be taken to suggest that in order to give health interests due weight in the decision-making process, a comprehensive effort at quantification of the health effects of CAP could and

should be undertaken. More generally, this chapter illustrates an approach to quantification that can be used in health impact assessment.

Acknowledgements

The authors would like to thank Frank Bunte of the LEI (Dutch national institute for agricultural economics) for his comments.

References

Barendregt, J. J., G. J. Van Oortmarssen, et al. (1998). "Coping with multiple morbidity in a life table." Math Popul Stud 7(1): 29-49, 109.

Bonneux, L. (2003). "Mortality reduction by breast-cancer screening." Lancet 362(9379): 245.

Bos, V., A. E. Kunst, et al. (2000). Nationale gegevens over sociaal-economische sterfteverschillen op basis van informatie over kleine geografische eenheden (National data on socio-economic mortality differences based on small area statistics). Rotterdam, Erasmus Universiteit Rotterdam, dept. PH: 54-55.

James, W. P. T., M. Nelson, et al. (1997). "Socioeconomic determinants of health: The contribution of nutrition to inequalities in health." BMJ 314(7093): 1545.

Jansen, M. C. J. F. and L. P. L. Van deVijver (2004). Fruits and vegetables in chronic disease prevention - another update. Zoetermeer, Holland Produce Promotion.

Joffe, M. and A. Robertson (2001). "The potential contribution of increased vegetable and fruit consumption to health gain in the European Union." Public Health Nutr 4(4): 893-901.

Kleinbaum, D. G., L. L. Kupper, et al. (1982). Epidemiologic research: principles and quantitative methods. Belmont, CA, Lifetime Learning Publications.

Lenthe, F. J. v., E. Gevers, et al. (2002). "Material and behavioral factors in the explanation of educational differences in incidence of acute myocardial infarction: the Globe study." Ann Epidemiol 12(8): 535-42.

Riboli, E. and T. Norat (2003). "Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk." Am J Clin Nutr 78(3 Suppl): 559S-569S.

Ritson, C. and D. Petrovici (2001). The Economics of Food Choice: Is Price Important? Food, People and Society. L. J. Frewer, E. Risvik and H. Schifferstein. Berlin, Springer Verlag: 339-363.

RIVM. (2004). "The National Public Health Compass." from http://www.rivm.nl/vtv/data/site_compass/introduction.htm.

Roos, G., L. Johansson, et al. (2001). "Disparities in vegetable and fruit consumption: European cases from the north to the south." Public Health Nutr 4(1): 35-43.

Rossum, C. T. v., H. v. d. Mheen, et al. (1999). "Socioeconomic differences in stroke among Dutch elderly women: the Rotterdam Study." Stroke 30(2): 357-62.

Schäfer Elinder, L. (2003). Public Health aspects of the EU Common Agricultural Policy. Developments and recommendations for change in four sectors: Fruit and vegetables, dairy, wine and tobacco. Stockholm, Swedish National Institute of Public Health.

Stouthard, M. E. A., M.-L. Essink-Bot, et al. (1997). Disability Weights for Diseases in The Netherlands. Rotterdam, Department of Public Health, Erasmus University Rotterdam.

Vianio, H. and F. Bianchini, Eds. (2003). IARC Handbooks of Cancer Prevention Vol. 8: Fruit and Vegetables. IARC Handbooks of Cancer Prevention. Lyon, IARC Press.

Voedingscentrum (1998). Zo eet Nederland (Results of the Dutch Food Consumption Survey). The Hague, Voedingscentrum.

WHO (2002). World Health Report 2002: Reducing Risks, Promoting Healthy Life. W. H. Organization. Geneva, World Health Organization.

WHO (2003). Diet, Nutrition and the Prevention of Chronic Diseases. Report of a joint FAO/WHO Expert Consultation. WHO Technical Report Series. WHO. Geneva, World Health Organization.

4

Using epidemiological models to estimate the health effects of diet behaviour change: the example of tailored fruit and vegetable promotion

Veerman JL, Barendregt JJ, Mackenbach JP, Brug J. Using epidemiological models to estimate the health effects of diet behaviour change: the example of tailored fruit and vegetable promotion. *Public Health Nutrition* 2006; 9 (4): 415-20.

Abstract

Background
To explore the use of epidemiological modelling for the estimation of health effects of behaviour change interventions, using the example of computer-tailored nutrition education aimed at fruit and vegetable consumption in The Netherlands.

Methods
The effects of the intervention on changes in consumption were obtained from an earlier evaluation study. The effect on health outcomes was estimated using an epidemiological multi-state life table model. Input data for the model consisted of relative risk estimates for cardiovascular disease and cancers, data on disease occurrence and mortality, and survey data on the consumption of fruits and vegetables.

Results
If the computer-tailored nutrition education reached the entire adult population and the effects were sustained, it could result in a mortality decrease of 0.4 to 0.7% and save 72 to 115 life-years per 100 000 persons aged 25 years or older. Healthy life expectancy is estimated to increase by 32.7 days for men and 25.3 days for women. The true effect is likely to lie between this theoretical maximum and zero effect, depending mostly on durability of behaviour change and reach of the intervention.

Conclusion
Epidemiological models can be used to estimate the health impact of health promotion interventions.

Introduction

Health promotion interventions are intended to improve health. Good practice requires that their effects be evaluated. Typically, health promotion interventions aim to promote health behaviours, i.e. behaviours that reduce the risk of disease, such as non-smoking, physical activity and fruit and vegetable intake. The degree to which this is achieved is assessed in impact evaluation. However, ultimately we are interested in health gain and not in promoting health behaviour change as such. Knowledge of the effects on health outcomes, such as the number of cases of disease prevented or years of life gained, makes the beneficial effects of health promotion visible and allows comparison of different interventions. This requires outcome evaluation, which is included in most health education and promotion planning models such as the Precede–Proceed Model (Carlson Gielen and McDonald 1997). However, in practice this may prove difficult. How, for example, does one determine the health effects of a programme that promotes the consumption of fruits and vegetables?

A major problem with outcome assessment is that most health effects of health behaviour change occur only after many years. Even apart from the delay, this makes measurement of health outcomes complicated and expensive, if not impossible. For example, a decrease in smoking prevalence is followed by a decrease in the incidence of chronic obstructive lung disease several years later, and prevalence and mortality lag even more as incident cases are prevalent for a number of years and then die. Public health professionals and policy-makers are not generally willing to wait that long before deciding whether an intervention is worth the investment.

As an alternative to actual measurement, estimates of health outcomes can be obtained using impact evaluations and epidemiological simulation models. These models are currently used to assess the burden of disease caused by specific risk factors and to estimate the effects of trends in risk factors, but have also been used to assess the health effects of public health interventions (Naidoo, Thorogood et al. 1997; Mooy and Gunning-Schepers 2001; Ezzati, Hoorn et al. 2003).

In this chapter we illustrate the use of an epidemiological model and estimate the potential effects on the Dutch population (totalling just over 16 million) of individual nutrition advice via computer-tailored nutrition education, a promising behaviour change intervention strategy (Brug, Oenema et al. 2003).

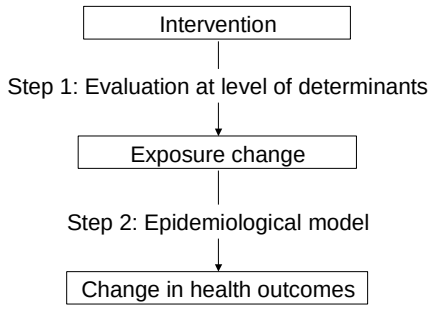
Methods

The method consists of a two-step process (Fig. 4.1). The first step is to estimate the effect of an intervention on the exposure of a population to a determinant of health (i.e. the consumption of fruits and vegetables). This is preferably based on a meta-analysis of evaluation studies of a behaviour change intervention approach. If such a meta-analysis is not available, estimates from single well-conducted studies can be used. In the second step an epidemiological model is used to estimate the change in health outcomes due to the intervention-induced change in exposure of a particular population. We shall first introduce the health promotion intervention and its impact evaluation (step 1), then briefly explain the structure and contents of the model (step 2), and discuss how we connected the two steps and what assumptions we made where data were not available.

Computer-tailored nutrition education

Individualisation of health education interventions via computer-tailoring is regarded as one of the most promising health education techniques, and is effective in inducing dietary behaviour change (Brug, Oenema et al. 2003). In computer-tailoring, an expert system is used to provide respondents with personally relevant dietary change information based on a personal assessment of dietary intake and potential mediators of change, such as intentions, attitudes and self-efficacy expectations.

Figure 4.1 Overview of methods.

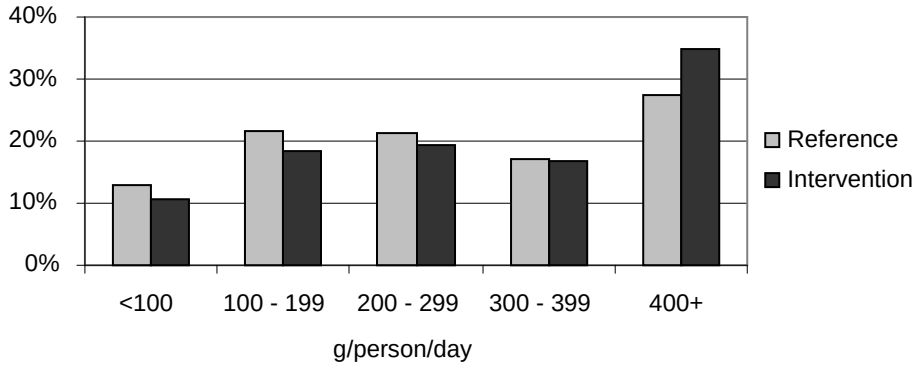


Brug *et al.* conducted a pre-test–post-test randomised trial of a computer-tailored feedback intervention on fat, fruit and vegetable intakes in Dutch adults (Brug, Glanz et al. 1998). The intervention group received computer-generated feedback letters tailored to their dietary intake, awareness of personal intake levels, intentions, attitudes and self-efficacy expectations. Participants subsequently received a second computer-tailored feedback letter tailored to the changes they made after the initial feedback. A post-test was conducted 4 weeks after this second feedback letter. The reference group received generic nutrition information. For the present study we compared pre- and post-test measurements of fruit and vegetable consumption and expressed this difference as a percentage of pre-test consumption. The intervention achieved consumption increases of 15% for fruits and 13.2% for vegetables.

The fruit and vegetable health model

The effect of a change in fruit and vegetable consumption on health outcomes was estimated using a proportional multi-state life table model (Barendregt, Van Oortmarssen et al. 1998). This model compares two populations: one that is modelled after the Dutch population and an identical population for which the fruit and vegetable intake can be manipulated. The mean consumption (in grams per day) is fitted to a Weibull distribution of which the mean is shifted upwards in the intervention population while keeping the shape parameter stable. This results in a new distribution pattern at a higher consumption level. Both curves are translated into discrete consumption categories (see Fig. 4.2).

Figure 4.2 Example of the estimated change of consumption of fruits and vegetables due to computer-tailored health promotion, assuming participation of the entire population (shown here is the effect on men aged 45-64).



Consumption levels influence the incidence of ischaemic heart disease, stroke and cancer of the oesophagus, stomach, colorectum, lung and female breast (Vianio and Bianchini 2003; Jansen and Van deVijver 2004). This effect is quantified via the potential impact fraction (PIF), which is defined as the proportional change in expected incidence as a consequence of a specified change in exposure level. PIF is calculated on the basis of age- and gender-specific exposure data and the relative risks of disease incidence at the corresponding levels of exposure (see Table 4.1). Changes in the incidence of a disease lead to corresponding changes in prevalence and mortality. Finally, the disease-specific data of the diseases in the model are integrated in a life table. From this life table, summary measures of population health such as life expectancy, years lost to disease and health-adjusted life-years (DALYs) can be derived (Murray and Acharya 1997). The difference in health outcomes of the two populations is attributed to the intervention, i.e. the change in consumption of fruits and vegetables. The data used to construct the model were the most recent estimates of disease frequency provided by the Dutch National Institute of Public Health and the Environment, population data and mortality rates from the Central Bureau of Statistics, consumption data from the National Food Consumption Survey 1997–1998, estimates of relative risk of coronary heart disease and stroke from a review and of cancer from a recent meta-analysis, and the Dutch disability weights (Stouthard, Essink-Bot et al. 1997; Riboli and Norat 2003; Jansen and Van deVijver 2004; RIVM 2004). The model was implemented in Microsoft® Excel. Ninety-five per cent uncertainty intervals (UI) that express the uncertainty in the protective effects of fruits and vegetables were calculated by Monte Carlo simulation, using the program @Risk version 4 (Palisade, London).

Assumptions

When estimating the effect of an intervention on a population, questions may arise for which an evaluation does not provide direct answers. For this study, we made the following supplementary assumptions. First, the participants in the study by Brug *et al.* were assumed to be representative of the Dutch adult population. Second, the effect of the intervention was defined as the difference between baseline measurement (T1) and 8 weeks later (T3, 4 weeks after last feedback). The control group was ignored because for this study we were interested in the effect of the intervention versus no intervention, not in the additional effect of tailoring and feedback over generic health education. Third, the results of the intervention were expressed as percentage increase in average total consumption of fruits and vegetables per person per day (as opposed to

Table 4.1 Relative risks of disease incidence for fruits and vegetable consumption.

	Relative risk (95% confidence interval)	Unit
Ischaemic Heart Disease	0.8 (0.65 to 0.90)*	400+ versus 0-99 g/day
Cerebrovascular Accident	0.8 (0.60 to 0.95)*	400+ versus 0-99 g/day
Oesophagus cancer	0.81 (0.72 to 0.90)	per 100 g/day increase
Stomach cancer	0.78 (0.72 to 0.84)	per 100 g/day increase
Colorectal cancer	0.93 (0.88 to 0.98)	per 100 g/day increase
Lung cancer	0.87 (0.80 to 0.93)	per 100 g/day increase
Breast cancer	0.98 (0.96 to 0.99)	per 100 g/day increase

* For ischaemic heart disease and CVA no confidence intervals were available but the values stated were interpreted as such.

using absolute numbers of grams, which would have been equally defensible). Fourth, the effects on fruits (+15.0%) and vegetables (+13.2%) combined are estimated to be the average of the two values (i.e. +14.1%). Finally, changes in consumption are assumed lifelong. This left us with the question what percentage of the Dutch population could be expected to participate in the programme. By way of sensitivity analysis, we calculated outcomes for participation rates of 100%, 25% and 10% of the Dutch population of age 25 years and over.

Results

The 14.1% increase in the average consumption of fruits and vegetables results in a shift in the consumption distribution of the Dutch population. As an illustration, Fig. 4.2 shows the effect on men aged 45–64 years. Assuming a participation rate of 100%, this change in consumption would result in a mortality decrease of about 786 (95% UI 587–979) deaths annually, or 0.4–0.7% in relative terms, over 10 000 life-years or DALYs (Table 4.2)). Healthy life expectancy would rise by almost 33 (95% UI 23–43) days for men and 25 (95% UI 16–34) days for women. When lower participation rates are assumed, health gain diminishes linearly (Table 4.3).

Table 4.2 Annual health gain due to computer-tailored health promotion if the total Dutch population aged 25 and over are reached, with 95% certainty intervals expressing uncertainty in relative risks in brackets. DALY = healthy life year; DALE = healthy life expectancy; DA = disability adjusted.

	Men		*Women*		*Total / average*	
Mortality decrease	472 (358 to 581)	0.7% (0.5 to 0.9)	314 (222 to 405)	0.5% (0.3 to 0.6)	-786 (587 to 979)	**0.6% (0.4 to 0.7)**
Life years gained	6217 (4750 to 7627)		4330 (3276 to 5356)		10547 (8109 to 12906)	
DALYs gained	7760 (5467 to 9997)		6030 (4043 to 7975)		13789 (9550 to 17893)	
Life-years gained per 100,000 person-years (aged 25+)	113 (86 to 139)		75 (57 to 93)		94 (72 to 115)	
DALYs gained per 100,000 person-years (aged 25+)	141 (99 to 182)		105 (70 to 138)		123 (85 to 159)	
Life expectancy at birth (days)	27.1 (20.6 to 33.3)		18.7 (13.8 to 23.3)		22.9 (17.4 to 28.1)	
DALE-expectancy at birth (days)	32.7 (22.6 to 42.6)		25.3 (16.1 to 33.9)		29.0 (19.4 to 38.1)	
Annual number of incident cases prevented						
Ischaemic heart disease	580 (193 to 968)	1.3% (0.4 to 2.1)	424 (141 to 712)	1.3% (0.4 to 2.1)	1004 (334 to 1681)	**1.3% (0.4 to 2.1)**
Stroke	195 (13 to 369)	1.3% (0.1 to 2.4)	219 (15 to 420)	1.3% (0.1 to 2.4)	414 (28 to 789)	**1.3% (0.1 to 2.4)**
Oesophagus cancer	36 (23 to 48)	4.8% (3.1 to 6.3)	15 (9 to 20)	4.8% (3.1 to 6.5)	51 (33 to 68)	**4.8% (3.1 to 6.4)**
Stomach cancer	78 (62 to 93)	10.2% (8.1 to 12.1)	44 (35 to 53)	14.2% (11.3 to 17.1)	122 (97 to 145)	**12.2% (9.7 to 14.6)**
Colorectal cancer	81 (43 to 118)	1.8% (0.9 to 2.6)	74(39 to 109)	1.8% (0.9 to 2.6)	156 (82 to 228)	**1.8% (0.9 to 2.6)**
Lung cancer	209 (132 to 280)	3.1% (2.0 to 4.2)	68 (43 to 91)	3.1% (2.0 to 4.2)	277 (175 to 371)	**3.1% (2.0 to 4.2)**
Breast cancer			58 (26 to 89)	0.6% (0.3 to 0.9)	58 (26 to 89)	**0.6% (0.3 to 0.9)**
Mortality decrease						
Ischaemic heart disease	109 (36 to 181)	1.0% (0.3 to 1.7)	76 (25 to 127)	1.0% (0.3 to 1.7)	185 (62 to 309)	**1.0% (0.3 to 1.7)**
Stroke	52 (4 to 99)	1.1% (0.1 to 2.2)	77 (5 to 146)	1.1% (0.1 to 2.2)	139 (9 to 245)	**1.1% (0.1 to 2.2)**
Oesophagus cancer	33 (21 to 44)	4.5% (2.9 to 6.0)	13 (8 to 18)	4.6% (2.9 to 6.1)	46 (30 to 62)	**4.5% (2.9 to 6.0)**
Stomach cancer	55 (44 to 65)	5.2% (4.2 to 6.2)	33 (26 to 40)	5.3% (4.2 to 6.4)	88 (70 to 105)	**5.3% (4.2 to 6.3)**
Colorectal cancer	37 (20 to 54)	1.7% (0.9 to 2.4)	35 (19 to 52)	1.6% (0.9 to 2.4)	72 (38 to 106)	**1.7% (0.9 to 2.4)**
Lung cancer	186 (118 to 250)	2.9% (1.9 to 4.0)	64 (38 to 60)	3.1% (1.8 to 4.0)	250 (156 to 331)	**2.9% (1.9 to 4.0)**
Breast cancer			**19 (9 to 30)**	**0.5% (0.2 to 0.8)**	**19 (9 to 30)**	**0.5% (0.2 to 0.8)**

Table 4.3 Annual health gain due to computer-tailored health promotion with 10%, 25% and 100% of the total Dutch population aged 25 and over participating. DALY = healthy life years; DALE = healthy life expectancy; DA = disability adjusted.

Participation rate	100%	25%	10%
mortality difference	-786	-196	-79
life years gained	10547	2637	1055
DALYs gained	13789	3447	1379
Life expectancy at birth (days)	22,9	5,7	2,3
DALE-expectancy at birth ("DA-days")	29,0	7,2	2,9

Discussion

The predicted increase in life expectancy of 3 weeks is about half that of lowering daily intake of salt by 6 g per person, but reaching that effect would probably require multiple interventions (Selmer, Kristiansen et al. 2000). The Dutch breast cancer screening programme increases female life expectancy by about 5 weeks, but is more demanding in terms of resources and emotions than a computer-tailored fruit and vegetable promotion (Bonneux 2003).

The results of this exercise are dependent on the quality of the data and methods used, as well as on a number of assumptions.

Uncertainty in the data

The results are most sensitive to two kinds of data uncertainties. First, the effect of the intervention was estimated on the basis of a single evaluation, which makes it vulnerable to bias. The use of a formal review of all similar interventions would be preferable, but at present none is available. Second, the estimates of the relative risks of disease for different levels of consumption of fruits and vegetables are subject to considerable uncertainty (Riboli and Norat 2003; Vianio and Bianchini 2003). This reflects the fact that it is not known exactly what the active components in fruits and vegetables are, the difficulty of accurately measuring fruit and vegetable intakes, and controversy over the best study designs to investigate possible preventive effects of fruit and vegetable consumption (Vianio and Bianchini 2003). Recent cohort studies generally give lower effect estimates than case–control studies; we used estimates based on reviews that included both designs (Table 4.1) (Riboli and Norat 2003; Jansen and Van deVijver 2004).

Uncertainty in the assumptions

In the analysis a number of assumptions have been made, of which especially the assumed participation rates and the sustainability of the effects influence the health outcomes. In studies on computer-tailored nutrition interventions that have been published, participation rates of up to 80% have been reported (Brug, Campbell et al. 1999). A tailored intervention targeting highly educated male employees had a participation of 74% (Brug, Steenhuis et al. 1996). These rates will be hard to match in the open Dutch population, although much will depend on the recruitment strategy. In our model, participation relates linearly to health outcomes. Because of this assumed linearity, health outcomes can be derived for any assumption regarding participation levels.

The results are also sensitive to assumptions regarding durability of the change in consumption. The effect estimate was obtained in an evaluation 4 weeks after the participants received their last feedback. The results presented are based on the assumption that they continue eating more fruits and vegetables. In one study the effects lasted for at least 8 weeks (Irvine, Ary et al. 2004). Short-term effects of short-term interventions are, however, often not sustained (De Vries and Brug 1999). Again, different assumptions on sustainability of intervention effects can be used as input in our model. If we assume, for example, that half of the participants fall back to their previous consumption levels soon after the intervention, health effects will also halve. Assessment of fruit and vegetable consumption some years after the intervention could help to determine what longer-term effect can realistically be expected.

Finally, we assumed that the population was representative of the open Dutch population. However, the study population was predominantly female and more highly educated than average, which may have led to overestimation of the effects. There is insufficient evidence to draw firm conclusions about the influence of educational level and gender on the effectiveness of tailored health education. There are indications that it is similarly effective among lower-educated women in the USA (Campbell, Tessaro et al. 2002), but a study among predominantly highly educated male employees reported only a 5% increase in fruit and vegetable intake (Brug, Steenhuis et al. 1996).

Limitations due to the structure of the model

Apart from the limitations due to the data and uncertainty in the assumptions, the time factor introduces uncertainties that our model does not fully address.

The first is that the model used in this study does not incorporate a time dimension, but instead gives causal effects as immediate. In reality there is a

time lag between change in consumption and incidence of disease, and between incidence and mortality. The health gain due to an increase in consumption in a particular year would in reality materialise over a number of subsequent years, but in the end the health gain will be equal to the size of the effect predicted by the model. If it is important to estimate when effects will occur, dynamic models are more appropriate. Dynamic models use the output of one year as input for the next, making the model more realistic but also more complex.

Second, the model uses the present occurrence of disease. As the disease pattern changes with time, so does the effect of the intervention, which means that the accuracy of predictions decreases the further we look into the future. However, patterns of disease in populations generally change slowly, so that accuracy for the first 10 or 20 years will not be biased much.

Finally, this is a macro-simulation model that ignores heterogeneity within the population, such as health differences linked to socio-economic status (SES). Because lower socio-economic groups consume less fruits and vegetables and have higher disease rates, they are likely to gain more health than those with higher SES. In general, micro-simulation models (which have individuals as basic unit) are better suited for the assessment of heterogeneity, but again at the cost of increased complexity and data requirements.

Possibilities for further analysis

A next step could be to estimate the costs of the intervention programme and perform a cost-effectiveness analysis. This kind of analysis would help decision-makers set priorities and decide whether to invest in a computer-tailored nutrition advice, a school fruit programme or e.g. a campaign to help smokers quit. For optimal comparability, similar methods should be used for all assessments.

The value of forecasting health effects

Despite the above limitations and the uncertainty inherent in any prediction, epidemiological models can be useful to estimate the effect of interventions on population health.

First, modelling makes the health effects of alternative interventions comparable. This can inform decisions on the allocation of resources for health and aid in setting priorities. Whereas the results of impact evaluations are often difficult to compare, with the help of models different interventions can be translated into forecasts of total health gain in generic terms (e.g. life-years or DALYs) (Murray, Ezzati et al. 2003).

Second, epidemiological models provide a logical framework in which evidence can be summarised and assessed for its consequences. This is useful even without actually running the model. As shown above, use of the model raises questions that need to be answered in order to estimate health effects of interventions at the population level. How strong is the protective effect of fruits and vegetables? What proportion of the target group will participate? The model splits up the general question 'what is the health effect' into sub-questions. For each of these, evidence needs to be sought and assessed for validity.

This leads to the third advantage of the use of models: it shows where the gaps in the evidence are, and where future research can contribute to reducing uncertainty in effect estimates for interventions.

Fourth, this stepwise approach makes the assessment process transparent. Each of the steps in the thought process can be judged for validity. Alternative assumptions can also be assessed for their consequences.

A fifth advantage is that even when there is little evidence available, at least the possible effects of an intervention can be explored. An uncertainty analysis can be done to establish likely boundaries of potential health effects, and a sensitivity analysis reveals how sensitive the results are to changes in input parameters. In our analysis for example, we included estimates of the health effect assuming a participation rate of 100%. This shows that the health effect of computer-tailored health education is unlikely to yield more than 10,000 life-years annually.

In conclusion, health promotion needs to move beyond process evaluation and start assessing its effects on population health. In this venture, epidemiological models can be used to estimate the health effects of health promotion interventions, building upon impact evaluation of these interventions. Since no prediction can be better than the information that it is based upon, thorough evaluation of interventions is crucial. These should preferably be conducted a considerable time after the intervention in order to estimate the durability of the behaviour changes.

Acknowledgement

We thank Elling Bere (Institute for Nutrition Research, Oslo) for his comments on an earlier version of this article.

References

Barendregt, J. J., G. J. Van Oortmarssen, et al. (1998). "Coping with multiple morbidity in a life table." Math Popul Stud **7**(1): 29-49, 109.

Bonneux, L. (2003). "Mortality reduction by breast-cancer screening." Lancet **362**(9379): 245.

Brug, J., M. Campbell, et al. (1999). "The application and impact of computer-generated personalized nutrition education: a review of the literature." Patient Educ Couns **36**(2): 145-56.

Brug, J., K. Glanz, et al. (1998). "The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake." Health Educ Behav **25**(4): 517-31.

Brug, J., A. Oenema, et al. (2003). "Past, present, and future of computer-tailored nutrition education." Am J Clin Nutr **77**(4 Suppl): 1028S-1034S.

Brug, J., I. Steenhuis, et al. (1996). "The impact of a computer-tailored nutrition intervention." Prev Med **25**(3): 236-42.

Campbell, M. K., I. Tessaro, et al. (2002). "Effects of a tailored health promotion program for female blue-collar workers: health works for women." Prev Med **34**(3): 313-23.

Carlson Gielen, A. and E. M. McDonald (1997). The precede-proceed planning model. Health behaviour and health education. K. Glanz, F. M. Lewis and B. K. Rimer, Jossey-Bass**:** 359-383.

De Vries, H. and J. Brug (1999). "Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach." Patient Educ Couns **36**(2): 99-105.

Ezzati, M., S. V. Hoorn, et al. (2003). "Estimates of global and regional potential health gains from reducing multiple major risk factors." Lancet **362**(9380): 271-80.

Irvine, A. B., D. V. Ary, et al. (2004). "The effectiveness of an interactive multimedia program to influence eating habits." Health Educ Res **19**(3): 290-305.

Jansen, M. C. J. F. and L. P. L. Van deVijver (2004). Fruits and vegetables in chronic disease prevention - another update. Zoetermeer, Holland Produce Promotion.

Mooy, J. M. and L. J. Gunning-Schepers (2001). "Computer-assisted health impact assessment for intersectoral health policy." Health Policy **57**(3): 169-77.

Murray, C. J. and A. K. Acharya (1997). "Understanding DALYs (disability-adjusted life years)." J Health Econ **16**(6): 703-30.

Murray, C. J., M. Ezzati, et al. (2003). "Comparative quantification of health risks conceptual framework and methodological issues." Popul Health Metr **1**(1): 1.

Naidoo, B., M. Thorogood, et al. (1997). "Modelling the effects of increased physical activity on coronary heart disease in England and Wales." J Epidemiol Community Health **51**(2): 144-50.

Riboli, E. and T. Norat (2003). "Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk." Am J Clin Nutr **78**(3 Suppl): 559S-569S.

RIVM. (2004). "The National Public Health Compass." from Http://www.rivm.nl/vtv/data/site_compass/introduction.htm (last accessed 1 October 2004).

Selmer, R. M., I. S. Kristiansen, et al. (2000). "Cost and health consequences of reducing the population intake of salt." J Epidemiol Community Health **54**(9): 697-702.

Stouthard, M. E. A., M.-L. Essink-Bot, et al. (1997). Disability Weights for Diseases in The Netherlands. Rotterdam, Department of Public Health, Erasmus University Rotterdam.

Vianio, H. and F. Bianchini, Eds. (2003). IARC Handbooks of Cancer Prevention Vol. 8: Fruit and Vegetables. IARC Handbooks of Cancer Prevention. Lyon, IARC Press.

5

Stemming the obesity epidemic: a tantalizing prospect

Veerman JL, Barendregt JJ, Van Beeck EF, Seidell JC, Mackenbach JP. Stemming the obesity epidemic: a tantalizing prospect. *Obesity, in press.*

Summary

Background
Obesity is a growing problem worldwide but there are no good methods to assess the future course of the epidemic and the potential influence of interventions. We explore the behaviour change needed to stop the obesity epidemic in the US.

Methods
We modelled the population distribution of body mass index (BMI) as a lognormal curve of which the mean shifts upward with time due to a positive population energy balance. Interventions that decrease food intake or increase physical activity result in more favourable trends in BMI.

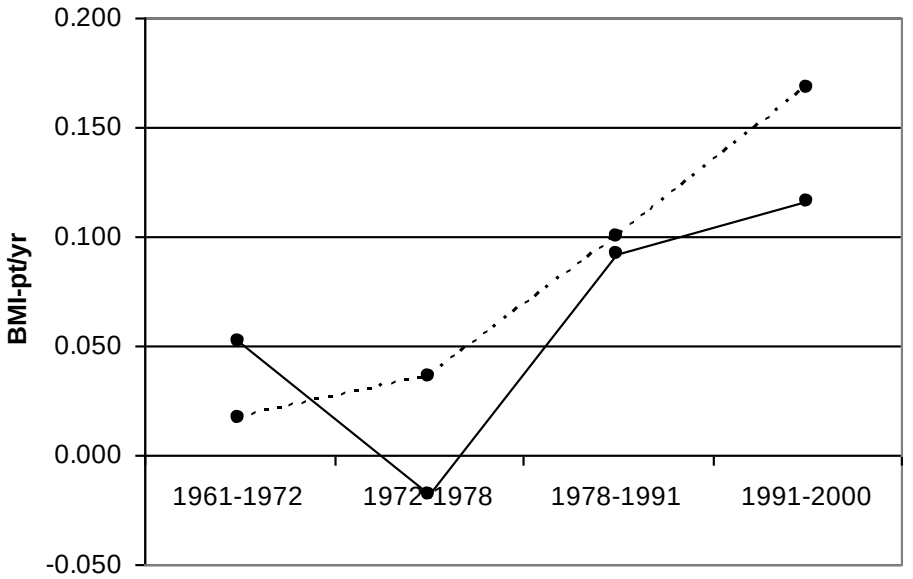
Results
The recently observed trend in average BMI implies that the average US adult overconsumes by about 10 kcal/day. If this trend continues unaltered, obesity prevalence will exceed 40% for men and 45% for women in 2015. To stop the epidemic, it suffices to decrease caloric consumption by about 10 kcal or walk an extra 2 to 3 minutes per day, on average.

Conclusion
This leads to a paradox: little behaviour change seems sufficient to halt the epidemic but in practice this proves hard to achieve. The obesogenic environment is the likely culprit. Individuals trying to maintain a healthy weight need to be supported by environments that stimulate physical activity and do not encourage overconsumption. Research should show what measures are effective.

Background

Obesity is a growing health problem worldwide. In the US the prevalence of overweight (defined as a body mass index of 25 to 30 kg/m^2) is about 35% while another 30% of the adult population is obese (BMI >30 kg/m^2) (Hedley, Ogden et al. 2004). As yet, there is no sign of an end of the epidemic. On the contrary, the rate of increase in body weight still seems in the accelerating phase (fig. 5.1), with adolescents and young adults having especially high rates. In a few decades this generation may experience obesity-related mortality to the extent of lowering life expectancy (Olshansky, Passaro et al. 2005). To make matters worse, there are indications that the relative risk of disease may increase with the duration of exposure (Peeters, Barendregt et al. 2003), though the uncertainty concerning the magnitude of the problem remains considerable (Mark 2005).

Figure 5.1 US trend in mean BMI, expressed as BMI-point increase per year, for men (solid line) and women (dashed line) aged 20 to 74 (Ogden, Fryar et al. 2004).



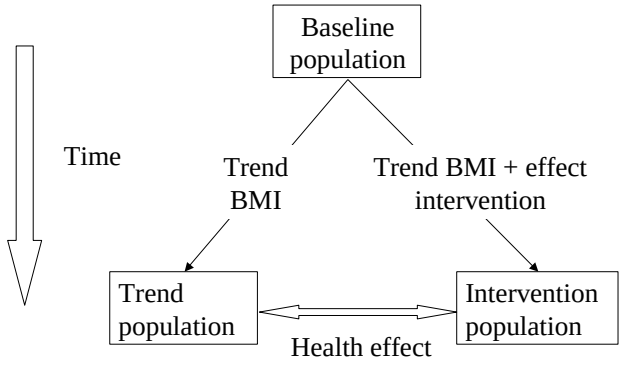
Obesity results from a mismatch between energy input (caloric intake) and output (of which physical activity is the main factor that can be influenced), or a mismatch between our biology that is geared towards the creation of energy stores for meagre times and our current environment in which food is abundant and in which physical activity can easily be avoided. Most experts agree that efforts to stem the epidemic must focus on the 'obesogenic' environment (Hill, Wyatt et al. 2003). In the US, the BMI distribution has been shown to shift to higher levels across the population, whereby the lower end changes little but the upper tail is increasingly skewed (Flegal and Troiano 2000). Though some subgroups are more affected than others, this persisting lognormality of the BMI distribution supports the notion that the obesogenic environment affects the whole US population.

For the planning of interventions and to estimate the need for health services, information about the future course of the obesity epidemic is required. There are methods to estimate the future prevalence of overweight and obesity (Hill, Wyatt et al. 2003; Arterburn, Crane et al. 2004; James, Jackson-Leach et al. 2004) but these offer no convenient framework for the assessment of the effect of interventions. We present a framework for effectiveness modelling that employs a population perspective, and present an example of a hypothetical US-wide intervention that results in an average of two extra minutes of walking per day. We also estimate the magnitude of the behaviour change needed to prevent further increases in the prevalence of obesity.

Methods

Underlying the prevalence of overweight and obesity is the population distribution of body mass index, of which the mean predicts the number of deviant individuals (Rose 1991). It therefore makes sense to relate changes in the prevalence of obesity and overweight to the mean BMI of a population (James, Jackson-Leach et al. 2004). We constructed a model that mathematically describes a population in terms of BMI. From this baseline population two secondary populations are derived which represent the same population after a specified number of years (fig. 5.2). The BMI-distribution of these populations can be changed by manipulation of mean BMI. With an increase in mean BMI, the variance also increases such that the lower end of the BMI distribution remains fixed for each age and sex group (Flegal and Troiano 2000). Mean BMI in turn is dependent on a population energy balance. Recent observed trends in BMI reflect the degree of imbalance.

Figure 5.2 Logical structure of the model.



The first of the secondary populations is exposed to this trend ("trend population") and acts as a reference scenario. In the second derived population, the observed trend is modified by interventions that affect caloric intake or physical activity via the energy balance ("intervention population"). We implemented the model in a spreadsheet.

Data

The baseline population was modelled after the US population in the year 2000. NHANES-C data (1999-2002) on the measured prevalence of overweight, obesity and extreme obesity in seven age groups (20-29, 30-39, … 80+) were fitted to a lognormal distribution using the least squares method. The recent trend in mean BMI (in BMI-points/year) for adults was calculated from NHANES-III and NHANES-C data (Ogden, Fryar et al. 2004). From the fitted mean BMI and data on the age-specific average height (Ogden, Fryar et al. 2004), we calculated the age-specific average annual increase in weight (i.e., taking a period perspective). Metabolic calculations of the American College of Sports Medicine, which assume a relationship of about 3500 kcal per pound of body weight (7700 kcal/kg), permitted to translate the trend into degrees of energy misbalance in kcal/year (American College of Sports Medicine 1995). We assumed that 10% of energy intake is spent in the digestive process and that it takes a further 10% to store energy as body weight (Westerterp, Donkers et al. 1995). Population numbers from the US Census 2000 were used to compute age-weighted prevalence of overweight and obesity.

Intervention

We constructed two scenarios for developments until the year 2015. In the baseline scenario, the recent trend in mean BMI remains unchanged. This scenario is compared to a second scenario in which a hypothetical policy is implemented that results in an increase in the average daily amount of physical activity by 2 minutes of walking (at 3.1 ml/h or 5 km/h) per day or an equivalent effort (depending on body weight this represents an average increase in energy expenditure of about 6 kcal). All else remains the same, e.g. people do not eat more to compensate for an increase in physical activity. This intervention is equivalent to a consumption decrease of about 8 kcal on average.

Uncertainty analysis

Bootstrapping is used to assess the cumulative uncertainty in the estimates that results from the BMI mean (Ogden, Fryar et al. 2004), the trend in average BMI (+/- 0.02 BMI-points/yr), and the measure linking caloric misbalance to body mass (+/- 10%). The results are expressed as 95% uncertainty intervals.

Results

The recent US trend in average BMI is about 0.116 BMI-point/year for men and 0.168 for women (fig. 5.1). Depending on age, this implies that the energy balance is positive by 7.2 to 7.7 kcal/day for the US male population and 8.8 to 9.5 kcal for women. In the baseline scenario this trend results in a prevalence of obesity of 40% in men and 47% in women in 2015, an increase of about 15% compared to the year 2000 (table 5.1).

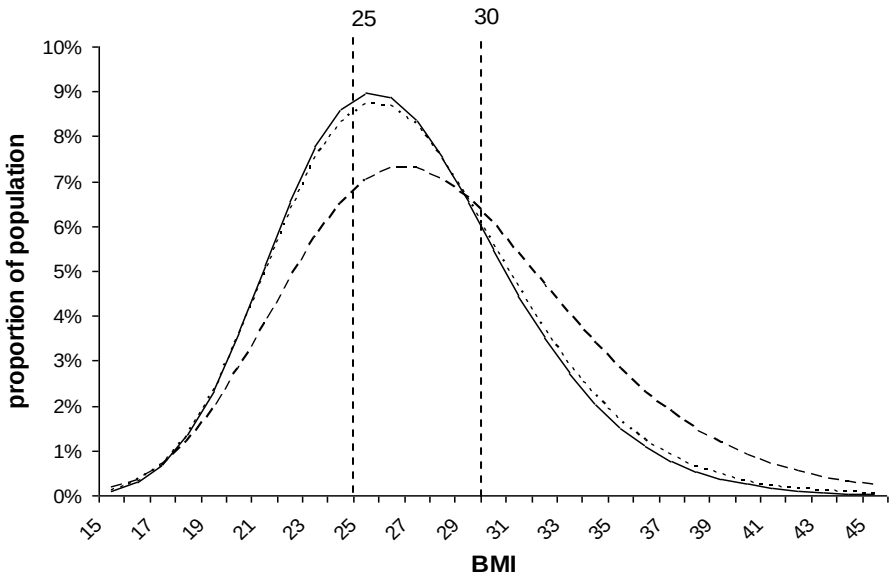
Table 5.1 Predicted US obesity prevalence in 2015. Modelled prevalence of overweight and obesity in the US population in the year 2000, with estimates for 2015 assuming the recently observed trend in average BMI continues and two alternative scenarios that result in lower trends, with 95% uncertainty intervals in brackets.

	2000	2015 – trend	2015 – 2 min. extra physical activity / day
Men			
% overweight	41.9 (41.2 – 42.4)	35.9 (34.8 – 37.0)	41.5 (40.3 – 42.6)
% obese	25.0 (23.5 – 26.5)	39.0 (36.2 – 41.6)	26.3 (22.7 – 29.3)
Women			
% overweight	28.0 (27.6 – 28.3)	23.9 (23.4 – 24.4)	26.5 (25.9 – 27.2)
% obese	32.2 (31.1 – 33.3)	44.5 (43.0 – 46.2)	37.1 (35.1 – 39.1)

Overall, it is the prevalence of obesity that is expected to rise while the prevalence of overweight tends to decrease with a higher expected upward trend in average BMI. Uncertainty in the estimates is higher for the prevalence of obesity than for the prevalence of overweight.

In the scenario in which the total population walks an average 2 minutes per day extra, adult obesity prevalence in 2015 rises to 26.3% for men and 37.1% for women. This is respectively 12.7 and 7.4 percentage points lower than otherwise expected if the current trend continues unabated (table 5.1). For men, this nearly amounts to a stabilization in the prevalence of obesity (fig. 5.3).

Figure 5.3 Modelled BMI distribution of US men aged 30-39. The continuous line depicts the situation in the year 2000, while the dashed line indicates the distribution in 2015 if the current trend in mean BMI continues. The dotted line in between shows the distribution in 2015 assuming this trend is mitigated by an intervention, in this case a 2-minute increase in average daily walking (at 3.1 mph). Intermittent vertical lines show the cut-off points between normal weight, overweight and obesity.



To stop the increase in obesity prevalence in the US, energy expenditure should increase, energy intake should decrease, or a combination of both. Physical activity has to increase by, on average, (an equivalent of) 2.2 (95% uncertainty interval (UI): 1.8 – 2.6) minutes of walking at 3.1 mph per day for men and 3.2 (2.7- 3.6) minutes for women, under the condition that this is not compensated for by extra energy intake. Conversely, assuming energy expenditure constant and taking into account the energy cost of digestion and conversion to body mass, energy intake has to decrease by 9.1 (7.3 – 10.9) kcal per day for men and 11.2 (9.6 – 12.9) for women – less than a can of soft drink per week.

Discussion

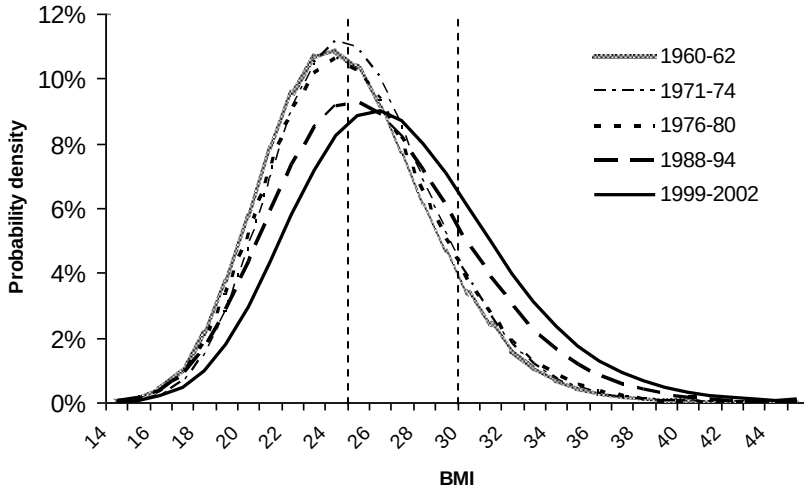
The assumptions of a lognormal distribution of BMI and an energy balance at the population level provide a framework for predictions of the prevalence of overweight and obesity. Applying this framework to the US adult population we estimate that the average American overconsumes by about 10 kcal per day. If current trends continue unchecked, in 2015 obesity prevalence could exceed 40% among US men and 45% among women. Modest increases in physical activity or decreases in caloric intake can mitigate this scenario or even reverse the upward trend in obesity prevalence.

Limitations

The two crucial assumptions in this model are the lognormality of the BMI distribution and the population-level energy balance. The lognormal distribution seems a reasonable approximation of the data, though it underestimates the prevalence of extreme obesity. Since we only look at three broad categories of BMI (<25, 25-29.9, 30+) this is unlikely to influence the results much. If anything, the method tends to underestimate the prevalence of obesity. This is also why the estimates of obesity prevalence in the year 2000 reported in table 5.1 are below CDC's estimates (Hedley, Ogden et al. 2004). We modelled the lower end of the BMI-distribution such that it remains fixed over time, which is in agreement with observations in the original data for most sex and age groups (fig. 5.4) (Flegal and Troiano 2000).

The use of a population-level energy balance is supported by the notion that environmental factors shift the BMI-distribution of entire populations. For individuals the validity of the energy balance and the assumption that it takes an overconsumption relative to energy expenditure of about 3500 kcal to gain a pound of bodyweight is generally accepted. The validity of this measure for use

Figure 5.4 Lognormally-fitted BMI curves for the US male population based on successive cross-sectional surveys (Flegal, Carroll et al. 1998; Hedley, Ogden et al. 2004). After about 1980 the curve shifts to the right with increased skew while the lower end remained 'fixed'.



on population mean BMI seems plausible, but is difficult if not impossible to support empirically. The uncertainty in the measurement of caloric intake and energy expenditure in the open population does not allow reliable estimation of the population energy balance, considering that we are looking for a mismatch in the order of a few kcal per day (Westerterp and Goris 2002). In the calculation of the uncertainty intervals we tentatively estimated the uncertainty in this parameter at +/- 10% (i.e. SD = 5%).

A further limitation of the model is that the age groups are modelled taking a period perspective and independently, whereas in reality individuals age and move to a higher category, taking their accumulated body mass with them. However, since the trend is similar for all age groups, it is not possible to distinguish cohort and period effects, and therefore this will not affect the results (Clayton and Schifflers 1987; Flegal, Carroll et al. 1998).

We use BMI as measure of obesity, whereas it seems visceral adiposity that confers most of the health risk. For an individual BMI is a rather poor indicator of that quantity, especially at young and old age (Deurenberg, Weststrate et al. 1991). However, for monitoring populations BMI is still useful. Firstly because it is reported widely and because it is easier to measure and less

error-prone than e.g. the waist circumference. Secondly, it is more valid for following populations through time than for comparing individuals. While muscular individuals with a relatively high BMI bias the results at the individual level when it comes to assessing health risks, at the population level we compare 50-year old men in 2000 with 50-year old men in 2015. In the absence of a clear trend in physical exercise at the population level, it seems reasonable to assume that most of the differences in body mass between these groups of 50-year olds would be due to a difference in fat mass, of which a proportion will be abdominal visceral fat.

It should be noted that a higher body mass requires a higher energy consumption in the order of 35 to 55 kcal per pound (15-25 kcal/kg), so that a surplus energy intake of 10 kcal per day results in an extra consumption of about 117 to 190 kcal at the end of a 15 year period, used mostly to maintain an extra body weight of about 15.5 pounds (7 kg) (Seidell and Rissanen 2004).

Lastly, our estimate of the average amount of extra physical activity needed may be a bit optimistic since activity may increase appetite and consumption, so that it may take more than a few minutes of extra activity to prevent weight gain.

Other predictions

Published models that predict the future course of the obesity epidemic are few. Arterburn et al. examined the obesity prevalence by birth cohort and predict a rise in the prevalence of obesity in elderly Americans from 32.0% in 2000 to 39.6% in 2010 if the current trends continue (Arterburn, Crane et al. 2004). Our model predicts a slightly higher prevalence of 40.6% (95% UI: 38.8 – 42.6; results not shown), which can be explained by the shape of the BMI distribution. Figure 5.3 shows that the increase in obesity prevalence with increasing mean BMI is not linear; a small increase in mean BMI pushes a disproportionate percentage of the population over the 30 kg/m^2 threshold. A second prediction-method is the formula used in the WHO Comparative Risk Assessment study (James, Jackson-Leach et al. 2004). This formula is not applicable to the future US population because it considers the proportion of overweight and obesity separately, without taking account of the underlying BMI-distribution. As mean BMI approaches or exceeds 30, this results in overestimation of the percentage of overweight individuals. In contrast, our model predicts the prevalence of overweight to decrease with increasing mean BMI when mean BMI exceeds 30. Figure 5.3 shows that as average BMI goes up, more overweight people shift over the upper boundary of 30 kg/m^2 to become obese than there are normal weight people crossing the lower boundary of 25 kg/m^2 to become overweight.

How much change is needed?

Our estimate of the magnitude of the behaviour change that the US population needs to make to stop the obesity epidemic is more optimistic than that of Hill et al (Hill, Wyatt et al. 2003). They estimate that affecting the energy balance by 100 kcal per day would suffice to stop weight gain in 90% of the population. Since they assume that only 50% of the excess caloric intake is accumulated as body mass where others assume that 90% is stored (Leibel, Rosenbaum et al. 1995) this is likely to be an overestimate of the behavioural change that the average US citizen needs to make. We assumed that digestion and conversion to body fat each cost 10% of the caloric intake. Since all relations in our model are linear, if 60% instead of 20% of excess energy intake is needed to store the excess intake, then the net excess energy that is stored is halved. Consequently the minutes walked to stop the increase should be doubled. And similarly with the cans of soft drink. Furthermore, the goal of Hill and colleagues is to stop weight gain in each individual whereas our approach aims to stabilize the population mean BMI. For example, we compare men aged 50 in the year 2000 with men aged 50 in 2015 while the comparable cohort of Hill and colleagues would be 65 years old. Since the average BMI increases up to about age 60 in the US population, our approach allows individuals some weight gain up to that age to keep the population mean BMI stable. In order to achieve weight loss, sustained behavioural changes larger than those cited here are required.

Paradox

This leads to an interesting paradox: little behaviour change seems sufficient to halt the epidemic but in practice this proves hard to achieve. Part of this paradox may be explained by the fact that the model works with averages; some individuals would have to change considerably more to avoid further weight gain, and this is compounded by the fact that the body resists weight loss more than it resists weight gain (World Health Organization 1998). Current trends are towards less physical activity and higher consumption levels, so achieving the small behaviour changes needed to stop the obesity epidemic requires a reversal of these trends. Underneath these small changes powerful environmental determinants make people eat more and/or move less than would be healthful. The difficulty individuals have in changing their behaviour on a permanent basis (Asp, Bjorntorp et al. 2002) points to the need for changes in our environment that encourage physical activity and decrease the stimulus for overconsumption. Such population targeted interventions have the additional benefit of not stigmatizing overweight individuals. Examples of such interventions are the construction of cycling lanes and safe walking routes, increasing fuel taxation to

discourage the use of cars, and regulating food advertising aimed at children (Poston and Foreyt 1999). The difficulty is that many potentially effective interventions demand sacrifices while their effectiveness is unclear. Little research effort has been made to assess the effects of population targeted interventions. The research record mainly shows that targeting individuals has little impact on the obesity epidemic; we now need research that shows how effective population interventions are (Jain 2005).

Clinical relevance

For an individual patient with overweight it is of little use to know that small changes in population averages could stop the obesity epidemic from advancing further, and a clinician's first responsibility is to provide optimal treatment and counselling to his or her patients. The medical profession, however, has a broader responsibility. Physicians should realize that the obesity epidemic will not be cured in the consultation room, and press for measures that make the living and working conditions of their patients less obesogenic.

Conclusion

The recent increase in the prevalence of overweight and obesity is expected to result in considerable morbidity and mortality in the future. Public health action is necessary to stem the current upward trend in body weight. The framework described in this chapter can aid planning by creating scenarios of future developments in obesity prevalence. The difference between the current upward trend in obesity and a downward trend is estimated to be about 10 calories or a 3 minute walk per day on average, but achieving this change proves a tantalizing challenge. Empowering individuals met with limited success so far and needs to be supplemented with changes in the environment that stimulate physical activity and a healthy diet. Increased research into the effectiveness of population-targeted interventions should guide this process of societal change.

References

American College of Sports Medicine (1995). ACSM's Guidelines for Exercise Testing and Prescription. Baltimore, Williams & Wilkins.

Arterburn, D. E., P. K. Crane, et al. (2004). "The coming epidemic of obesity in elderly Americans." J Am Geriatr Soc 52(11): 1907-12.

Asp, N. G., P. Bjorntorp, et al. (2002). Obesity - problems and interventions. Stockholm, Swedish Council on Technology Assessment in Health Care (SBU).

Clayton, D. and E. Schifflers (1987). "Models for temporal variation in cancer rates. II: Age-period-cohort models." Stat Med 6(4): 469-81.

Deurenberg, P., J. A. Weststrate, et al. (1991). "Body mass index as a measure of body fatness: age- and sex-specific prediction formulas." Br J Nutr 65(2): 105-14.

Flegal, K. M., M. D. Carroll, et al. (1998). "Overweight and obesity in the United States: prevalence and trends, 1960-1994." Int J Obes Relat Metab Disord 22(1): 39-47.

Flegal, K. M. and R. P. Troiano (2000). "Changes in the distribution of body mass index of adults and children in the US population." Int J Obes Relat Metab Disord 24(7): 807-18.

Hedley, A. A., C. L. Ogden, et al. (2004). "Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002." Jama 291(23): 2847-50.

Hill, J. O., H. R. Wyatt, et al. (2003). "Obesity and the environment: where do we go from here?" Science 299(5608): 853-5.

Jain, A. (2005). "Treating obesity in individuals and populations." Bmj 331(7529): 1387-90.

James, W. P. T., R. Jackson-Leach, et al. (2004). Overweight and obesity. Comparative quantification of health risks. M. Ezzati, A. D. Lopez, A. Rodgers and C. J. L. Murray. Geneva, WHO: 497-596.

Leibel, R. L., M. Rosenbaum, et al. (1995). "Changes in energy expenditure resulting from altered body weight." N Engl J Med 332(10): 621-8.

Mark, D. H. (2005). "Deaths attributable to obesity." Jama 293(15): 1918-9.

Ogden, C. L., C. D. Fryar, et al. (2004). "Mean body weight, height, and body mass index, United States 1960-2002." Adv Data(347): 1-17.

Olshansky, S. J., D. J. Passaro, et al. (2005). "A potential decline in life expectancy in the United States in the 21st century." N Engl J Med 352(11): 1138-45.

Peeters, A., J. J. Barendregt, et al. (2003). "Obesity in adulthood and its consequences for life expectancy: a life-table analysis." Ann Intern Med 138(1): 24-32.

Poston, W. S., 2nd and J. P. Foreyt (1999). "Obesity is an environmental issue." Atherosclerosis 146(2): 201-9.

Rose, G. (1991). "Population distributions of risk and disease." Nutr Metab Cardiovasc Dis 1: 37-40.

Seidell, J. C. and A. M. Rissanen (2004). Prevalence of Obesity in Adults: The Global Epidemic. Handbook of Obesity. G. A. Bray and C. Bouchard. New York, Marcel Dekker, Inc.

Westerterp, K. R., J. H. Donkers, et al. (1995). "Energy intake, physical activity and body weight: a simulation model." Br J Nutr 73(3): 337-47.

Westerterp, K. R. and A. H. Goris (2002). "Validity of the assessment of dietary intake: problems of misreporting." Curr Opin Clin Nutr Metab Care 5(5): 489-93.

World Health Organization (1998). Obesity: Preventing and Managing the Epidemic. Geneva, WHO.

6

Childhood obesity: the potential effectiveness of limiting food advertising on television

Veerman JL, Van Beeck EF, Barendregt JJ, Mackenbach JP. Childhood obesity: the potential effectiveness of limiting food advertising on television. *Manuscript*.

Abstract

Background

Childhood obesity rates are rising. Once these children age this increases their risk of diabetes, CVD, cancer and arthritis and lowers their life expectancy. Recent reviews support the notion that food advertising contributes to the rise in childhood obesity, but the size of this contribution is unclear. We attempt to quantify the potential effects of measures that limit food advertising on television on the prevalence of overweight and obesity among 6 to 12-year old Dutch children.

Methods

Based on available evidence we created a mathematical simulation model that links the average exposure to food commercials, via caloric consumption, to body weight and the prevalence of overweight and obesity. In a second analysis we used a Delphi study to estimate the effect of television advertising on total caloric intake.

Results

We found that an effective ban on food advertising would reduce the prevalence of overweight and obesity among Dutch children by about 1.5 percentage-point.

Discussion

Even this modest effect might not materialise when restrictions apply only to advertising on television and the same marketing budgets are used to reach children via other channels. Interventions that aim to limit obesogenic food advertising therefore should not be confined to televised commercials but also target other marketing channels, and form part of a broader effort to improve children's diets and physical activity patterns. Like the struggle against smoking, lowering obesity rates will require many small steps and the gradual building of a societal consensus to take more drastic measures.

Introduction

Childhood obesity rates are rising worldwide. This is alarming because once gained, it is very difficult to get rid of extra weight, and obesity is associated with an increased risk of CVD, cancer and with a lower life expectancy (Peeters, Barendregt et al. 2003; James, Jackson-Leach et al. 2004).

There are indications that food advertising causes weight gain in children (Hastings, Stead et al. 2003; IOM 2005; Lobstein and Dibb 2005). Advertising influences food preferences and purchasing behaviour, while there is limited evidence for an effect on total consumption patterns and body weight (Hastings, Stead et al. 2003). How much of the overweight and obesity in children is attributable to food advertising is unknown, and there is no accepted method to estimate the potential effect of measures that limit food advertising targeting children. We undertook to estimate the potential effects on overweight and obesity prevalence among 6 to 12-year old Dutch children of restrictions on food advertising on TV. Based on published literature and a Delphi study, we created a mathematical simulation model.

In this chapter we will first explore a scenario in which exposure of children to food advertising on television is reduced to zero and all else remains the same. The results are compared to the situation in 1980 when the rise in obesity prevalence had not yet started. However, in reality all else may not remain the same. Even if it is possible to reduce exposure of children to food advertising on TV to zero, marketing budgets that cannot be spent on television advertising might well be spent on other ways to influence children and their parents, such as in-school marketing, sponsorship, product placement, internet marketing and sales promotions (Hawkes 2004). We therefore also asked the experts how much of the effect that is lost by restricting TV advertising they estimated would be made up for by an increase in marketing via other marketing. This second scenario gives an estimate of the ‘net’ effect of banning food advertising on television.

Methods

Our analytical framework consists of four steps. First, an intervention or policy lowers exposure to food commercials. Dutch children aged 3-12 see about 12 minutes of advertising per day, of which about 1.8 minutes for food products (personal communication STER).

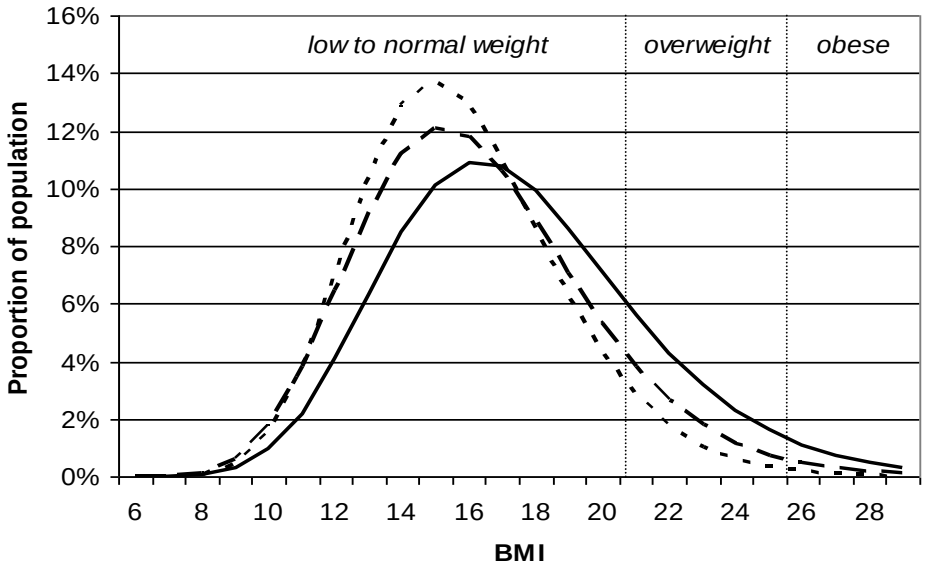
Second, a change in exposure lowers total daily caloric consumption. We searched the literature for studies that quantified the effect of advertising on caloric intake. We found only one. Based on cross-sectional data and a structural equation model Bolton estimates that an increase in food advertising exposure by an additional 10 minutes per week would cause a child to consume one additional snack per week, which would increase the child's caloric intake by approximately 0.6 % (Bolton 1983).

Third, the lower consumption leads to a lower average body weight. For this step we applied the results of a study by Swinburn and colleagues who estimated that two populations of children with a 10% difference in mean energy intake would have a 4.5% difference in mean weight (Swinburn, Jolley et al. 2006).

Finally, we applied Rose's theorem that the mean predicts the number of deviant individuals (Rose and Day 1990) and assumed that the average BMI predicts the number of overweight and obese. The development of the obesity epidemic can be visualised as a shifting population distribution of body mass index (BMI) (Chapter 5). Figure 6.2 illustrates this with an historic example of 11-year old Dutch girls. Over the years, the BMI-curve has shifted to higher values with increased skew to the right. To mimic the changes observed in the data, we fitted the measured prevalence of overweight and obesity (Hurk, Dommelen et al. 2006) to a lognormal curve using the least squares method. We fixed the lower end of the curve in our model. This is consistent with the data and theoretically plausible: under a certain BMI no life is possible. We modelled boys and girls separately in one-year age categories, using age-dependent cut-off points of overweight and obesity (Cole, Bellizzi et al. 2000). The entire BMI distribution can be manipulated by changing its mean value. Shifting the average upwards increases the variance and the rightward skew, which pushes a higher proportion of the population above the thresholds for overweight and obesity. The model was implemented in a spreadsheet (MS Excel).

The uncertainty in this calculation was quite large. The uncertainty in the dose-response relationship between TV advertising and total caloric intake contributed greatly to that. It is based on only one study, which did not quantify the uncertainty. In view of this paucity of evidence, we decided to obtain expert opinion and conducted a Delphi study to estimate the impact of advertising on consumption (Jones and Hunter 1995). We invited 33 academic experts to participate. Twenty five were selected via PubMed because of recent publications on the subject, and 8 were known by the authors to have expertise on the subject. Eight panellists completed the two rounds of questions, a response of 24%. After briefly presenting the Bolton study, we asked for a lower

Figure 6.2 Modelled distribution of the BMI of 11-year old Dutch girls. The dotted line represents the distribution in 1980, the dashed line 1997 and the solid line 2003. The area under the curve represents the total population of Dutch 11-year old girls. Over the years a greater proportion of the distribution lies above the age-specific cut-off points for obesity (26.05) and overweight (21.20), reflecting the rising prevalence of overweight and obesity (Cole, Bellizzi et al. 2000; Hirasing, Fredriks et al. 2001; Hurk, Dommelen et al. 2006).



and higher boundary and a central estimate of the effect.

The degree of uncertainty in the basic scenario (which was based on the Bolton study) was assessed by simply varying relationship between the number of kcal consumed and body mass to reflect the 95% confidence interval of the source study (Swinburn, Jolley et al. 2006). In the scenario that uses the results of the Delphi study we applied simultaneous parametric and nonparametric bootstrapping using the programme Ersatz (Barendregt JJ, Brisbane 2007) to assess uncertainty in the two parameters with the highest uncertainty (Efron and Tibshirani 1994). For the dose-response relation between exposure to advertising and caloric intake we used nonparametric bootstrap on the central estimates of the experts, and for the link between the number of kcal consumed and body mass we used a parametric bootstrap assuming a normal distribution. The number of iterations was 5000.

Results

The model predicts that reducing the exposure of Dutch children to zero would result in a decrease in the prevalence of obesity by 0.2 (95% confidence interval 0.1 to 0.2) percent-points for boys and 0.2% (0.2 to 0.2) for girls. The prevalence of overweight would be reduced by 0.2% (0.2 to 0.3) and 0.4 (0.3 to 0.4), respectively (table 6.1).

The Delphi panel estimated that reducing the exposure of children to food advertising on television by 10 minutes per week would reduce the total caloric intake by 1.4% (95% uncertainty interval: 0.8 to 2.1) (table 6.2).

Using the central estimate and its uncertainty interval as input, the model predicts that reducing the exposure of Dutch children to zero would result in a decrease in the prevalence of obesity by 0.4 (0.2, 0.7) percent-point for boys and 0.6 (0.3, 0.8) for girls. The prevalence of overweight would be reduced by 0.7% (0.4, 0.9) and 0.9% (0.5, 1.4), respectively (table 6.3).

Table 6.1 Prevalence of overweight and obesity (%) of Dutch children aged 6 to 12 in 2003 and in a hypothetical situation in which children are not exposed to food advertising on television, compared to 1980. The dose-response relationship between advertising and caloric intake is based on the Bolton study. (95% confidence interval in brackets)

		2003	*2003, no TV commercials*	1980
Boys	Overweight	11.5	11.3 (11.2 to 11.3)	3.4
	Obesity	3.2	3.0 (3.0 to 3.1)	0.1
Girls	Overweight	15.3	14.9 (14.9 to 15.0)	6.4
	Obesity	4.1	3.9 (3.9 to 3.9)	0.5

Table 6.2 Results of the Delphi study and bootstrap

		Mean	95% UI*
Bolton c.s. found that 10 minutes per week of food advertising increases caloric intake among children by 0.556%. How would you estimate the effect of 10 minutes of TV food advertising per week?	Lower boundary	0.2	0.0 to 0.4
	Higher boundary	4.2	2.4 to 6.3
	Central estimate	1.4	0.8 to 2.1

* UI = uncertainty interval

Table 6.3 Prevalence of overweight and obesity (%) of Dutch children aged 6 to 12 in 2003 and in a hypothetical situation in which children are not exposed to food advertising on television, using the Delphi results for the relationship advertising – caloric intake. (95% uncertainty interval in brackets)

		2003	*2003, no TV commercials*	1980
Boys	Overweight	11.5	10.8 (10.4, 11.1)	3.4
	Obesity	3.2	2.8 (2.5, 3.0)	0.1
Girls	Overweight	15.3	14.4 (13.9, 14.8)	6.4
	Obesity	4.1	3.5 (3.3, 3.8)	0.5

Discussion

Restrictions on food advertising on television may be expected to reduce the prevalence of overweight and obesity among Dutch children by about 1.5%-point at maximum.

Limitations

The uncertainty in the size of these effects reflects the paucity of quantified data in this field of inquiry. The two parameters that add most uncertainty in our model were the dose-response relation between advertising and total caloric intake and the relationship between total caloric intake and body mass in children. We will discuss these factors in the following paragraphs. Compared to the uncertainty in the data, the uncertainty that is inherent in the model itself is likely to be small.

We based our estimate of the number of calories/day that is attributable to food advertising on TV on the Bolton study. In addition, we asked a Delphi panel to estimate the same relationship, using the Bolton study as a starting point for the discussion. Both the Bolton study and the Delphi study have their limitations. In the Bolton study, children in the US in 1977 formed the study population. We apply the results to the population of Dutch children in 2003, so the content (of advertising) and context differs. This may introduce a bias of which the direction is unknown. Secondly, the effect of advertising may be confounded with 'snacking-while-watching', because the time children were exposed to TV advertising was not adjusted for the time spent watching television. If children eat more when the watch TV than they would during alternative activities regardless of whether they see commercials for foods, this

results in the attribution of extra consumption to the influence of advertising when the real cause is the snacking-while-watching. A third possible bias works in the other direction. The study may underestimate the effect of commercials because it corrects for parental influence. But adults cannot be assumed immune to advertising. Longstanding exposure to the influence of advertising may have shifted parents' idea of what constitutes a 'normal' diet for children, even if there is no evidence of an influence on their conception of a healthy diet (Hastings, Stead et al. 2003). Advertising does not so much prompt people to try new things, but it plays a role in reinforcing and normalising behaviour (Hoek and Gendall 2006). Because this issue is important in our analysis and the scarce empirical evidence is difficult to interpret, we performed a Delphi study in which we asked experts with experience in the study of the effects of advertising on obesity. The response was rather low at 24%, and the possibility of a self-selection bias cannot be excluded. This bias would probably be in the direction of an overestimation of the effects of restrictions on advertising, since people who are passionate about the subject can be expected to be more likely to respond. The respondents found the questions difficult to answer, but the resulting effect estimate is over twice that of Bolton. One argument given was that marketing has evolved over the years and is probably more effective now than it was in 1977. Perceived limitations of the Bolton study and (indirect) evidence from other studies also played a role.

For the relationship between consumption to body weight we used the estimate of Swinburn and colleagues (Swinburn, Jolley et al. 2006). They analysed cross-sectional data in which total energy expenditure is precisely measured and assumed to be equal to total energy intake ("EnFlux"), and validate the outcomes with longitudinal data. The authors found that a higher total energy expenditure is associated with a higher body mass, and estimate that two populations with a 10% difference in mean EnFlux would have a 4.5% (95% CI: 3.8%, 5.1%) difference in mean weight. Interestingly, the application of this study results in very high maintenance cost for extra body mass in young children: they would need to consume 200 kcal/day extra per kilogram, whereas for adults this is estimated to be about 15-25 kcal/day/kg (Seidell and Rissanen 2004). The fact that children are more physically active than adults may account for some of this difference but at face value seems insufficient to explain all of it. A higher dose-response relationship would result in larger effect estimates.

Other studies

Few studies give a quantified estimate of the effect of food advertising on childhood obesity, or of the effectiveness of measures to limit the exposure of

children to advertising. In a recent ecological study, Lobstein & Dibbs conclude that advertising could explain up to half of the variation between countries' overweight prevalence figures, which suggests a much larger effect than we found (Lobstein and Dibb 2005). Apart from the above limitations of our study, one likely explanation for this discrepancy is the possibility that their exposure parameter, the number of obesogenic advertisements per hour, acts as proxy for marketing pressure more in general, and perhaps also for a general attitude of the public (including parents) towards advertising.

While few other studies estimate the magnitude of the effect of food advertising, television watching as such has been linked to obesity. In a 4-year longitudinal study Dietz & Gortmaker found a dose-response relationship of about 0.6% increase in obesity prevalence per extra hour of television, after correcting for past obesity and several socio-economic characteristics (Dietz and Gortmaker 1985). In a later study they conclude that more than 60% of overweight incidence in a representative sample of US children could be linked to excess television viewing time (Gortmaker, Must et al. 1996). Robinson conducted a small-scale experiment that reduced the time 9-year old children spent watching television from about 15 to 9 hours per week. After 7 months the BMI of the intervention group was 0.45 lower than that of the control group.

Three pathways have been postulated to explain the association between TV viewing and obesity: the sedentary nature of watching television, the influence of advertising and a tendency to snack while watching TV (Robinson 1999). What proportion should be attributed to advertising remains unclear.

Logically, the influence of television viewing must be larger than that of television advertising. Nonetheless, Haby et al. estimate that reducing TV advertising would be more effective in terms of the prevention of childhood obesity (Haby, Vos et al. 2006). This is probably due to the assumption that reducing advertising benefits a larger age group than the one-time intervention on which they base the effect of reducing TV viewing time, because the reported BMI-reduction per child is smaller for a reduction in advertising than for a reduction in television viewing.

If advertising on television were banned (either by voluntary agreement with the industry or by legal measures), would the projected reductions in obesity prevalence really materialize? Probably not fully. Marketing budgets might not be reduced in size, but merely used to deliver the same message via different channels. At least part of the obesity-preventing effect of restrictions on television food advertising could thus be compensated by increased marketing efforts via other channels. For example, the internet opens new ways to reach children with marketing messages (e.g. via 'advergaming') so the total

commercial pressure on children's diets is likely to be increasing. Around the year 2000 about 75% of the marketing budgets in European countries were spent on television advertising, but the market share of television seems to be diminishing (European Heart Network 2004). Bans on television advertising would hasten this diversification of marketing efforts. However, it seems unlikely that this would fully compensate the loss of TV as an advertising medium. Marketers use a mix of strategies and channels to influence their target audience, and it seems reasonable to assume that if the opportunities to reach children diminish, some of its power will be lost.

The total effect of marketing on childhood obesity rates is likely to be considerably larger than the results of this study may seem to indicate. Firstly because there are other marketing channels besides television advertising. And secondly because marketing directed at adults may indirectly affect children's diets. If parents regard frequent consumption of fast-food and snacks to be part of a normal eating pattern, their children are likely to share in their views and diets.

Policy options

In practice, crafting effective policies to limit the obesity-generating effect of food advertising is likely to prove difficult. In many countries the effects of legal measures can be undercut by broadcasting from abroad, and by a shift of marketing budgets to other channels. For the European situation, the effect of broadcasting from abroad may be remedied by EU-wide or global measures (European Heart Network 2004; Knai, Lobstein et al. 2005). Preventing an increase of marketing via alternative channels requires a policy that has a broader focus than television alone, but it may be challenging to either design practical legislation or, alternatively, to convince the industry to stop marketing energy-dense food on a voluntary basis. The modest effect that can be expected from limiting TV food advertising as an isolated measure also underscores that the promotion of healthier eating in childhood will need to involve a range of measures in a variety of settings, not just changing the ways foods are marketed to children. Restrictions on advertising may well be part of this package. Another option may be to counterbalance the weight-promoting effect of commercials by advertising healthier foods (Lobstein and Dibb 2005). This is a more positive approach that may meet with more enthusiasm among policy makers and more progressive minds in the commercial sector, though in view of the above the expectations of its effectiveness should not be put up too high. Finally, reducing the time children watch television is likely to reduce the

prevalence of childhood obesity. There is, however, little research that shows how to achieve it.

How should we look at the size of the obesity-preventive effect of limits to TV advertising, bearing in mind how difficult it may be to take effective measures? At maximum 1.5% less overweight and obese children may not seems overly impressive, but in absolute terms it still concerns about 14,000 children in the Netherlands. In the struggle against tobacco smoking, many of the measures taken each had a small effect, if an effect could be demonstrated at all. However, the combined effects are becoming visible in a reduced smoking prevalence in most western countries. The effect of some measures may have been symbolic, but it is by getting the problem in the public attention again and again that a society-wide consensus may grow that further steps are needed. Limiting exposure of children to marketing of unhealthy foods can be one step in the building of a society-wide determination to spare children the hazards and stigma of obesity.

Strengthening the evidence base

More quantitative evidence on the effect of marketing on childhood obesity might strengthen the case for policy measures that limit (commercial) freedom and go against powerful vested interests. However, the ultimate proof will not be attainable. Experimental studies in a realistic setting are virtually impossible, and observational studies are hampered by the fact that exposure to television advertising is difficult to measure and highly correlated with the time spent watching television and permissive parenting styles. Compared with clinical decision-making where the evidence base is dominated by randomised controlled trials with high internal validity, the evidence base for obesity prevention needs many different types of evidence, often including the informed opinions of stakeholders to ensure external validity and contextual relevance. A framework for translating evidence in the field of obesity prevention into action is available (Swinburn, Gill et al. 2005), but how much evidence justifies action will always be a matter of judgment.

Conclusions

Restrictions on food advertising on television may reduce the prevalence of overweight and obesity among children. Effective measures are likely to be more comprehensive than only TV and, in the case of the Netherlands, EU-wide. They could nonetheless be an element in a broader effort to make children's diets healthier.

References

Bolton, R. (1983). "Modeling the impact of television food advertising on children's diets." Current Issues and Research in Advertising 6(1): 173-199.

Cole, T. J., M. C. Bellizzi, et al. (2000). "Establishing a standard definition for child overweight and obesity worldwide: international survey." Bmj 320(7244): 1240-3.

Dietz, W. H., Jr. and S. L. Gortmaker (1985). "Do we fatten our children at the television set? Obesity and television viewing in children and adolescents." Pediatrics 75(5): 807-12.

Efron, B. and R. Tibshirani (1994). An Introduction to the Bootstrap, Chapman & Hall/CRC.

European Heart Network (2004). The marketing of unhealthy food to children in Europe. Brussels.

Gortmaker, S. L., A. Must, et al. (1996). "Television viewing as a cause of increasing obesity among children in the United States, 1986-1990." Arch Pediatr Adolesc Med 150(4): 356-62.

Haby, M. M., T. Vos, et al. (2006). "A new approach to assessing the health benefit from obesity interventions in children and adolescents: the assessing cost-effectiveness in obesity project." Int J Obes (Lond) 30(10): 1463-75.

Hastings, G., M. Stead, et al. (2003). Review of research on the effects of food promotion to children. Glasgow, University of Strathclyde.

Hawkes, C. (2004). Marketing food to children: The Global Regulatory Environment. Geneva, WHO.

Hirasing, R. A., A. M. Fredriks, et al. (2001). "[Increased prevalence of overweight and obesity in Dutch children, and the detection of overweight and obesity using international criteria and new reference diagrams]
Toegenomen prevalentie van overgewicht en obesitas bij Nederlandse kinderen en signalering daarvan aan de hand van internationale normen en nieuwe referentiediagrammen." Ned Tijdschr Geneeskd 145(27): 1303-8.

Hoek, J. and P. Gendall (2006). "Advertising and obesity: a behavioral perspective." J Health Commun 11(4): 409-23.

Hurk, K. V. d., P. v. Dommelen, et al. (2006). Prevalentie van overgewicht en obesitas onder jeugdigen 4-15 jaar in de periode 2002-2004. Leiden, TNO Kwaliteit van Leven.

IOM (2005). Food Marketing to Children and Youth: Threat or Opportunity? Washington, D.C., The National Academies Press.

James, W. P. T., R. Jackson-Leach, et al. (2004). Overweight and obesity (high body mass index). Comparative quantification of health risks. Geneva, World Health Organization.

Jones, J. and D. Hunter (1995). "Consensus methods for medical and health services research." Bmj 311(7001): 376-80.

Knai, C., T. Lobstein, et al. (2005). "Childhood obesity: the case for binding international legislation." Eur J Public Health 15(6): 559-61.

Lobstein, T. and S. Dibb (2005). "Evidence of a possible link between obesogenic food advertising and child overweight." Obes Rev 6(3): 203-8.

Peeters, A., J. J. Barendregt, et al. (2003). "Obesity in adulthood and its consequences for life expectancy: a life-table analysis." Ann Intern Med 138(1): 24-32.

Robinson, T. N. (1999). "Reducing children's television viewing to prevent obesity: a randomized controlled trial." Jama 282(16): 1561-7.

Rose, G. and S. Day (1990). "The population mean predicts the number of deviant individuals." Bmj 301(6759): 1031-4.

Seidell, J. C. and A. M. Rissanen (2004). Prevalence of Obesity in Adults: The Global Epidemic. Handbook of Obesity. G. A. Bray and C. Bouchard. New York, Marcel Dekker, Inc.

Swinburn, B., T. Gill, et al. (2005). "Obesity prevention: a proposed framework for translating evidence into action." Obes Rev 6(1): 23-33.

Swinburn, B. A., D. Jolley, et al. (2006). "Estimating the effects of energy imbalance on changes in body weight in children." Am J Clin Nutr 83(4): 859-63.

Veerman, J. L., J. J. Barendregt, et al. (in press). "Stemming the obesity epidemic: a tantalizing prospect." Obesity.

PART C

Assessing validity in HIA

7

The validity of predictions in Health Impact Assessment

Veerman JL, Barendregt JJ, Van Beeck EF, Mackenbach JP. Validity in Health Impact Assessment. *Journal of Epidemiology and Community Health.* 2007 Apr; 61 (4): 362-6.

Abstract

Introduction
An essential characteristic of health impact assessment (HIA) is that it seeks to predict the future consequences of possible decisions for health. These predictions have to be valid, but as yet it is unclear how validity should be defined in HIA.

Aims
To examine the philosophical basis for predictions and the relevance of different forms of validity to HIA.

Conclusions
HIA is valid if formal validity, plausibility and predictive validity are in order. Both formal validity and plausibility can usually be established, but establishing predictive validity implies outcome evaluation of HIA. This is seldom feasible owing to long time lags, migration, measurement problems, a lack of data and sensitive indicators, and the fact that predictions may influence subsequent events. Predictive validity most often is not attainable in HIA and we have to make do with formal validity and plausibility. However, in political science, this is by no means exceptional.

Introduction

There are various definitions of health impact assessment (HIA) and a wide variety of activities have been termed HIA. In agreement with Kemm (Kemm 2003) we would argue that in essence, HIA seeks to (1) predict the future consequences of possible decisions regarding projects, programmes or policies for health; and (2) inform policy decisions on the basis of these predictions. This restricts HIA to what some would call "prospective" HIA. Generally, the HIA procedure starts with screening for policies with potentially modifiable health consequences, followed by "scoping" to determine who should perform the assessment, and how. The results of the assessment are communicated to relevant parties, and finally evaluation and monitoring of health effects takes place. Discussions about evaluation of HIA to date tend to focus on their effects on decision making and participation, whereas less attention has been paid to the validity of predictions (Quigley and Taylor 2004; Parry and Kemm 2005). Notwithstanding the futility of the exercise if it exerts no influence on policy, in order to be valuable to policy makers and stakeholders, the predictions in HIA need to be valid. Validity is the expression of the degree to which a measurement measures what it purports to measure (Nieuwenhuijsen 2003). As yet, it is unclear how the validity of predictions should be defined in HIA, and how it can be assessed. In this chapter, we discuss the assessment of the validity of predictions in HIA. The objective of this chapter is twofold: it aims to 1) discuss the assessment of validity of predictions in HIA; and 2) propose a checklist to establish the validity of predictions in HIA. Where possible, we think that these predictions would preferably be quantified, but qualitative work can be judged by the same standards. The chapter is structured in three sections. The first section presents the conceptual basis for predictions in HIA by referring to the work of Popper. In the second section, the concepts of validity are reviewed and discussed. In the third section, we apply Popper's logical structure to construct a checklist to establish the validity of predictions in HIA and critically discuss issues in assessing validity.

Figure 7.1 The logical structure of Popper's unity of method in all theoretical or generalizing sciences. Initial conditions are linked to outcomes by a theoretical framework. When two of the three are provisionally assumed valid, the third can be inferred.

A conceptual basis for predictions

Before elaborating on the concept of validity for HIA, we will take a closer look at predictions. For an understanding of predictions it is useful to look at the work of Popper (Popper 1957). He asserts that explanation, prediction and testing share three common elements: initial conditions, outcomes, and a theoretical framework (fig. 7.1). When we are testing a hypothesis, we accept our initial conditions and outcomes and make inferences about the theoretical framework. For explanation, the outcomes and the theoretical framework or the initial conditions are (provisionally) considered reliable, which allows us to make inferences about either the theoretical framework or the initial conditions. Prediction requires that the initial conditions and the theoretical framework are considered valid. (According to Popper, this is a case in which we apply our scientific results.) One cannot do two things at a time—that is, testing a theory and at the same time making predictions is logically impossible.

Thus, HIA assesses the initial conditions and policy plans, and uses the theoretical framework of science (e.g. epidemiology, demography, physics, and also the social sciences) to make predictions. In essence, HIA is a deductive activity: it uses (general) theory to make statements about specific situations. Though results may be extrapolated to similar policies in other situations, HIA is not, in the first place, concerned with generating knowledge that is generalisable; it aims to assess the effects of a specific policy on a specific population in a specified environment, using a theoretical framework that has resulted from previous scientific work. A HIA project can be compared with the daily weather forecast, which is not meant to check whether the model behind it is correct but to advise on what coat to wear tomorrow.

Concepts of validity in HIA

When can we consider the predictions in HIA valid? We will examine different forms of validity and consider their applicability to HIA and illustrate this with reference to the study on an aspect of the European Union (EU) policy on fruits and vegetables (FV) (Chapter 3 of this thesis). This policy guarantees producers a minimum price for their products by withdrawing FV from the market when prices drop below a specified level. On the basis of a simulation model with data on the EU agricultural policy, FV consumption and health, our study concluded that reform might, at maximum, result in modest health gains for the Dutch population: an estimated annual gain of 1930 disability-adjusted life years or an increase in life expectancy by 3.8 days for men and 2.6 days for women.

In epidemiological studies, internal and external validity are important concepts (Campbell and Stanley 1963; Campbell 1969). Internal validity indicates the degree to which results of research support or refute a causal relationship between the dependent and independent variables. A HIA must be based on a theoretical framework that ultimately rests on research that is internally valid. HIA itself however is not primarily intended to investigate causal relationships; as we have seen, these simply have to be assumed valid in order to make prediction possible. Therefore the concept of internal validity does not directly apply to HIA.

External validity refers to the degree to which the theoretical knowledge resulting from research can be generalised to other populations. However, in a HIA, we are trying to do the reverse: established generalisable knowledge is applied to a specific population. HIA is, as it were, at the receiving end of external validity: the theory used to make the predictions must of course be relevant for the population concerned. External validity of a HIA itself is not of primary concern. Nonetheless, conclusions may be generalisable to similar situations. For example, if the validity of our study on the EU FV policy is accepted, it would be reasonable to expect a similar effect of abolishing withdrawal support on life expectancy in the UK. (However, since the British consume less FV they might benefit slightly more.)

In psychometric research Cronbach's concepts of validity are in wide use. In contrast with internal and external validity, they are more appropriate for measurement instruments than for complete studies. Cronbach distinguishes face validity, content validity, criterion validity and construct validity. Face validity or plausibility is the degree to which an observer deems that the theoretical framework is understandable, applicable and plausible (Cronbach 1971; Sargent

1996). It is closely related to credibility, which is the confidence that (potential) users have in a theory or model. Plausibility should clearly be considered relevant to HIA. The causality of the relationships in a HIA must be credible, both in qualitative terms (is there a likely mechanism between cause and effect?) and in quantitative terms (is the strength of the association plausible?). In our example of the EU FV policy, the question would be whether abolishing withdrawal support would result in higher consumption of FV. We postulated that, at maximum, the amount currently withdrawn would enter the market, that this would lower prices, which in turn would increase sales and ultimately consumption. But alternative scenarios are conceivable—for example, if producers decide to produce something else—and the assumption that consumption rises in equal proportion to the amounts now taken off of market can only be an overestimate.

Plausibility may look vague and arbitrary, but then so is causality. Causation cannot be proven, but ultimately rests on judgement, properly supported by evidence. Two generally accepted minimum conditions for a causal relationship are that the cause precedes the consequence and that there is a correlation between the two (Rothman and Greenland 1998). A third requirement is that there is a plausible mechanism. The difficulty with this requirement is: who decides what is plausible? De Groot posits that ultimately, the forum of the scientific community decides (De Groot 1969). The requirement of plausibility can therefore be translated as the obligation to convince one's peers, and this can be done by arguments based on logical inference and empirical data. A HIA therefore has to present evidence as to why the predictions are likely to be correct, especially as it is not intended to test hypotheses but to inform a policy process of (accepted) scientific knowledge. Plausibility in this definition clearly is not a superficial matter. Its synonym face validity seems to suggest otherwise and is therefore best avoided.

Criterion validity is the degree to which outcomes are confirmed by a "gold standard". For HIA studies as a whole there are no such standards, but there may be for measurement instruments used in HIA. Criterion validity is sometimes subdivided into "predictive" and "concurrent", depending on when the gold standard is measured (Gliner and Morgan 2000). Predictive validity is the degree to which predictions are confirmed by facts. We would propose to turn things around and argue that the concept of criterion validity is redundant if one accepts the idea of predictive validity. Predictive validity should be established, and this can be performed using 'gold standard' tests to the degree that these tests accurately measure the concept of interest.

Content validity is concerned with the question of whether all aspects of the phenomenon to be measured are represented in the appropriate proportions. Translated to HIA, the question is whether all the relevant determinants and health effects have been included in a plausible order of magnitude. This is a matter of judgement and can therefore be considered part of plausibility, removing the need for content validity as a separate form of validity in HIA.

Construct validity is the degree to which the outcomes correlate with those of other instruments that purport to measure the same construct. It applies to hypothetical concepts that cannot be measured directly. HIA should reflect the current scientific understanding, and so would in principle avoid using methods or concepts of which the construct validity has not been established in other research. We do not see an important role for construct validity in HIA as such.

Formal validity concerns how well an argument conforms to the rules of logic to arrive at a conclusion that must be true, assuming that the premises are true (Verlinden 1998). Though not always explicitly, formal validity plays a role in any research. Besides argumentation it is also about the correctness of calculations and other methodological aspects of scientific endeavour. Applied to HIA, formal validity is concerned with the correct application of correct methods. Clearly, this must be in order for a prediction to be valid.

We therefore propose that three types of validity are relevant for HIA: plausibility, formal validity and predictive validity, whereby plausibility broadly refers to the subject matter, formal validity to the method, and predictive validity to direct empirical evidence. Other types of validity can either be considered redundant or are unimportant in HIA.

Establishing validity of predictions in HIA

The predictions in a HIA can be considered valid if plausibility, formal validity and predictive validity are in order. In Annex I, we use Popper's logical structure (fig. 7.1) to construct a list of aspects of an HIA that need to be examined in order to determine its validity. This checklist helps to systematically examine a HIA study by subsequently focusing on a number of questions regarding the plausibility and formal validity of the assessment of the initial conditions and of the theoretical framework that was applied, and on the predictive validity of the study. Annex II illustrates the use of the checklist by highlighting some of the points that an independent assessment of the validity of the EU FV study could focus on.

Assessing validity is not a problem for plausibility and formal validity. Formal validity can be checked, though it often requires time, effort and expertise. For example, in the case of the EU FV study, the data and calculations in the spreadsheet used could be checked, though this would require some understanding of life table analysis. Plausibility will superficially be assessed by policy-makers and stakeholders, but should really be checked by independent experts in the relevant scientific discipline (which will increase face validity for the first group as well). For the study on the EU FV policy this would include specialists in agricultural economics and epidemiology.

In operational research, which is concerned with building simulation models of (military, industrial and economic) processes in order to optimise outcomes, an established approach to validation of models is "independent verification and validation" (IV&V) (Sargent 1996). In IV&V, an independent party examines the model and judges its validity. Verification should be understood as checking the formal validity (are all calculations correct?), whereas validation refers to the examination of plausibility and predictive validity. Although a complete IV&V can be costly, it could well be considered a means to establish formal validity and plausibility in HIA.

In HIA, the predictive validity of entire studies usually cannot be established. This would require outcome evaluation of completed HIA studies, which is difficult for a number of reasons.

In the first place, there is often a long time lag between a change in policy and the corresponding health effects. For example, a decision to stop using asbestos in The Netherlands was taken decades ago but the incidence of mesothelioma is still rising (Swuste, Burdorf et al. 2004). Besides having to wait a long time before measurement is at all possible, this time lag makes it expensive and increases the loss to follow-up. The latter is made worse by migration, which may be differential: those experiencing most inconvenience by a development may be the first to move away (Parry and Stevens 2001).

Second, many factors influence the same outcomes that the HIA considers, often to a much greater extent. This may obscure any effect of the policy decision under consideration. For example, a trend in smoking may obscure any effect of changes in the EU FV policy on cardiovascular disease and cancer.

Third, many health problems are hard to measure. Routine data may not be available at the appropriate geographical scale or may not be measured frequently enough to pick up changes due to the policy under scrutiny.

Fourth, HIA intends to influence policy, but if successful it invalidates its own predictions. This is what Popper called the "Oedipus effect" after the

mythical figure who killed his father whom he had never seen because of a prophecy that had caused his father to abandon him as a child (Popper 1957). (With regard to the EU agricultural policy there is little risk of this effect occurring, if only because no direct communication with stakeholders took place—which distinguishes that study from a HIA exercise.)

Finally, in most cases a control group is lacking. One intervention group and no control does not make for a strong research design (Parry and Stevens 2001).

When complete assessment of predictive validity is not possible, sometimes partial predictive validation is feasible by focusing on intermediary outcomes (Lindholm and Rosen 2000). For example, the effect of a price change on the consumption of FV is measurable shortly after a tax reduction whereas the outcome of interest (cardiovascular disease and cancer) is unlikely to ever be measurable.

Predictive validity can also be supported by using knowledge from initial conditions and outcomes in the past: this allows testing of the theory (historical data validation) (Sargent 1996).

Historical data validation would be confronted with the same problems as predictive validity in general, except that it saves a long wait for the results. Although theoretically possible, we know of no example in the literature on HIA.

External reviews of HIA exercises have been carried out, but few focus on the validity of predictions. We know of no example in which the predictive validity was established. Formal validity and, to a lesser extent, plausibility have been assessed. In practice, these two forms of validity are closely connected. For example, the evaluation of the Alconbury Airfield HIA assessed formal validity among others by checking that reasons for including and excluding determinants are clearly stated. The recommendation to pay more attention to impacts that are hard to quantify can be considered as an element of plausibility (Close 2001).

Finally, even in a properly validated HIA, unanticipated adverse effects may arise. Science cannot (ever) claim to provide all knowledge needed, so if not for establishing predictive validity, HIA studies should make recommendations for the monitoring of health outcomes to aid early detection.

Conclusion

Predictions are at the core of HIA, but predictive validity will most often prove unattainable. Instead, we have to make do with less than the gold standard and

assess HIA studies and methods for plausibility and formal validity only. It may be of comfort to know that in political science this is by no means exceptional. Few decisions can be taken with the confident knowledge of relevant and thorough outcome evaluations.

Acknowledgement

We thank dr. E.F. van Beeck for his helpful comments.

References

Campbell, D. T. (1969). "Reforms as experiments." American Psychologist **24**: 409-429.

Campbell, D. T. and J. C. Stanley (1963). Experimental and quasi-experimental designs for research. Chicago, Rand McNally.

Close, N. (2001). Alconbury Airfield Development HIA evaluation report.

Cronbach, L. J. (1971). Test validation. Educational Measurement (2nd Ed.). R. L. Thorndike. Washington, D. C., American Council on Education.

De Groot, A. D. (1969). Methodology: foundations of inference and research in the behavioral sciences. The Hague, Mouton.

Gliner, J. A. and G. A. Morgan (2000). Research methods in applied settings: an integrated approach to design and analysis, Mahwah, N.J. [etc.] : Lawrence Erlbaum Associates.

Kemm, J. (2003). "Perspectives on health impact assessment." Bull World Health Organ **81**(6): 387.

Lindholm, L. and M. Rosen (2000). "What is the "golden standard" for assessing population-based interventions?--problems of dilution bias." J Epidemiol Community Health **54**(8): 617-22.

Nieuwenhuijsen, M. J., Ed. (2003). Exposure assessment in occupational and environmental epidemiology. Oxford Medical Publications. New York, Oxford University Press.

Parry, J. and A. Stevens (2001). "Prospective health impact assessment: pitfalls, problems, and possible ways forward." Bmj **323**(7322): 1177-82.

Parry, J. M. and J. R. Kemm (2005). "Criteria for use in the evaluation of health impact assessments." Public Health **119**(12): 1122-9.

Popper, K. R. (1957). The Unity of Method. The Poverty of Historicism. London, Routledge & Kegan Paul**:** 130-43.

Quigley, R. J. and L. C. Taylor (2004). "Evaluating health impact assessment." Public Health **118**(8): 544-52.

Rothman, K. J. and S. Greenland (1998). Causation and Causal Inference. Modern Epidemiology. K. J. Rothman and S. Greenland. Philadelphia, Lippincott Williams & Wilkins**:** 7-28.

Sargent, R. G. (1996). Verification and validation of simulation models. Proceedings of 1996 Winter Simulation Conference**:** 55-64.

Swuste, P., A. Burdorf, et al. (2004). "Asbestos, asbestos-related diseases, and compensation claims in The Netherlands." Int J Occup Environ Health **10**(2): 159-65.

Verlinden, J. (1998). "Formal validity review." from http://www.humboldt.edu/~act/HTML/tests/validity/review.html (Accessed 19 December 2005).

ANNEX I

HOW TO ESTABLISH VALIDITY OF PREDICTIONS IN HIA – A CHECKLIST

Plausibility

Definition: degree to which an observer deems that the theoretical framework is understandable, applicable and plausible.
To be established by researchers, external experts and stakeholders. (Stakeholders' judgement has to be interpreted with caution, bearing in mind the interests the stakeholder may have in the outcomes.)

Initial conditions

- Is the policy plan/project described accurately?
- Is the description of the baseline situation accurate?
- Has uncertainty in the initial conditions been assessed?
- How robust is the model to (foreseeable) changes in the initial conditions? For example if increases in air transport are likely, have these been included in the assessment of the health consequences of building housing near an airport?

Theoretical framework

- Is the causal web underlying the analysis valid according to the state of the pertaining scientific field?
- Is the order of magnitude of the causal relationships in concurrence with current scientific knowledge?
- Has the degree of certainty of the causal relationships been described?
- Are all exposures to determinants of health that are likely to result from the intended policy/project included in the analysis?
- Of the exposures included, have all plausible health outcomes been included?
- Have all populations likely to be affected by the policy been included in the analysis?
- If available, how do the results of similar exercises compare with the predicted effects in this HIA? Can any differences be satisfactorily explained by differences in the initial conditions (including intervening events during the period of analysis) or lack of formal validity of the previous analyses?

Formal validity (= verification)

Definition: the degree to which correct methods have been applied correctly.
To be established by researchers and external experts.

Initial conditions
Have the right methods been applied to:

- describe the policy proposal; and
- describe the baseline situation,

and have both these sets of methods been applied correctly?

Theoretical framework
Have the right methods been applied to:

- construct the causal framework;
- estimate the order of magnitude of the causal relationships;
- estimate degree of certainty of the causal relationships;
- find all significant determinants of health of which the exposure changes as a result of the proposed policy;
- find all health outcomes that result from changes in exposure;
- identify populations likely to be affected by the policy been included in the analysis;

and have these methods been applied correctly?

Predictive validity

Definition: the degree to which predictions are confirmed by facts
To be established by researchers and external experts.

Historical predictive validity

- Are historical data on initial conditions and subsequent outcomes available on which the model underlying the HIA can be tested?
- If testing has been performed, how well does the model "postdict" these outcomes, and can any differences between model and empirical data be explained satisfactorily by differences in the initial conditions or uncertainty in initial conditions (including intervening events during the period of analysis) and/or outcomes?

In retrospect

- To what extent did the predictions materialise?

ANNEX II

THE USE OF THE CHECKLIST OF VALIDITY OF PREDICTIONS IN HIA – THE EXAMPLE OF THE EU WITHDRAWAL SUBSIDIES FOR FRUITS AND VEGETABLES

The use of the checklist is illustrated with the example of an HIA we conducted on the EU Common Agricultural Policy to withdraw fruits and vegetables (FV) from the market when prices drop below an intervention-threshold (Chapter 3). The withdrawn products are mostly composted. The assumption underlying the assessment was that ending withdrawal support would maximally lead to a proportional increase in consumption equal to the increase in availability of FV. The health gain for the Dutch population in this scenario was estimated at 1930 DALY per year or an increase in life expectancy by 3.8 days for men and 2.6 days for women. Below, we highlight some of the points an independent assessment of the validity of our study could focus on, without of course intending to relieve the assessors from their responsibility to make their own judgements.

Plausibility

The plausibility of this study is best assessed by a team that includes epidemiologists and economists specialised in international agricultural trade and econometrics.

Initial conditions

Is the policy plan/project described accurately ?	Leaving aside difficult issues of political feasibility, the brief descriptions of the current EU policy and the proposed intervention should be judged. One question would be how soon such policy change could enter into force.
Is the description of the baseline situation accurate?	The data on amounts withdrawn and FV consumption were relatively old. The amounts may change rapidly while consumption patterns are likely to remain stable.
Has uncertainty in the initial conditions been assessed?	Only with respect to the amount of FV withdrawn from the market, not consumption or health outcomes. Is this justified by the assumption that uncertainty in those factors is relatively minor?
How robust is the model to	Maximum amounts and compensation for

(foreseeable) changes in the initial conditions?	withdrawals were to be lowered for a number of products over the years; this was not taken into account. Would this lead to overestimation of the potential effects? We ignored trends in FV consumption and disease occurrence. Is this justified?
Theoretical framework	
Is the causal web underlying the analysis valid according to the state of the pertaining scientific field?	The association between FV consumption and health is generally accepted. Validity assessment could focus on the plausibility of the CAP withdrawal policy influencing FV consumption.
Is the order of magnitude of the causal relations in concurrence with current scientific knowledge?	The effects of changes in FV consumption on health as these were taken from recent reviews. RRs were not age-adjusted and applied uniformly to all ages. Would this bias results? The effect of policy change on FV consumption was only explored in a 'maximum effect' scenario. A more realistic scenario would require an econometric equilibrium model. We did not find one, and we also found no similar analysis in the literature.
Has the degree of certainty of the causal relations been described?	Two sources of uncertainty have been taken into account: the amount withdrawn and the relative risks of disease for changes in FV consumption. However, additional uncertainty remains, especially in the effect of policy change on FV consumption. Was the present analysis sufficient?
Are all exposures to determinants of health that are likely to result from the intended policy/project included in the analysis?	Other effects (e.g. on FV producers) are conceivable, but were estimated to be negligible compared to the effects on FV consumption. Is this justified?
Of the exposures included, have all plausible health outcomes been included?	We included CVD and cancer at a number of sites. For other diseases the evidence was deemed insufficient by the authors of the reviews we based the analysis on.
Have all populations likely to be affected by the policy been included in the analysis?	The analysis was restricted to the general Dutch population, with a qualitative comment that reform is likely to benefit low SES groups more than proportionally.
If available, how do the results of similar exercises compare with the predicted	To our knowledge no similar exercise has been conducted, but an independent assessor might know of similar work.

effects in this HIA? Can any differences be satisfactorily explained by differences in the initial conditions (including intervening events during the period of analysis) or lack of formal validity of the previous analyses?		

Formal validity (verification)

Formal validity could be assessed by the same team that assessed plausibility. This team would probably want to have the spreadsheets used for the analysis.

Initial conditions

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Description of policy proposal	We based our brief description mainly on a previous health-focused analysis (Schäfer Elinder 2003). It might have been more elegant to refer to EU documents and reports.	Assessors could check for inaccuracies in the description.
Description baseline situation	See above. Would an independent assessor agree with our choice of the source of data on FV consumption and disease occurrence?	Assessors could check the spreadsheet and the paper for inaccuracies in the numbers.

Theoretical framework

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Construction of causal framework	Lacking an econometric model, the method to reason from policy change to FV consumption was rather simple. A life table approach was used to model the health effects of changes in FV consumption.	Though it had limitations, the analysis was straightforward.
Estimation of magnitude of causal relations	For the relation policy - consumption simple assumptions were made. We conducted a PubMed search for recent meta-analyses or reviews on the relative risks of FV to disease, and contacted authors of older work if no result.	Perhaps other estimates for the RRs for cardiovascular disease would have been found if other authors had been approached.
Estimation of degree of certainty of causal relations	Uncertainty in the effect of policy on consumption could only partly be included; for the RRs confidence intervals in reviews were used. Bootstrapping was used to assess the overall uncertainty.	Check the procedures used and re-run a bootstrap procedure.
Search for significant determinants of health of which exposure changes as a result of the proposed policy	No formal search was conducted because substantial other effects were not deemed plausible.	Not applicable.
Search for health outcomes that result from changes in exposure	Relied on reviews of the effect of FV consumption on CVD and cancer. Restricting inclusion to diseases with statistically significant relationship with FV may lead to underestimation of effect.	Was the search strategy correct?
Search for populations likely to be affected by the policy	No search; restricted to general Dutch population.	Not applicable.

Predictive validity

Historical predictive validity

Are historical data on initial conditions and subsequent outcomes available on which the model underlying the HIA can be tested?	Such data would need to link changes in agricultural policy to FV consumption and health. We do not know of any.
If testing has been performed, how well does the model 'postdict' these outcomes, and can any differences between model and empirical data be explained satisfactorily by differences in the initial conditions or uncertainty in initial conditions (including intervening events during the period of analysis) and/or outcomes?	Not applicable.

In retrospect

To what extent did the predictions materialise?	Even if the proposed policy change would be effectuated, it would be impossible to measure any effect on population health. It might be possible to measure changes in FV consumption but even this would require a large sample size to detect the modest changes that are expected.

Literature

Schäfer Elinder, L. (2003). Public Health aspects of the EU Common Agricultural Policy. Developments and recommendations for change in four sectors: Fruit and vegetables, dairy, wine and tobacco. Stockholm, Swedish National Institute of Public Health.

8

General discussion

Introduction

This thesis set out to answer the following questions:

1. What methods have been used in previous studies to make quantified predictions of the health impact of policy outside the health sector?
2. Can a combination of potential impact fractions and multi-state life tables be used to predict changes in summary measures of population health due to policies outside the health sector?
3. How can the validity of predictions in HIA be established?

In the following section we will summarise the findings of the studies presented in chapters 2 to 7 and answer the study questions. We subsequently present some reflections about quantitative prediction for HIA, and end with conclusions and recommendations.

Summary of findings

What methods have been used in previous studies to make quantified predictions of the health impact of policy outside the health sector?

The results from the review of HIA case studies presented in part A of this thesis suggest that quantification is comparatively rare in HIA. Methods are available in the areas of environmental health and, to a lesser extent, traffic accidents, infectious diseases and behavioural factors. The methods can be divided into two sets, the first of which concerns methods to estimate the impact of a policy on exposure to determinants of health. This was mostly done on the basis of routine data and measurements. Various kinds of modelling also played a role, particularly in the fields of traffic related and environmental factors. Where data were not available, experts' estimates and authors' assumptions were used.

The second set of methods calculated health outcomes based on changes in exposure of populations to determinants of health. Here, variants of epidemiological and toxicological risk assessment were used. Sometimes these formed part of more elaborate mathematical models. The effects of increased employment and income were quantified in two studies, but on the whole, quantification was mainly limited to proximal determinants of health, i.e., factors that influence health fairly directly.

The validity of the methods has not in all cases been clearly established. The degree of certainty of the estimates was seldom quantified. Time horizons

and outcome measures differed by method and study. Summary measures of population health were not used.

Our review has several limitations. For example, we may have missed studies that qualify as HIA but did not show up in our search of the (largely grey) literature, though we tried to be systematic. In addition, there must be many methods in different fields of scientific endeavour that could be used in HIA but haven't yet been used. Furthermore, the study provides a broad overview but has little room for in-depth judgement of quality. The strengths of our review are that it covers diverse scientific areas, which use different methods and normally have little mutual contact, and that it offers criteria by which to judge (the reporting of) quantitative HIA.

The scientific community would do well to establish a common core of methods and minimum requirements for the reporting of health effects in HIA. For HIA studies that take an experimental or economic perspective (Øvretveit 1998) one might expect a development similar to the one that has taken place (and is still taking place) in health economic evaluation, where authoritative guidelines increasingly standardise practice (Gold, Siegel et al. 1996).

Can a combination of potential impact fractions and multi-state life tables be used to predict changes in summary measures of population health due to policies outside the health sector?

To answer this question we will analyse at the four case studies discussed in part B of this thesis. As explained in chapter 1, our approach to the effect estimation in HIA consists of a two-step process. In the first step we assess the effect of the policy on exposure to determinants of health. In the second step the results are translated into health outcomes using the combination of the potential impact fraction and the proportional multi-state life table. We will now first summarise the results of the case studies and analyse them, bearing in mind the results on chapter 2 (which reviewed past efforts at quantification in HIA) and chapter 7 (which provides a checklist on the validity of predictions in HIA), and proceed to answer the study question.

EU policy on fruits and vegetables

Most Europeans consume less fruits and vegetables than recommended by the World Health Organization (WHO 2003). The European Union Common Agricultural Policy (CAP) keeps the prices of fruits and vegetables above a pre-set minimum by limiting the availability of fruits and vegetables on the European market. Withdrawal subsidies are given to take (perfectly edible) fruits and vegetables off the market. Most of this 'excess' produce is destroyed

(Schäfer Elinder 2003). In Chapter 3 of this thesis we estimate that ending these withdrawal subsidies might increase the average consumption of fruits and vegetables by up to 1.80% (95% uncertainty interval 1.12–2.73), with a resulting increase in life expectancy of 3.8 (2.2–5.9) days for Dutch men and 2.6 (1.5–4.2) days for Dutch women. This equates with an annual gain of 1930 health-adjusted life years (DALYs) in the Dutch population. The reform is also likely to decrease socio-economic inequalities in health, albeit to a modest degree.

Key strengths of the study are that it is the first to quantify the potential health gain of reforming this part of the CAP and that it presents health outcomes at both the disease-specific and the generic level.

The study has several limitations. First, it gave only 'maximum effect-scenario' but no estimate of the most likely effect. Secondly, though uncertainty intervals are given, these reflect the uncertainty around this maximum effect, not around the most likely effect, and the uncertainty intervals only include the uncertainty in the amount of extra fruits and vegetables that would be available and the uncertainty in the health effects of fruits and vegetable consumption. A third limitation is that the data we used were somewhat dated and we implicitly assumed that the status quo would continue. In fact, recent data show that the amounts of fruits and vegetables withdrawn from the market were much lower in subsequent years. In other words, the model as we used it was not sufficiently robust to changes in the initial conditions. Fourthly, our analysis was limited to the effects on the Dutch population, while the total population of the European Union would benefit from the proposed reform. Fifth, no time horizon is reported. None of these limitations have bearing on the PIF/MSLT method; the difficulties lie in the assessment of exposure change.

Now that the withdrawal subsidies have been reduced, an upcoming issue for public health is the import barriers to fruits and vegetables from outside the EU. Lifting all those might have three times the effect on consumption that we reported as maximum effect for abolishing withdrawal subsidies (personal communication Frank Bunte, LEI / OECD, September 2005). This might annually avert about 5000 DALYs in the Dutch population and many times that number in the European population. This is a very rough estimate that deserves further investigation.

Tailored fruit and vegetable promotion

In a second case study we explored the use of epidemiological modelling for the estimation of health effects of behaviour change interventions, using the example of computer-tailored nutrition education aimed at fruit and vegetable consumption in the Netherlands. The effects of the intervention on changes in

consumption were obtained from an earlier evaluation study (Brug, Glanz et al. 1998); the effect of consumption change on population health was modelled. The results indicate that if the intervention reached the entire adult population and the effects were sustained, it could result in a mortality decrease of 0.4 to 0.7% and save 72 to 115 life years per 100,000 persons aged 25 or older. Health-adjusted life expectancy is estimated to increase by 32.7 days for men and 25.3 days for women. The true effect is likely to lie between this theoretical maximum and zero effect, depending mostly on the durability of behaviour change and reach of the intervention. Since these two factors are assumed to be linearly related to the health outcomes, estimates of the health effect can easily be generated for any expected level of participation and durability.

Like the study on the EU policy on fruits and vegetables, this study is limited to a maximum-effect scenario. However, it also presents a simple method to make more precise estimates once more information on participation rates and durability becomes available. As in the previous study, the major part of the uncertainty originates in the estimation of change in exposure, not in the PIF/MSLT method.

Prediction of the future course of the obesity epidemic in the US

The third case study, described in chapter 5, seeks to predict the future prevalence of obesity and overweight in the United States of America. It also provides a framework for assessment of the effects of population-targeted interventions that aim to lower the prevalence of obesity. Currently about 25% of men and 32% of women in the USA are obese. By extrapolating recent observed trends in average body mass index and assuming that BMI is always lognormally distributed in the US population, we estimate that the obesity prevalence will exceed 40% for men and 45% for women in 2015. A newly developed ‘population energy balance’ indicates that in order to stop the increase in obesity prevalence, men should walk about 2.2 extra minutes per day (95% uncertainty interval: 1.8 – 2.6) and women 3.2 (2.7- 3.6) minutes, under the condition that this is not compensated for by extra energy intake. The same result would be obtained if daily energy intake decreases by 9.1 (7.3 – 10.9) kcal per day for men and 11.2 (9.6 – 12.9) for women – less than a can of soft drink per week.

Limitations of this study include the assumptions of lognormality and the concept of a ‘population energy balance’, and the use of BMI as measure of obesity. The assumption that BMI is always lognormally distributed, at least in the general US population, is an approximation of reality. The data generally seem to support this assumption, although the method does underestimate the

prevalence of extreme obesity. The assumption of a population energy balance is also new. For individuals the energy balance is an accepted concept and its validity for populations seems plausible, but in the absence of sufficiently accurate methods to measure physical activity and energy intake it is impossible to validate this concept empirically. (Let alone to establish the predictive validity of the model in its entirety).

BMI is not a very good measure of adiposity at the individual level, and is an even worse measure of the amount of visceral abdominal fat, which seems to confer the most health risk. For trends in populations the use of BMI is likely to be more valid, because genetic composition and many relevant lifestyle characteristics can be assumed to be more stable. Still, it might be interesting to perform a similar exercise using waist circumference as measure of (intermediate) outcome, should sufficient data become available.

Judged against our criteria for HIA, two more remarks can be made. In the first place the interventions are not described in much detail. People should eat a few calories less or walk a few minutes more every day, but the true difficulty lies in how to bring about this change in behaviour. This will require societal action and the cooperation of policy makers, and perhaps HIA can facilitate this process (Bekker 2007). Secondly, the study presents no health outcomes but stops at the level of obesity prevalence, which is a (proximal) determinant. Adding a multi-state life table component could extend the analysis to the level of summary measures of population health. Lastly, the continuous lognormal distribution was translated into the rather crude categories of normal weight, overweight and obesity. Further refinement could be added by using a variant of the PIF formula that uses continuous distributions in combination with a continuous risk function.

Food advertising and childhood obesity

In our last case study we quantify the potential effects of measures that limit food advertising on television on the prevalence of overweight and obesity among 6 to 12-year old Dutch children (chapter 6). Based on available evidence we created a mathematical simulation model that links the average exposure to food commercials, via caloric consumption to body weight and the prevalence of overweight and obesity. In a second analysis we used a Delphi study to estimate the effect of television advertising on total caloric intake, instead of the meagre results of our literature search. We found that an effective ban on food advertising might reduce the prevalence of overweight and obesity among Dutch children by about 1.5%, from about 17% to 15.5%. Furthermore, this effect

might not materialise when restrictions apply only to advertising on television, while the same marketing budgets are used to reach children via other channels.

This study uses unorthodox methods and has to cope with a paucity of data, so the results should not be regarded as the final answer to the question. Most of the uncertainty relates to the plausibility of the theoretical framework, especially the order of magnitude of the effect of a change in exposure to advertising. Two parameters that probably contribute most to the total degree of uncertainty are the dose-response relation between advertising and total caloric intake, and the relationship between total caloric intake and body mass in children. In the baseline assessment only the statistical variance of the latter factor was quantified, and the rest of the uncertainty remained unknown. In a second assessment we performed a Delphi study to get experts' estimates on the dose-response relationship between advertising and total caloric intake. This also permitted the inclusion of the statistical variance in the experts' responses into the uncertainty interval. This leaves a considerable but unknown amount of uncertainty unaccounted for. On the positive side, the study provides more information than a general statement that there is likely to be an effect, and to some extent it shows where research can contribute most to the reduction of uncertainty. Again, the main difficulties lie in the estimation of exposure change. Adding the PIF/MSLT to extend the analysis to summary measures of population health would be straightforward.

Can a combination of potential impact fractions and multi-state life tables be used to predict changes in summary measures of population health due to policies outside the health sector?

The above case studies show that the potential impact fraction and the multi-state life table can indeed be used to predict changes in summary measures of population health (SMPH) due to some policies outside the health sector, at least in the analysed policies that impact on nutrition and obesity. Although the studies that focus on obesity did not use the MSLT, the model structure for doing so was present and with some additional effort and data the analysis could have been extended to include health outcomes in terms of SMPH. An important reason for not doing this was that the purpose of the work was to advise the ministry of health, which had formulated its policy goals in terms of obesity prevalence; extending the analysis would have had limited policy relevance. It would also have taken additional time and increased the complexity of the analysis beyond what can be comfortably described in a scientific paper.

The significance of the ability to use the PIF/MSLT method is that it enables health changes to be expressed in terms of summary measures of

population health, which makes the health effects of policy changes comparable across policy options and policy fields. A limitation that is evident from the case studies is that it is frequently difficult to quantify the effects that a policy will have on the exposure of populations. Furthermore, we only tested the method on policies that influence a single determinant of health for which reliable risk estimates are available. Methods for integrating the health effects of changes in multiple determinants are available, but reliable risk estimates are always essential.

How can the validity of predictions in HIA be established?

In part C we argued that an HIA can be considered valid if formal validity, plausibility and predictive validity are in order. Formal validity concerns how well an argument conforms to the rules of logic to arrive at a conclusion that must be true, assuming the premises are true. Plausibility is the degree to which an observer deems the theoretical framework is understandable, applicable and plausible. Predictive validity is the degree to which predictions are confirmed by facts. Both formal validity and plausibility can usually be established, but establishing predictive validity is more demanding. It requires outcome evaluation of HIA. This is seldom feasible due to long time lags, migration, measurement problems and a lack of data and sensitive indicators, and the fact that predictions may influence subsequent events ('Oedipus-effect'). Since predictive validity most often cannot be established in HIA, we have to rely on testing for formal validity and plausibility. This may seem like a meagre basis for conclusions, but in policy making it is well accepted that it is better to take decisions based on the best data available rather than to wait for the best imaginable data.

We further proposed a checklist that guides an assessment of the validity of prediction in an HIA study (p. 122-124). It consists of a series of questions that systematically guide the assessor through all aspects of formal validity, plausibility and predictive validity. Once completed it delivers an overview of the arguments underpinning the prediction and signals at what points validity is not ensured. The instrument can be used by the team that performs an HIA study to improve the assessment, or by an independent team to review the quality of an HIA study and its reporting.

Though it can structure an assessment of the validity of predictions in an HIA, a checklist cannot replace background knowledge. Checking formal validity requires knowledge of how methods ought to be applied, and checking

plausibility is an essentially subjective affair that gains in validity with the expertise of the assessor(s).

Reflections

The second question this thesis aims to answer is: Can a combination of potential impact fractions and multi-state life tables be used to predict changes in summary measures of population health due to policies outside the health sector? We have shown that this is possible with a few specific policies in the field of nutrition (fruits and vegetables) and obesity, but to what extent is this generalisable to HIA in other policy areas, and under what conditions would these models be useful?

Before answering that question, we must first conclude that none of the case studies fully qualifies as health impact assessment. They were all predictions, but not all up to the level of health outcomes, and most importantly, they did not all contribute directly to a policy decision. While we can draw some tentative conclusions about the possibilities to make valid, quantified predictions

Table 8.1 Overview of the case studies in this thesis and the most important scientific and policy circumstances that impacted on them.

	Methods	Scientific circumstances	Policy circumstances
EU agricultural policy (Ch. 3)	Policy analysis, multi-state life table	Maximum effect could be estimated. Collaboration with agricultural economist did not yield the desired link with econometric model.	Not HIA in a strict sense because no concrete policy process was informed
Computer-tailored HE (Ch. 4)	Previous evaluation, multi-state life table	Impact evaluation did not provide all input needed to enable estimation of health outcomes, but maximum effect could be estimated.	Extension of evaluation of a health promotion intervention rather than HIA. No direct link with a policy process.
US obesity (Ch. 5)	Trend analysis, deterministic model	Good data available	More a scenario study, not HIA in a strict sense. No direct link with a policy process.
Covenant Overweight and food advertising targeting children (Ch. 6)	Equilibrium model, Delphi panel	Paucity of evidence on several of the input parameters. Low response of panel.	No formal HIA process from the start. Initiative on our side instead of that of policy makers. Ministry waited for results and then decided not to act on them directly.

of health effects, we therefore have to be careful in drawing conclusions about the usefulness of these methods in actual HIA practice, where the policy circumstances may have more influence on the assessment.

We distinguish scientific conditions and policy conditions that influence the scope for quantitative health impact assessment (table 8.1). The scientific conditions involve the availability of the data and the methods necessary to assess the health consequences of the policy proposal at stake. The policy circumstances refer to the availability of time, money and other resources, but also to the possibilities to influence the decision making process and the possibility that the policy actors influence the scientific process.

Scientific conditions

The scientific conditions refer to the availability of the data and the methods necessary to quantify the health consequences of a policy proposal. When can the scientific conditions be considered fulfilled? And were they fulfilled in the case studies?

Scientific conditions for quantitative HIA

Little guidance is available on the minimum conditions for quantification in HIA. A UK discussion group provides some guidelines on what information is needed and mentions some difficulties that may be encountered (Mindell, Hansell et al. 2001). The group expresses a preference for a "broad model of health that integrates social, genetic and environmental factors, [which] is likely to identify impacts that are important but cannot be readily quantified." They argue that any assumptions need to be made explicit "rather than relying upon 'common sense' assumptions". The discussion group further recommends that when mathematical models are used, their predictions should be compared with empirical data, and the robustness of assumptions tested in sensitivity analyses. Lastly, when recommendations to maximize health benefit and mitigate risks are obvious as soon as the impacts are identified, time spent on quantification would be unnecessary and wasteful.

So the first question that must be answered is: in view of the decisions to be taken, is it worthwhile to invest in quantification? If it is worthwhile then two issues are important: the validity of the evidence and methods used, and the reporting of uncertainty.

Can general demands for evidence and methods for quantitative predictions in HIA be formulated? In principle, quantitative estimates can

always be made. Gaps in the evidence base can be filled with assumptions. However, when there is little empirical evidence and the uncertainty is large, there comes a point beyond which quantifying is not useful. Where this point lies also depends on whether the uncertainty can be quantified. Quantification may still confer more information than qualitative statements like 'there may be an effect' if the amount of uncertainty can be indicated (e.g. by a 95% uncertainty interval). A quantitative statement with an unknown degree of certainty merely indicates (or conceals) ignorance.

The minimum scientific conditions for quantitative HIA could therefore be (1) that methods and data enable a quantified estimate of the health impact of a proposed policy decision, and (2) that the uncertainty in the estimate can be quantified to an acceptable degree. The word 'acceptable' implies a judgement. This judgement should be made by the scientific community (De Groot 1969).

The general statement that methods and data must enable a quantified estimate of health impact and quantification of the uncertainty around that estimate can be divided into sub-conditions along the lines of the validity-checklist described in chapter 7. That is, there must be sufficient data to adequately describe the initial conditions (policy plan, current exposure pattern and health situation) and the theoretical framework must be quantifiable - both the part that leads from policy to exposure change and the part that connects exposure change to health outcomes. For the latter part the PIF/MSLT method may prove useful (as demonstrated in this thesis), while for the former the methods will depend on the characteristics of the policy and the determinants it influences.

Did the studies chapters 3 to 6 of this thesis meet these criteria? We tried to use the best available evidence and methods, discussed the most important assumptions we had to make, included a probabilistic uncertainty analysis that covered the main sources of (parameter) uncertainty, and generally tried not to overstate our case in the formulation of conclusions. From the fact that the studies presented in chapters 3, 4 and 5 of this thesis have been published in peer-reviewed journals we may conclude that they passed the first test of validity by the wider scientific community. (The study in chapter 6 has not yet reached that stage as it has only recently been submitted.) However, whether the evidence is strong enough to justify the amendment of policy decisions also depends on the nature of that policy decision, as we shall discuss below.

The effect of policy on exposure is a major source of uncertainty

The case studies show that the uncertainty in the effect of a policy decision on the exposure of populations to determinants of health is often much larger than the uncertainty in the effect of these changes in exposure on health outcomes. For the EU agricultural policy (Chapter 3) and a hypothetical national computer-tailored health education programme (Chapter 4) we could only present estimates of the maximum effect. For measures to limit children's exposure to food advertising (Chapter 6) we provided an estimate of the most likely effect, but it has a rather wide uncertainty interval.

Estimates of the effect of policy on determinants of health often concern issues that are not within the realm of the health sciences, such as how a pollutant disperses from a point source, how markets respond to changes in policies, or how much extra employment and income a development project yields. This necessitates collaboration with specialists from the relevant fields of science. It requires investment in finding the right specialist and, once a suitable party is found, investment in communication to overcome the barriers that frequently hamper interdisciplinary work. In our study of the European Union policy on fruits and vegetables (chapter 3) we worked together with an agricultural economist, but though this gave us valuable ideas and feedback, it did not result in a combined analysis using his econometric equilibrium model for the EU market for fruits and vegetables and our own model. It proved surprisingly difficult to tailor the output of the econometric model to what we needed as input for our MSLT model. In addition, the fact that we had no budget to offer in return and no clear policy process to advise is likely to have played a role. In the setting of a formal HIA both health scientists and economists should be involved from the start as part of the project team.

Not all problems can be solved by good intersectoral cooperation, however. It is inherently difficult to estimate the consequences of policy decisions. Policy decisions and the policy environment to which they apply are complex and cannot easily be standardised. For that reason, and because of their scale, they are seldom open to experimental research. This leaves observational research designs as the only empirical option, but these are less appropriate for establishing causality because of bias and confounding. By the standards used in medical (and pharmaceutical) research, this means that the effects of policy decisions can never be proven to an acceptable degree. The 'hierarchy of evidence' as applied to medical treatments is of little use in public health interventions (Petticrew and Roberts 2003), and even less in HIA where health effects are often unintended. Observational studies are often the best that is available, and sometimes there is little more than common sense to base

estimates on. In that case caution is warranted and quantification may not be wise (Mindell, Hansell et al. 2001), though others might argue that also the publication of informed "guesstimates" of health effects can help the field forward, provided that the assumptions on which they are based are very clearly stated (Kemm 2005). What evidence is acceptable may also depend on the policy context, as we shall discuss below.

What limits the validity of the studies in this thesis and in HIA in general?

The validity of predictions is inherently difficult to establish. To assess the validity of our work, we applied the validity checklist (chapter 7) to the studies in chapters 3 to 6. Of course, the exercise is limited by the fact that we judge our own work; a poor representation of the scientific community that, we argue, should make the judgement. The results are reported in an annex to this chapter. In general, they confirm the limitations we mentioned in the relevant chapters. The frameworks seem plausible (also to external reviewers, in the case of chapters 3, 4 and 5), and we found no reason to doubt their formal validity. However, the uncertainty in the magnitude of some of the causal relationships in the models is considerable, particularly in those that link policy to exposure (as discussed above). Furthermore, not all of that uncertainty could be expressed in quantitative terms (i.e., an uncertainty interval) as some of it originated in the nature of the research that produced the evidence. There were a few possibilities to test the historical predictive validity, but it was beyond our means to use them. We found little evidence about the effects of comparable interventions.

The application of the validity-checklist provides a few general observations about the current state of knowledge in the fields relevant to HIA. In the first place, there is a lack of empirical data in these areas. This limits the ability to predict effects in HIA, and to judge the validity of these predictions. HIA can only develop as a scientific discipline if more is invested in, particularly, the collection of data on the impact of non-health care policies on health. Secondly, there is not much previous research of similar interventions to compare the results with. The assessment of predictive validity would be aided by the creation of reconstructions of health consequences attributable to policy decisions in the past. An example of work in this field is a study on the health effects of urban regeneration programmes in the UK. A review of evaluations found little evidence of the impact of national urban regeneration investment on socioeconomic or health outcomes. Where impacts had been assessed, these were often small and positive but adverse impacts have also occurred (Thomson, Atkinson et al. 2006). Establishing causality in this kind of research is often difficult, but compared to prospective research, it has the advantage of saving

time (Chapter 7). However, prospective research is not always impossible. For example, a prospective controlled study with 100 participants showed no significant change in perceived health due to moving to new housing in an urban regeneration project (Thomson, Morrison et al. 2007). The rather poor state of scientific development of HIA limits the testing of validity, which in turn reflects negatively on the credibility of HIA as an instrument for policy makers. This vicious circle can only be broken by performing more HIA studies and systematically learning from the experience, for example in a European context. Fortunately, HIA has been (and still is) a priority area in the European Union's Public Health programme.

The place of the PIF/MSLT method in HIA

In the method used in this thesis, the potential impact fraction is pivotal: it connects the present situation to the situation after an intervention. The PIF requires input of a change in exposure to a determinant, and the relative risks that link the determinant to disease incidence (or mortality). Our case studies focused on the consumption of fruits and vegetables and on obesity, for which relative risk estimates are indeed available. The main challenge in terms of uncertainty was to estimate the effect of the interventions on exposure of the populations, but we succeeded in providing those. The full answer to our second study question is therefore probably: yes, a combination of potential impact fractions and multi-state life tables be used to predict changes in summary measures of population health due to policies outside the health sector, provided that reliable relative risk estimates for the determinants concerned are available, and that the effect of the policy on exposure can be quantified.

To what extent are these conditions satisfied in the HIA field? In chapter 2 we expressed the expectation that the degree of standardisation achieved in environmental HIA will be hard to match in HIA that focuses on other policy areas, because the health effects are context dependent. This leads to an inability to obtain the required reliable relative risk measures for the PIF calculation. This might be challenging in the case of determinants like employment, job strain, or education. More frequently, however, the use of the PIF/MSLT method will be hampered by the inability to quantify the effect of a policy proposal on exposure. Many policy plans do not sufficiently specify the implementation of any measures, either because they are the result of compromise, or to permit flexibility in their execution. Furthermore, they may impact on complex systems in which volition (human choices that have not yet been made; (Raskin, Banuri et al. 2002)) may play a role. In such cases, the use

of the PIF/MSLT method (or other quantitative methods) will be limited to the exploration of plausible scenarios.

The studies in this thesis are also not typical of HIA because they each focus on one determinant, while many policies influence multiple determinants. For example, a regional transport plan may result in changes in air quality, physical activity, traffic accidents, land use, economic opportunities for certain types of businesses (car-related, bikes, public transport, etc.), employment, gross economic growth, and so on. The PIF/MSLT method can be helpful in quantifying some of these impacts and validly add up the results, but it is unlikely to enable quantification of all health effects.

Further limitations to the PIF/MSLT method as we used it in this thesis are that it does not take account of demographic change and that time trends are not reckoned with. The MSLT cannot model time and cohort dimensions simultaneously (chapter 1). In our analysis, we ignored the factor time and modelled the 'steady state' that would arise if the population would be exposed to the present epidemiological situation for a generation or more (corrected for current population numbers), in combination with either the current exposure to the risk factor or an alternative exposure. If the size of the age cohorts in a population varies considerably and the time horizon is not short, this may influence the effect estimates. Significant trends in exposure or disease frequency may produce similar effects. Incorporating these effects into the analysis requires the inclusion of a time factor in the model. This can be done either by switching to a dynamic model, or by creating a separate variant of the MSLT model for every age group and subsequently adding the results. This enables inclusion of the effects of demographic change and time trends, as well as discounting (i.e. lowering the value of health effects that occur further into the future). Of course, this requires knowledge of past trends and assumptions about future trends, and complicates the analysis.

These limits do not imply that if prediction is not possible, the methods and models used in this thesis cannot be of use in HIA. If a policy is deemed likely to influence exposure to a risk factor, but the size of this impact cannot be estimated, scenarios can still confer an idea of what could be the quantitative consequences under different assumptions about what may happen. Examples in this thesis are the chapters on the EU common agricultural policy where our ignorance limited the analysis to the calculation of a maximum effect of reform, and the computer-tailored health education where we did not know how many people would participate and how long effects would last. In spite of this lack of data we could still estimate the maximum plausible health impact.

It takes time to prepare the model, so either the policy-process that the HIA is intended to inform should allow for that time or the model has to be prepared in advance. If suitable 'off-the-shelf' models are available, the HIA-analyst can focus on assessing the effect of the proposed policy on exposure to determinants. For HIA the following determinants might be most useful: physical activity, obesity, and small ambient particles (fine dust). These determinants have important population health effects and are influenced by many non-health care policies at the local, national and international level. In the ideal case, a model is available that already incorporates all the necessary information for the target population, including the exposure to the determinant in question and the disease patterns. This requires that the model is regularly updated. In the Dutch situation the RIVM Chronic Disease Model could provide that function if it were adapted for the purpose and made available to HIA analysts. If a model does not fully fit the population of interest, the HIA researcher may need to adapt the model to reflect the exposure and disease pattern of that population. In that case the model still provides a framework for the analysis and many elements can be used without adaptation.

The place of the Delphi method in predictions for HIA

In Chapter 6 we ran into the problem that no directly applicable information was available about the size of the effect of television food advertising on body mass in children. We did find one study that quantified the effect on caloric intake, but it was not completely clear how the results should be adapted to fit the actual Dutch situation. Rather than accepting the results of that one study without amendments or using our own interpretation of the study in our calculations, we consulted a panel of experts in a Delphi process. We selected 25 authors from recent papers on advertising and childhood obesity and 8 experts we personally knew to be active in that field, and asked them to fill in a questionnaire. The results were communicated back to the panel with the request to revise or restate their responses. The response was rather low and the possibility of bias could not be excluded. However, one would expect this bias to be resulting in an exaggeration of the effects of measures to reduce exposure of children to television advertising while our results indicated very modest effects.

Can the Delphi method be useful in HIA? Expert opinion does not rank high in the hierarchy of evidence. On classical evidence-based medicine lists it does not even feature, and in hierarchies that are designed to assess the strength of the evidence for the effectiveness of public health interventions it ranks lowest (Petticrew and Roberts 2003). Still, the estimates of a group of experts obtained in a structured process may be preferable to the guesses of a few

authors or the uncritical application of the results of only one study. In HIA as in obesity research, it is the best available evidence that should be used. Waiting for the best possible evidence is often the worse option (Swinburn, Gill et al. 2005). Given the paucity of data often encountered the Delphi method certainly has a place in predictions for HIA.

Health economic evaluation and HIA

A HIA that aims to quantify health effects comes to resemble the 'effect-side' of a health economic evaluation. Accordingly, quantitative HIA might use methods that have been developed for economic evaluation. In the following paragraphs we will discuss the use of guidelines and a 'reference case' scenario, the discounting of health effects, the 'value of information' literature, and how to integrate health and non-health effects of interventions.

While there are many different guidelines for HIA, economic evaluation is well ahead of HIA in terms of the development of guidelines on how to perform and report quantitative analyses (Drummond and McGuire 2001; Tan-Torres Edejer, Baltussen et al. 2003). Though most of the recommendations concern the costing side of the cost-effectiveness equation, the guidelines also deal with issues such as what effects to include, what outcome measures to report, and the need to perform subgroup analyses if differential effects are likely. Also potentially useful is the formulation of a 'reference case analysis' (Gold, Siegel et al. 1996). This is a standardised way of analysing and reporting an economic evaluation which should be performed with every study, alongside the main results that may be tailored to match the needs or views of a particular decision maker. The reference case enables comparison between studies. Both the guidelines and the reference case may be adapted for use in HIA.

In economic evaluation, both costs and health effects are discounted, usually at the same rate (a rate of 3-5% per annum is used most frequently). There are three possible reasons to discount health effects: (1) because a calamity may prevent one from enjoying a later benefit, or new technology makes the intervention obsolete; (2) because one simply prefers to enjoy a benefit earlier rather than later; and (3) because we get richer all the time and the benefit of extra wealth therefore diminishes with time (Tan-Torres Edejer, Baltussen et al. 2003). For the discounting of health effects the first two of these reasons also apply, but attaching less value to health gains in the future does not express much solidarity with future generations. So while discounting health seems the most logical option for individuals, it may make sense not to discount health for societies. Technical arguments that health should be discounted at the same rate as costs have been shown to be unrealistic (Tan-Torres Edejer,

Baltussen et al. 2003). In the context of HIA, discounting of health effects can be motivated by the first argument: uncertainty of effects increases with distance into the future. A preference for health gain now rather than later is more debatable. An argument against discounting future health effects can be found in the value of sustainability that many see to underpin HIA. A sustainable development is one that "meets the needs of the present without compromising the ability of future generations to meet their own needs." (United Nations (Brundtland Commission) 1987) Clearly, discounting future health does not sit comfortably in that framework. In sum, moral arguments can be made both for and against time discounting in HIA. If quantitative HIA grows, debate on this point should establish consensus, and this debate may well link in with the same debate in health economics.

In this thesis, we did not discount health effects. One of the reasons is that we aimed to develop a rather simple tool and limited ourselves to the estimation of the size of health effects. The MSLT has no time dimension. Models that do have a time dimension are more complex, require additional assumptions and are less transparent. This is a disadvantage in HIA, which takes place in interaction with decision makers, who will often lack prior knowledge about modelling. If implementation of an intervention is seriously considered and additional information is necessary, the next step would be to use a dynamic model, and to discount health effects if it has been decided to do so.

If the uncertainty about the best decision is large, it may be wise to postpone the decision and gather new evidence. A rather new development in health economic evaluation to assist in this kind of decisions is the assessment of the 'value of information'. This enables the identification of the kind of evidence that is most useful for reducing the probability of making a 'wrong' decision that will lead to the implementation of an intervention that falsely appears to be cost-effective. It also estimates the maximum amount of money that should be invested in such research (Briggs, Claxton et al. 2006). The application of the value of information methods requires a choice of 'threshold for cost-effectiveness' below which an intervention is deemed worth its cost. Moreover, the application of this approach to HIA requires that it incorporates *all* the uncertainty that pertains to a decision, not only the uncertainty in the health impact. It necessitates the use of a cost-benefit framework, in which health effects are translated to monetary terms. This will make the method unattractive to many.

Furthermore, a cost-benefit framework requires the application of equity-weighting and other kinds of weighting to the health gain, to incorporate all relevant values (Stolk, van Donselaar et al. 2004). The policy option with the

highest gain is then chosen. However, while this approach helps to make tradeoffs between policy options, it hides value judgments and differences in impact between population groups (Mindell, Hansell et al. 2001). Another option is to structure the balancing of different aspects and effects of a policy proposal in multi-criteria decision analysis (MCDA). In MCDA, decision makers assign weighting factors to aspects they value, after which the relative weight of different options is calculated. Again, the policy option with the highest gain is chosen (Baltussen and Niessen 2006). Multi-criteria decision analysis is flexible, explicitly involves decision makers, and does not require the transformation of health outcomes.

While economic evaluation of the whole proposal will seldom be feasible or useful, such an analysis of potential measures to mitigate harm or enhance benefit may be (Mindell, Hansell et al. 2001).

In summary, HIA could learn from the literature on economic evaluation for the development of guidelines for quantification of predictions in HIA, and for methods to weight different outcomes and values pertaining to a policy proposal. It may also refer to the economic literature when discussing the merits of discounting future health. The 'value of information' literature does not seem applicable to HIA at this point.

Policy conditions

As stated above, the policy circumstances refer to the availability of time, money and other resources, but also to the possibilities to influence the decision making process and the possibility that the policy actors influence the scientific process. This thesis has a counterpart in the administrative sciences that analyses the role of HIA in the policy process (Bekker 2007). Here we briefly present issues that impact directly on the scope for making predictions for HIA.

Health research can influence policy in several ways. Indirectly via the media and the opinions and actions of the general public, more directly by research councils providing reports at request, and most directly by health impact assessment, which tries to influence policy by presenting predictions of health consequences for a specific policy decision on a specific population.

Not all policy processes are open to the kind of direct information HIA provides. Policy making often is a complicated business in which many parties are trying to influence the outcome and in which the health aspect is only one of many. Uncertainty about the effects of different possible decisions on several of these aspects is often considerable. Policy makers generally try to reduce the complexities and uncertainties. In contrast, HIA risks adding more aspects and

stakeholders to the process - health aspects and health authorities, but also the public that is expected to experience the health consequences (Bekker 2007).

With that in mind, it is hardly surprising that we had difficulties in finding ongoing policy processes to advise. Only the study on food advertising and childhood obesity (Chapter 6) more or less directly informed policy makers, albeit not to a concrete policy decision. Several factors contributed to this difficulty. Firstly, we were looking for policies to advise, rather than being requested to perform a study, as would be the case in most HIA-settings. Understandably, population health not being among their responsibilities, some policy makers we contacted did not feel a pressing need to complicate matters by taking health considerations into account. Secondly, the duration of the project was limited and we could only deliver a one-time assessment, while policy makers expressed a need for longer-term consultation and commitment. Linked to that, we had no political or even technical responsibility for the health of the populations concerned. Such responsibility might have given us an argument to interfere with a policy process. And lastly, there is no legal basis for HIA so its inclusion in the policy process is dependent on the goodwill of the parties involved. HIA is not a well-known tool for most policy makers, and is viewed by some who do know it as an additional bureaucratic hurdle in an already complex policy process. A lack of goodwill may also have contributed to the fact that none of the studies in this thesis has all the features of HIA (table 10.1). When the research project was initiated the government was interested in HIA, but before the start of the project a new government had come to power. The new government clearly had different priorities.

A potential solution to the problem that policy makers who have no formal responsibility for health will often be reluctant to participate in HIA is to make the assessment mandatory, as has been done with environmental impact assessment. It would likely be difficult to define the type of decisions that would require HIA, and certainly in the beginning it does not guarantee that the assessments would be of good quality, but it would probably boost the development of methods, guidelines and quality criteria. This might be a topic worthy of study by the administrative sciences. Studies into the effectiveness of HIA as an instrument to improve population health can provide useful input (Bekker 2007).

The fact that HIA is so close to the policy process also has its potential problems. It aims to influence decisions, and this provides a stimulus for other stakeholders to try to influence the HIA. This may jeopardise the scientific validity of the study, but the most prominent risk is that it may limit the scope of the study and bias the presentation of results. Policy makers may try to increase

the amount of interaction with the policy process if the conclusions support their favoured outcome, or to limit this interaction in the case of unfavourable results. We noticed this in studies we conducted for the Ministry of Health, Welfare and Sports, which were the only studies in which we interacted directly with policy makers. The Ministry had taken the initiative to a 'Covenant Overweight' in which industry and other societal stakeholders would commit to measures to stem the increase in obesity prevalence, which was one of the policy goals of the government. We offered to estimate the obesity-related health effects of a limited number of interventions under consideration. The steering group formed by the Ministry showed little enthusiasm during the difficult early phase of the policy process. The representatives of the Ministry with the obesity policy in their portfolio were more interested, so we decided to consider these to be our primary target audience, hoping to reach the Covenant group at a later stage. The study on food advertising on television and childhood obesity (chapter 6) resulted from the contacts with the Ministry. At first, the Ministry did not want to be seen to ask for a study on this topic because it was deemed politically too sensitive. The results of a voluntary code of conduct of the industry first had to be awaited, and requesting research into the effectiveness of measures that would go beyond that code was not deemed prudent. The fact that we did study the topic resulted mostly from our own interest; had we been on the payroll of the Ministry, the study might not have happened. The preliminary results, which were based on the Bolton publication and much lower than the results based on the Delphi exercise, showed rather small effects of measures to limit food advertising on childhood obesity. By that time we still did not have certainty about whether we would have the opportunity to present the results to the members of the covenant process, and knowing the results, the Ministry did not see a reason to present them. There was no proposal about the topic under discussion, and the Ministry interpreted the results as to indicate that it was better to try to limit the time children watch TV – a topic we included in order to compare it with the effects of limiting food advertising. While logical in terms of the policy process, the fact that the outcomes of the study influenced its dissemination could be argued to amount to publication bias in scientific terms. Researchers often reserve the right to publish in scientific journals from the outset. Analogously, in HIA it might be advisable to agree before the start of the study on the dissemination of results.

While this interference with the research work may be a source of bias, on the other hand the policy process *should* influence the conduct of an HIA because the assessment has to provide answers to the questions policy makers and stakeholders have. This may have consequences for the scientific methods

and data that can be used, and it may mean that quantification of health effects is not possible. If a policy decision is highly contested, some of the stakeholders may try to discard the results of the HIA as invalid. A good example of this is the discussion around climate change. The results of the International Panel on Climate Change were (and still are) heavily contested by stakeholders who were reluctant to make sacrifices on conflicting values such as access to cheap fuel and short-term economic growth. This contestation may lead to more 'conservative' predictions and a more meticulous underpinning of the conclusions and recommendations than would otherwise be the case, and some methods and types of evidence (e.g. expert opinion) may not be acceptable. For example, one of the reasons the Ministry did not feel like presenting the results of the study on the effect of limiting food advertising on TV was the fear that a Delphi study would likely be criticised as an opinion poll with insufficient scientific validity (Bekker 2007).

This shows that the answer to the question when there is enough evidence to make a quantified prediction of health impact has to include an analysis of the policy situation. How much evidence is enough to justify a particular action depends on the policy environment and on the costs of the intervention (in financial, political or other terms).

Did the case studies presented in chapters 3 to 6 of this thesis meet the minimum policy criteria for quantitative HIA? This question can only be answered for chapter 6, which looked into the obesity-preventive effects of limiting food advertising to children. This was the only study that was directly discussed with policy makers. They judged that the study did not meet their standards, at least not those for putting a controversial proposal up for discussion with stakeholders.

Quantification in perspective

Not everything that can be counted counts, and not everything that counts can be counted (Albert Einstein). The impact of a policy on health inequalities, for example, is often hard to quantify, but should not be neglected. There is a risk that both researchers and their audience pay more attention to quantified effects than to effects that could not be quantified, which undermines the validity of the assessment. However, rather than concluding that quantification should not be practiced in HIA, it calls for the development of methods to quantify the presently unquantifiable. Provided all assumptions and uncertainties are clearly stated, science is better served by publishing new approaches to the quantification of health effects than by limiting itself to methods of which the

validity has already been proven (Kemm 2005). If the policy conditions do not leave enough room for that in the context of HIA, this development of methods should be taken as a specific subject of scientific inquiry.

Conclusions and recommendations

General conclusions

- Quantification is comparatively rare in HIA, and there is little uniformity between HIA studies in the conduct and reporting of predictions.

- The ability to make predictions about the effect of policy decisions on exposure to determinants of health is limited by a paucity of data and results of prior research.

- For determinants of which reliable risk estimates are available, 'off the shelf'-models could be created. This would permit to translate changes in exposure to changes in health outcomes, expressed in summary measures of population health. The PIF/MSLT models tested in this thesis are an example of such models.

- A HIA can be considered valid if formal validity, plausibility and predictive validity are in order. Formal validity and plausibility can be systematically assessed with the checklist presented in chapter 7 of this thesis. Predictive validity can rarely be assessed.

Recommendations for research and development

Define criteria for the conduct and reporting of predictions in HIA
Uniformity of HIA practice and reporting will facilitate comparison and the assessment of validity and so improve quality. This thesis presents a framework and a checklist for the establishment of the validity of the predictions of health effects. These should be evaluated in HIA practice. The thesis also examines past quantitative HIA studies against a set of criteria. It concludes that a wide variety of methods and reporting formats is used, which makes it difficult to compare the studies and judge their worth. As HIA matures as a research

discipline, pressure will rise to formulate criteria on its conduct and reporting, analogous to developments in health economics. The studies in this thesis might play a role in the definition of these criteria. HIA could also benefit from economic evaluation about how to integrate different values and outcomes in decision making, and about the (de)merits of discounting future effects.

Develop and standardise methods for quantification of effects of policy on exposure to determinants of health

This will often involve mathematical modelling and will need to be done for different policy fields separately, by different (non-health) disciplines such as environmental risk assessment, transport, marketing, etc. It is not so much health research that is needed here, but rather research in other scientific fields that is geared to deliver the outcomes that HIA requires. Furthermore, many useful methods may already exist but have never been linked to health. For example, for the study on the EU agricultural policy we came into contact with an agricultural economist who had the model that might have been useful with comparatively little effort.

Develop ready-for-use simulation models for the quantification of the effects of changing exposure on health outcomes

To facilitate quantitative HIA, ready-for-use models that include data on common risk factors in specified populations should be prepared. The PIF/MSLT method provides such models. On a European scale a project to develop suitable models and adapt them for use in different countries is underway. The models need to be filled with up-to-date epidemiological information, preferably on a scale that matches the scale of the policy processes that need to be advised. In the Netherlands the RIVM (the Dutch National Institute for Public Health and the Environment) Chronic Disease Model already contains data tailored to the Dutch situation for 18 determinants and their health consequences. The use of these data in combination with a model designed to suit the needs of HIA is likely to be an efficient way to stimulate quantitative HIA in the Netherlands.

Recommendations for policy

Stimulate interdisciplinary research into the effects of (past) policy decisions on determinants of health

The starting point of HIA is that policies outside the health care sector can have a major, often unintended, influence on health. To inform future policy (e.g. by means of HIA), evaluation of the health effects of (past) policy decisions is needed. Because health is not at the core of many health-relevant disciplines, this type of research to date does not receive the interest it merits.

Apply health impact assessment in practice

Authorities would do well to screen policies for potential amenable health consequences (or have others do this), commission health impact assessments where indicated, and evaluate their use. This will stimulate the development of methods and ensure policy relevance. An example of a policy field where HIA might usefully be applied is the EU Common Agricultural Policy.

Use simulation modelling to estimate the health gains of health promotion efforts

Their evaluation at the level of changes in exposure to determinants of health could be complemented with simulation modelling to assess their consequences at the level of health outcomes and life expectancy. In a further step, their cost-effectiveness could be estimated.

Monitor the obesity 'epidemic' with a population perspective in mind

Together with the growing evidence that trying to change the behaviour of individuals while leaving their environment intact meets with little success, this may help to provide support of much-needed environmental interventions.

References

Baltussen, R. and L. Niessen (2006). "Priority setting of health interventions: the need for multi-criteria decision analysis." Cost Eff Resour Alloc **4**: 14.

Bekker, M. P. M. (2007). Lost in translation. Reframing Health Impact Assessment for health policy integration. Institute of Health Policy and Management. Rotterdam, Erasmus University Rotterdam.

Briggs, A., K. Claxton, et al. (2006). Decision modelling for health economic evaluation. Oxford, Oxford University Press.

Brug, J., K. Glanz, et al. (1998). "The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake." Health Educ Behav **25**(4): 517-31.

De Groot, A. D. (1969). Methodology: foundations of inference and research in the behavioral sciences. The Hague, Mouton.

Drummond, M. E. and A. McGuire, Eds. (2001). Economic evaluation in health care: merging theory with practice. Oxford, Oxford University Press.

Gold, M. R., J. E. Siegel, et al., Eds. (1996). Cost-effectiveness in health and medicine. New York, Oxford University Press.

Kemm, J. (2005). "The future challenges for HIA." Environmental Impact Assessment **25**: 799-807.

Mindell, J., A. Hansell, et al. (2001). "What do we need for robust, quantitative health impact assessment?" J Public Health Med **23**(3): 173-8.

Øvretveit, J. (1998). Evaluating Health Interventions: an introduction to evaluation of health treatments, services, policies, and organizational interventions. Buckingham, Open University Press.

Petticrew, M. and H. Roberts (2003). "Evidence, hierarchies, and typologies: horses for courses." J Epidemiol Community Health **57**(7): 527-9.

Raskin, P., T. Banuri, et al. (2002). Great Transition - The Promise and Lure of the Times Ahead. SEI PoleStar Series. G. Initiative. Boston, Stockholm Environment Institute.

Schäfer Elinder, L. (2003). Public Health aspects of the EU Common Agricultural Policy. Developments and recommendations for change in four sectors: Fruit and vegetables, dairy, wine and tobacco. Stockholm, Swedish National Institute of Public Health.

Stolk, E. A., G. van Donselaar, et al. (2004). "Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall." Pharmacoeconomics **22**(17): 1097-107.

Swinburn, B., T. Gill, et al. (2005). "Obesity prevention: a proposed framework for translating evidence into action." Obes Rev **6**(1): 23-33.

Tan-Torres Edejer, T., R. Baltussen, et al., Eds. (2003). Making choices in health: WHO guide to cost-effectiveness analysis. Geneva, World Health Organization.

Thomson, H., R. Atkinson, et al. (2006). "Do urban regeneration programmes improve public health and reduce health inequalities? A synthesis of the evidence from UK policy and practice (1980-2004)." J Epidemiol Community Health **60**(2): 108-15.

Thomson, H., D. Morrison, et al. (2007). "The health impacts of housing-led regeneration: a prospective controlled study." J Epidemiol Community Health **61**(3): 211-4.

United Nations (Brundtland Commission) (1987). Report of the World Commission on environment and development (Brundtland Commission report). New York.

WHO (2003). Diet, Nutrition and the Prevention of Chronic Diseases. Report of a joint FAO/WHO Expert Consultation. WHO Technical Report Series. WHO. Geneva, World Health Organization.

ANNEX I

VALIDITY CHECKLIST APPLIED TO CHAPTERS 4, 5 AND 6.

(The validity of Chapter 3 has been assessed in the annex to Chapter 7.)

Chapter 4 – Implementation of a programme of computer-tailored health education with a focus on fruits and vegetables

Plausibility

The plausibility of this study is best assessed by a team that includes epidemiologists and economists specialised in international agricultural trade and econometrics.

Initial conditions

Is the policy plan / project described accurately ?	The intervention itself was described briefly, but a reference to more information was included. However, the chapter does not specify an implementation strategy, which results in an inability to estimate a 'most likely' effect of rolling out the intervention on a national scale.
Is the description of the baseline situation accurate?	The data on the consumption levels of FV were rather dated, but these change gradually and the influence on the results would likely be small.
Has uncertainty in the initial conditions been assessed?	No uncertainty around baseline estimates of FV consumption and frequency of disease. But the influence on the final uncertainty intervals should be modest.
How robust is the model to (foreseeable) changes in the initial conditions?	Quite; any changes are likely to be gradual and, therefore, small.

Theoretical framework

Is the causal web underlying the analysis valid according to the state of the pertaining scientific field?	The association between FV consumption and health is generally accepted. The influence of the programme was established by an RCT, so the evidence is quite strong.
Is the order of magnitude of the causal relations in concurrence with current scientific	The effects of changes in FV consumption on health as these were taken from recent reviews. RRs were not age-adjusted and applied

knowledge?	uniformly to all ages, which might lead to underestimation of disease frequency at lower ages and overestimation at higher ages, with the total effect remaining uncertain. The effect of the programme on FV consumption was only explored in a 'maximum effect' scenario. A more realistic scenario would have required a proposed implementation strategy and evidence of its likely participation rates. Furthermore, no assumption was made regarding the duration of the effect because of a lack of evidence on this point. On both fronts, the addition of a 'most likely' scenario might have been informative. This could be based on expert opinion if no stronger evidence was available.
Has the degree of certainty of the causal relations been described?	Two sources of uncertainty have been taken into account: the effect of the intervention (the 95% CI) and the relative risks of disease for changes in FV consumption. However, some additional uncertainty remains. The participants in the trial were not very representative of the Dutch population.
Are all exposures to determinants of health that are likely to result from the intended policy/project included in the analysis?	Other effects (e.g. on FV producers) are conceivable, but were estimated to be negligible compared to the effects on FV consumption.
Of the exposures included, have all plausible health outcomes been included?	We included CVD and cancer at a number of sites. For other diseases the evidence was deemed insufficient by the authors of the reviews we based the analysis on.
Have all populations likely to be affected by the policy been included in the analysis?	The analysis was applied to the general Dutch population. No subgroups were specified.
If available, how do the results of similar exercises compare with the predicted effects in this HIA? Can any differences be satisfactorily explained by differences in the initial conditions (including intervening events during the period of analysis) or lack of formal validity of the previous analyses?	To our knowledge no similar exercise has been conducted. Modelling work by the RIVM yielded slightly higher estimates of the lifetime lost due to sub-optimal consumption of FV, compared to our model (van Kreijl and Knaap 2004).

Formal validity *(verification)*

Formal validity could be assessed by the same team that assessed plausibility. This team would probably want to have the spreadsheets used for the analysis.

Initial conditions

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Description of policy proposal	Brief description with reference to further information. Seems to meet the needs.	Unlikely to be problematic.
Description baseline situation	Latest available data on FV consumption (TNO) and disease frequency (RIVM). Should be adequate.	Assessors could check the spreadsheet and the paper for inaccuracies in the numbers. Gross errors are unlikely, at least we received no comments after publication.

Theoretical framework

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Construction of causal framework	A single RCT is level I evidence. A multi-state life table approach was used to model the health effects of changes in FV consumption. The basic structure of this model had previously been published in a peer-reviewed paper.	Though it had limitations, the analysis was straightforward.
Estimation of magnitude of causal relations	For the effect of the intervention the results of the RCT were used, but no estimate was made of the participation rates or the duration of the effects. We conducted a PubMed search for recent meta-analyses or reviews on the relative risks of FV to disease, and contacted authors of older work if no result.	Perhaps other estimates for the RRs for cardiovascular disease would have been found if other authors had been approached.

Estimation of degree of certainty of causal relations	Uncertainty in the effect of policy on consumption could only partly be included; for the RRs confidence intervals in reviews were used (cancer), or the boundaries of likely values were interpreted as such with the consent of the authors (CVD). Bootstrapping was used to assess the overall uncertainty.	We detected no errors, but an independent check the procedures used and re-run of the bootstrap procedure would add credibility.
Search for significant determinants of health of which exposure changes as a result of the proposed policy	No formal search was conducted because substantial other effects did not seem plausible.	Not applicable.
Search for health outcomes that result from changes in exposure	Relied on reviews of the effect of FV consumption on CVD and cancer. Restricting inclusion to diseases with statistically significant relationship with FV may lead to underestimation of effect.	We have no reason to believe that the strategy was improperly executed.
Search for populations likely to be affected by the policy	No search; restricted to general Dutch population.	Not applicable.

Predictive validity

Historical predictive validity

Are historical data on initial conditions and subsequent outcomes available on which the model underlying the HIA can be tested?	This particular intervention has not been tested elsewhere. A comprehensive search might well reveal evaluations of similar interventions, but conducting this search was beyond our means.
If testing has been performed, how well does the model 'postdict' these outcomes, and can any differences between model and empirical data be explained satisfactorily by differences in the initial conditions or uncertainty in initial conditions (including intervening events during the period of analysis) and/or outcomes?	Not applicable.

In retrospect

To what extent did the predictions materialise?	Not applicable. And even if the proposed policy change would be effectuated, measuring its effect on population health outcomes would be impossible. It might be possible to measure changes in FV consumption, but causal attribution would not be easy.

Chapter 5 – Prediction of the US obesity prevalence in 2015 and the behaviour change needed to stop the increase in prevalence

This study has a different structure from the others in that it does not evaluate a specific intervention, and it also does not extend the analysis to health outcomes. Actually, it presents a baseline scenario for expected changes in obesity prevalence, and assesses a dose-response relationship between physical activity and energy consumption on the one hand, and obesity prevalence on the other.

Plausibility

The plausibility of this study is best assessed by a team that includes epidemiologists, nutritionists and specialists in metabolism and sports.

Initial conditions

Is the policy plan / project described accurately ?	The paper does not (aim to) describe an exact intervention, but extrapolates current developments into the future and assesses what (behavioural) change is necessary to stop the increase in the rise of obesity prevalence.
Is the description of the baseline situation accurate?	It is based on good data (NHANES studies) and looks plausible (also to external reviewers).
Has uncertainty in the initial conditions been assessed?	No uncertainty around baseline estimates of obesity prevalence. The uncertainty around the trend was arbitrarily estimated at 0.02 BMI-point/year. There might have been statistical tools to estimate this on the basis of the NHANES data, but for the purpose of the study this would require too large an investment.
How robust is the model to (foreseeable) changes in the initial conditions?	Not very robust. The crucial issue is whether the current trend in mean BMI per year would continue at its recent pace. We show that in recent years the pace has been accelerating, so our estimate may be

	an underestimate. But we might be looking at a peak in the trend. Since it is not well known what exactly drives this trend, it is impossible to account for it in the model.
Theoretical framework	
Is the causal web underlying the analysis valid according to the state of the pertaining scientific field?	The general idea that obesity prevalence changes with trends in energy intake and physical activity is well accepted. Our assumption that BMI is always lognormally distributed over a population (with its lower tail at a fixed point) may be more contentious, but any error here is mitigated by the fact that we partition up this continuous distribution in three groups (normal, overweight, obese). We included factors for the energy required for digestion, and for storage as body mass. In the literature, there is little doubt that not 100% of energy consumed can be transformed into body mass.
Is the order of magnitude of the causal relations in concurrence with current scientific knowledge?	For the metabolic calculations we used an authoritative (US) source. However, it seems a rather crude calculation and reality is probably more complex, with the energy balance depending on body mass or other factors. We further assumed that 10% of energy intake is spent in the digestive process and that it takes a further 10% to store energy as body weight. We found that the literature gives rather divergent estimates of these factors, and account for this by including uncertainty around the measure linking caloric misbalance to body mass (+/- 10%) in the model.
Has the degree of certainty of the causal relations been described?	Two sources of uncertainty have been taken into account: the trend in BMI (+/- 0.02 BMI-point/year), and the measure linking caloric misbalance to body mass (+/- 10%) in the model. There is additional uncertainty, however, in the behaviour of people. In particular, we assume a scenario in which people move more but do not eat more (and conversely, that they eat less but do not move less because of it). It is unclear how valid this is, but at least the assumptions are clearly stated.
Are all exposures to determinants of health that are likely to result from the intended policy/project included in the analysis?	(Apart from the fact that not one specific intervention as assessed:) No. Increasing physical activity levels is a determinant in itself, independent of obesity. But this is stated in the discussion.
Of the exposures included, have all plausible health outcomes been included?	No health outcomes have been included unless obesity is counted as health outcome (which we would not).

Have all populations likely to be affected by the policy been included in the analysis?	The total US population was included in the analysis, but no subgroups were modelled.
If available, how do the results of similar exercises compare with the predicted effects in this HIA? Can any differences be satisfactorily explained by differences in the initial conditions (including intervening events during the period of analysis) or lack of formal validity of the previous analyses?	We state a study that performed a similar exercise for older US citizens, and show that our model yields similar results, the difference likely being due to our use of a lognormal distribution (or their lack of it). We also compare our estimate of the amount of kcal change necessary to a previous (higher) estimate, and try to explain the difference.

Formal validity *(verification)*

Formal validity could be assessed by the same team that assessed plausibility. This team would probably want to have the spreadsheets used for the analysis.

Initial conditions

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Description of policy proposal	Not applicable.	Not applicable.
Description baseline situation	Based on authoritative data (NHANES).	We think so. An external check on our trend calculations could remove any doubt.

Theoretical framework

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Construction of causal framework	Rather new collation of methods and data, but we argue that it is a valid model.	We see no mistakes (and neither did the reviewers).

	Correct method	*Correct application*
Estimation of magnitude of causal relations	Authoritative source of metabolic calculations (US College of Sports Medicine). Limited literature search and assumptions for losses due to digestion and storage. This latter might have been further investigated, e.g. by formal consultation of experts.	We have no reason to believe otherwise.
Estimation of degree of certainty of causal relations	The uncertainty in both the trend in BMI (+/- 0.02 BMI-point/year) and the measure linking caloric misbalance to body mass (+/- 10%) were estimated rather arbitrarily. Better methods might be available. The bootstrap (probabilistic sensitivity analysis) is an accepted method.	We have no reason to believe otherwise.
Search for significant determinants of health of which exposure changes as a result of the proposed policy	We had no specific intervention and therefore no formal search was conducted.	Not applicable.
Search for health outcomes that result from changes in exposure	The paper did not aim to go to that level, so not applicable.	Not applicable.
Search for populations likely to be affected by the policy	No search; restricted to general Dutch population.	Not applicable.

Predictive validity

Historical predictive validity

Are historical data on initial conditions and subsequent outcomes available on which the model underlying the HIA can be tested?	Partly. For example, it might be possible to examine the correlation between results of our model with historical trends, and the past prevalence of overweight and obesity. It would also be possible to examine data on consumption and physical activity, but these are unlikely to be sufficiently precise and valid to permit validation of the model.

If testing has been performed, how well does the model 'postdict' these outcomes, and can any differences between model and empirical data be explained satisfactorily by differences in the initial conditions or uncertainty in initial conditions (including intervening events during the period of analysis) and/or outcomes?	Not applicable.

In retrospect

To what extent did the predictions materialise?	Not applicable.

Chapter 6 – Food advertising and childhood obesity

This study looks at the effect of limiting exposure of children to television advertising, but it does not specify a specific intervention. It also does not extend the analysis to health outcomes, but quantifies up to the intermediate level of obesity prevalence.

Plausibility

The plausibility of this study is best assessed by a team that includes epidemiologists and specialists in marketing.

Initial conditions

Is the policy plan / project described accurately ?	The paper does not (aim to) describe an exact intervention and assesses the maximum effect that can theoretically be reached by reducing exposure.
Is the description of the baseline situation accurate?	It is based on recent, reliable data on children's BMI (TNO-KvL). The data look plausible enough. A bit more uncertain is our fitting of them to a lognormal curve; this is not standard practice. However, this does not affect the baseline description but rather, the predictions of future prevalence. More uncertain is the estimate of exposure to TV food advertising, which is based on STER data of one month (October 2005). I would have expected a higher average level of exposure. It would have been nice to have more sources of corroborating estimates.

Has uncertainty in the initial conditions been assessed?	No uncertainty around baseline estimates of obesity prevalence, nor around current exposure.
How robust is the model to (foreseeable) changes in the initial conditions?	Quite robust. Mean BMI shifts gradually over the years, and so does exposure to TV advertising.

Theoretical framework

Is the causal web underlying the analysis valid according to the state of the pertaining scientific field?	It is. We cite a few recent, high-profile reviews that link food advertising to childhood obesity.
Is the order of magnitude of the causal relations in concurrence with current scientific knowledge?	As clearly stated, the uncertainty around the dose-response relationship is considerable. But it seems to be in keeping with what little current scientific knowledge there is.
Has the degree of certainty of the causal relations been described?	The literature (Bolton study) does not present a quantified estimate of the degree of uncertainty. This was one of the reasons to undertake a Delphi study, which did permit to give a quantified estimate of the degree of uncertainty. However, the response was rather low (24%) and the possibility of (self-)selection bias cannot be excluded. Whether the range in estimates of the experts is an adequate measure of the total uncertainty in the parameter is debatable, but we know of no good techniques to refine this.
Are all exposures to determinants of health that are likely to result from the intended policy/project included in the analysis?	We limited ourselves to this one. But concrete measures could also influence TV viewing in general, diet, or physical activity, depending on their nature.
Of the exposures included, have all plausible health outcomes been included?	No health outcomes have been included unless obesity is counted as health outcome (which we would not).
Have all populations likely to be affected by the policy been included in the analysis?	Dutch children aged 6-12 were included in the analysis, but no subgroups were distinguished.
If available, how do the results of similar exercises compare with the predicted effects in this HIA? Can any differences be satisfactorily explained by differences in the initial conditions (including intervening events during the period of analysis) or lack of formal validity of the previous analyses?	We cite an ecological study that finds that 'up to' half the rise in childhood obesity prevalence can be attributed to advertising. We give likely reasons for the different outcome (mainly that the other study fails to correct for confounding).

Formal validity *(verification)*

Formal validity could be assessed by the same team that assessed plausibility. This team would probably want to have the spreadsheets used for the analysis.

Initial conditions

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Description of policy proposal	Not applicable.	Not applicable.
Description baseline situation	Based on authoritative data (TNO) and data provided by STER. The validity of the latter is a bit doubtful, but it is the best estimate we had.	We fitted the BMI-data on a lognormal curve. While we think this was done accurately, an external review of the calculations might add strength to this supposition.

Theoretical framework

Have the right methods been applied and have these methods been applied correctly?

	Correct method	*Correct application*
Construction of causal framework	Rather new collation of methods and data, but we argue that it is a valid model.	We see no mistakes.
Estimation of magnitude of causal relations	The literature search was a correct instrument, but it yielded little. In that circumstance, a Delphi study seems an adequate way to get more evidence at short term.	We have no reason to believe the literature search was inadequately executed, though it might have been extended to grey literature with the help of marketing experts.
Estimation of degree of certainty of causal relations	No quantified indication of uncertainty was given by the literature. For the Delphi study we used non-parametric bootstrapping to incorporate uncertainty. This is the best we could do with these data, since this method does not pre-suppose a certain form of the functional distribution of the data.	We have no reason to believe the application was not correct. For the bootstrap of the Delphi data we used a new programme, Ersatz. The validity of this programme, while unlikely to be problematic, has not been externally assessed.

Search for significant determinants of health of which exposure changes as a result of the proposed policy	We had no specific intervention and therefore no formal search was conducted.	Not applicable.
Search for health outcomes that result from changes in exposure	The paper did not aim to go to that level, so not applicable.	Not applicable.
Search for populations likely to be affected by the policy	No search; restricted to general Dutch population.	Not applicable.

Predictive validity

Historical predictive validity

Are historical data on initial conditions and subsequent outcomes available on which the model underlying the HIA can be tested?	Historical estimates of the exposure of children to food advertising are available for some populations and periods, and so are estimates of childhood obesity. However, it would be difficult or impossible to correct for all confounders, given the current state of knowledge.
If testing has been performed, how well does the model 'postdict' these outcomes, and can any differences between model and empirical data be explained satisfactorily by differences in the initial conditions or uncertainty in initial conditions (including intervening events during the period of analysis) and/or outcomes?	No such testing has been attempted.

In retrospect

To what extent did the predictions materialise?	Not applicable.

Reference

van Kreijl, C. F. and A. G. A. C. Knaap, Eds. (2004). Ons eten gemeten. Gezonde voeding en veilig voedsel in Nederland. Bilthoven, RIVM.

Summary

Health impact assessment

Health impact assessment (HIA) aims to influence policy decisions outside the health field to protect and promote the health of the populations concerned. It does that by making predictions about the potential health consequences and communicating these to the decision makers and other stakeholders. HIA evolved in the nineties from environmental impact assessment and a stream in public health that stresses the importance of social determinants of health. The field is still very much in development and there is little agreement on methodological issues.

Most methods for HIA do share a core of five steps. In the *screening* phase policy proposals are judged by their potential to influence health and if so, if an HIA has any chance of influencing the decisions taken. If an HIA is deemed potentially useful, *scoping* determines how the study will be conducted, by whom, and which parties will be involved. The *effect analysis* estimates the nature and size of the health impacts, as well as its distribution over different groups in the population. Recommendations are drafted and, together with the results, *communicated* to the policymakers and stakeholders. Once a decision has been made, it may be decided to *monitor* developments in health or in exposure to determinants of health, and evaluation of the HIA may take place.

This thesis

This thesis focuses on how to make proper predictions, to be used in the effect analysis in HIA. The basic premise is that it is desirable to quantify the estimates of the potential health effects. Quantified estimates permit to distinguish important effects from trivial ones. Furthermore, adding up all positive and negative health effects into a net effect permits the use of economic instruments such as cost effectiveness analysis. This does require that all health effects are comparable. To this end, summary measures of population health (SMPH) have been developed. SMPH are measures that combine information on mortality and non-fatal health outcomes, such as health-adjusted life years (DALYs or QALYs) or health-adjusted life expectancy (e.g. DALE).

This thesis first reviews what methods have been used in quantitative HIA to date, then explores the use of a method based on the potential impact fraction and multi-state life tables, and finally looks at the validity of predictions in HIA.

What methods have been used in previous studies to make quantified predictions of the health impact of policy outside the health sector?

The results from the review of HIA case studies presented in **Chapter 2** of this thesis suggest that quantification is comparatively rare in HIA. Methods are available in the areas of environmental health and, to a lesser extent, traffic accidents, infectious diseases and behavioural factors.

The methods can be divided into two sets. First, there are methods that estimate the impact of a policy on exposure to determinants of health. These typically use data and measurements that are routinely available. Mathematical models are used for traffic related and environmental factors and infectious diseases. Where data are not available, experts' estimates and authors' assumptions are used.

The second set of methods relates changes in exposure to health outcomes. Epidemiological and toxicological risk assessment is used, sometimes as part of mathematical models. Quantification is mainly limited to proximal determinants of health, i.e., factors that influence health fairly directly.

The validity of the methods has not in all cases been clearly established, and the degree of certainty of the estimates was seldom quantified. Time horizons and outcome measures differed by method and study, and summary measures of population health were not used. To some extent the wide variety in methods and reporting reflects the differences in the policies that were evaluated and the needs of the policy makers, but it also points to a degree of immaturity of the field as a scientific enterprise and suggests methodological development can improve quality.

Can a combination of potential impact fractions and multi-state life tables be used to predict changes in summary measures of population health due to policies outside the health sector?

The effect of changes in the exposure of populations to determinants of health is calculated using the concept of the Potential Impact Fraction (PIF). PIF is defined as the proportional change in the incidence (or mortality) as a function of a change in exposure. PIF can be incorporated in mathematical simulation models. For the work in this thesis we used PIF in the 'proportional multi-state life table' (MSLT). This is a relatively uncomplicated model that can provide outcomes in both disease-specific measures (incidence, prevalence, mortality) and summary measures of population health. It is implemented in a spreadsheet. The addition of bootstrapping permits to estimate the overall uncertainty that

results from multiple elements in the model. The model has two dimensions, age and either cohort or period, and can be interpreted as a one-year cohort that ages with time, or alternatively as an entire population in a 'steady state' of exposure to determinants of health. In this thesis we took the latter perspective, which permits to estimate the size of the health effects but not when they will materialise. Dynamic models can combine the two perspectives but are more complex and require more data. The choice of the MSLT reflects a trade-off between accuracy and parsimony.

The PIF/MSLT method translates changes in exposure to changes in health, and can in principle be applied to a wide range of determinants. However, HIA also needs methods that answer the question of how a particular policy decision influences exposure. We did not pre-select methods for this part of the analysis.

To explore the potential of this method we conducted four case studies. **Chapter 3** focuses on the EU policy on fruits and vegetables (FV). Current policy guarantees producers a minimal price by subsidising the withdrawal of (perfectly edible) FV from the market. Most of this 'excess' produce is destroyed. We assessed the potential health effects of ending these withdrawal subsidies, taking the Dutch population as an example. In the most optimistic scenario, the quantities withdrawn in recent years would have been consumed. This would increase the average consumption of FV by up to 1.80% (95% uncertainty interval 1.12–2.73). We fitted data on the present consumption on a mathematical (Weibull) distribution and shifted the mean of this distribution 1.8% to the right. The PIF/MSLT method (with a bootstrap to estimate the combined uncertainty in the amounts of fruits and vegetables withdrawn/consumed extra and the protective effects of FV) then predicts an increase in life expectancy of 3.8 (2.2–5.9) days for Dutch men and 2.6 (1.5–4.2) days for Dutch women. This equates with an annual gain of about 1930 health-adjusted life years (DALYs) in the total Dutch population. The PIF/MSLT method was useful, but in the absence of better methods to estimate the effect of the policy change on FV consumption, we could only estimate the maximum effect of the policy change. We could not assess the most likely effect.

Chapter 4 explores the use of the PIF/MSLT method for the estimation of health effects of behaviour change interventions, using the example of computer-tailored nutrition education aimed at fruit and vegetable consumption in the Netherlands. The effects of the intervention on changes in consumption were obtained from an earlier evaluation study. The effect of consumption change on population health was modelled. The results indicate that if the intervention would reach the entire adult population and the effects were

sustained, it could result in a mortality decrease of 0.4 to 0.7% and save 72 to 115 life years per 100,000 persons aged 25 or older. Health-adjusted life expectancy is estimated to increase by 32.7 days for men and 25.3 days for women. However, this is obviously an overly optimistic scenario. The true effect is likely to lie between this theoretical maximum and zero effect, depending mostly on the reach of the intervention and the durability of behaviour change it achieves. Since these two factors are linearly related to the health outcomes, estimates of the health effect can easily be generated for any expected level of participation and durability. Like the study on the EU policy on fruits and vegetables, this study is limited to a maximum-effect scenario, and again this is due to missing information on the effects of the intervention on exposure (i.e. consumption of FV).

Chapter 5 seeks to predict the future prevalence of obesity and overweight in the United States of America, and estimates the average behaviour change that would end the rise in obesity prevalence. It also provides a framework for the assessment of the effects of population-targeted interventions that aim to lower the prevalence of obesity. Currently about 25% of men and 32% of women in the USA are obese. We show that body mass index (BMI) can be fitted reasonably well to a lognormal distribution across time, and how the entire distribution shifts with changes in the average BMI value. Assuming that this remains true in the future and that recent trends in average BMI continue, in 2015 the obesity prevalence will exceed 40% for men and 45% for women. Calculation of the energy balance indicates that in order to stop the increase in obesity prevalence, the average male US citizen should walk about 2.2 extra minutes per day (95% uncertainty interval (UI): 1.8 – 2.6) and the average woman 3.2 (2.7- 3.6) minutes, under the condition that this is not compensated for by extra energy intake. The same result would be obtained if daily energy intake decreases by 9.1 (7.3 – 10.9) kcal per day for men and 11.2 (9.6 – 12.9) for women – less than a can of soft drink per week.

In **Chapter 6** we quantify the potential effects of measures that limit food advertising on television on the prevalence of overweight and obesity among 6 to 12-year old Dutch children. Using the same framework described in chapter 5, we created a mathematical simulation model that links the average exposure to food commercials via caloric consumption to body weight and the prevalence of overweight and obesity. In view of the meagre results of our literature search, we performed a second analysis in which we used a Delphi study to obtain experts' estimates of the effect of television advertising on total caloric intake. We found that an effective ban on food advertising might reduce the prevalence of overweight and obesity among Dutch children by about 1.5

percentage-points, from about 17% to 15.5%. However, this effect might not materialise when restrictions apply only to advertising on television and the same marketing budgets are used to reach children via other channels. Though the results have to be interpreted with caution, they do suggest that limiting exposure to advertising alone is unlikely to have a large impact on the prevalence of childhood obesity.

The above case studies show that the potential impact fraction and the multi-state life table can indeed be used to predict changes in summary measures of population health (SMPH) due to some policies outside the health sector, at least in the analysed policies that impact on nutrition and obesity. Although the studies that focus on obesity did not use the MSLT, the model structure for doing so was present and with some additional effort and data the analysis could have been extended to include health outcomes in terms of SMPH. However, it was often hard to estimate the effects a policy will have on the exposure of populations. Furthermore, we only tested the method on policies that influence a single determinant of health of which reliable risk estimates are available. Methods for integrating the health effects of changes in multiple determinants are available, but good risk estimates may not be available for all determinants.

The significance of the ability to use the PIF/MSLT method is that it permits to express health changes in terms of summary measures of population health, which makes the health effects of policy changes comparable across policy options and policy fields.

How can the validity of predictions in HIA be established?

In **Chapter 7** we argue that an HIA can be considered valid if formal validity, plausibility and predictive validity are in order. Formal validity concerns how well an argument conforms to the rules of logic to arrive at a conclusion that must be true, assuming the premises are true. Plausibility is the degree to which an observer deems the theoretical framework is understandable, applicable, plausible and complete. Predictive validity is the degree to which predictions are confirmed by facts. Both formal validity and plausibility can usually be established, but establishing predictive validity is more demanding. It requires outcome evaluation of HIA. This is seldom feasible due to long time lags, migration, measurement problems and a lack of data and sensitive indicators, and the fact that predictions may influence subsequent events (‘Oedipus-effect’). Since predictive validity most often cannot be established in HIA, we have to rely on testing formal validity and plausibility. This may seem a meagre basis

for conclusions, but in policy making it is well accepted that it is better to take decisions based on the best available data rather than to wait for the best imaginable data.

Discussion

In **Chapter 8** the findings are summarised and discussed. We conclude that the possibilities for quantification in HIA depend on the state of the relevant scientific fields and the availability of data on the one hand, and on the policy conditions on the other. The political environment determines how much time and money is available, what policy options are up for discussion, and has an influence on what constitutes evidence. The focus in this thesis is on the scientific issues. We saw that quantification of health effects is rare in HIA practice. Most of the difficulty seems to be in the estimation of the effects of a policy on determinants of health. The methods and data needed are often outside the remit of the health sciences, which necessitates intersectoral cooperation. Standardisation of methods is possible only by policy area. Still, for some types of policy–determinant combinations methods do exist and data may be available, depending on the case. Where this is the case and effects on exposure to determinants are quantified, the analysis can often be carried on to health effects by using a combination of potential impact fraction and multi-state life tables, or similar methods. Using a 'comparative risk assessment'-like framework, the methods and data can be standardised across determinants and policy areas, and the health effects can be expressed in terms of summary measures of population health. This permits comparison between policy options, and between policies in different fields. The degree of certainty of the predicted health effects should also be quantified. For the uncertainty in parameters of the models this can be done using bootstrapping. Almost inevitably, in the course of the prediction process assumptions have to be made that are not comfortably supported by empirical evidence. This is acceptable if the assumptions are clearly reported. It increases the importance of sensitivity analyses and uncertainty estimation. Quantitative HIA might benefit from knowledge of economic evaluation, notably on how to value health outcomes (discounting, equity-weighting, decision analysis). The validity of predictions in HIA can be assessed by critically examining their formal validity, plausibility and predictive validity.

The main conclusions and recommendation from this thesis are as follows:

- Criteria for the conduct and reporting of predictions in HIA should be defined to facilitate comparison, judge validity and so improve quality. This thesis, in particular chapters 2 and 7, aims to contribute to this enterprise.

- Methods for the quantification of the effects of policies on exposure to determinants of health should be developed and standardised by policy or scientific field, in collaboration with experts of relevant non-health fields.

- Ready-for-use simulation models for the quantification of the effects of changing exposure on health outcomes should be developed for use in HIA. These models can be based on a combination of potential impact fraction calculations and the multi-state life table in some form. The data to fill the models should be systematically collected, ideally on a scale that matches the scale of the policy processes that need to be advised. At the Dutch national level, the data collection of the National Institute for Public Health and the Environment (RIVM) provides a good basis.

- Interdisciplinary research into the effects of (past) policy decisions on determinants of health can help to inform future health impact assessment studies.

- Authorities should screen their policies for potential amenable health consequences (or have others do this), commission health impact assessments where indicated, and evaluate their use. This will stimulate the development of methods and ensure policy relevance.

Samenvatting

Gezondheidseffectschatting

Gezondheidseffectschatting (GES) heeft als doel beslissingen over beleid buiten de gezondheidssector te beinvloeden om de gezondheid van de betrokken bevolking te beschermen en/of bevorderen. Het doet dat door het maken van voorspellingen over gezondheidseffecten en het communiceren daarvan naar beleidsmakers en andere belanghebbenden. GES is in de jaren '90 ontstaan uit milieu-effectschatting en een stroming in de publieke gezondheid die het belang van sociale determinanten van gezondheid benadrukt. Het veld is nog volop in ontwikkeling en er is weinig overeenstemming over methodologie.

De meeste methoden delen wel een kern van vijf stappen. In de *screening* fase worden beleidsvoorstellen beoordeeld op hun potentiële invloed op gezondheid en wordt gekeken of een GES het beleid zou kunnen beïnvloeden. Als een GES zinvol geacht wordt, bepaalt *scoping* hoe de studie wordt verricht, door wie, en welke partijen erbij worden betrokken. De *effect-analyse* schat de aard en omvang van de gezondheidseffecten, evenals de verdeling over verschillende populatiegroepen. Aanbevelingen worden opgesteld en samen met de resultaten *gecommuniceerd* naar beleidsmakers en belanghebbenden. Nadat een besluit is genomen kan worden besloten tot het *monitoren* van ontwikkelingen in gezondheid of in blootstelling aan determinanten van gezondheid, en evaluatie van de GES kan plaatsvinden.

Dit proefschrift

Dit proefschrift richt zich op het maken van valide voorspellingen voor gebruik in de effect-analyse van GES. Daarbij wordt ervan uitgegaan dat het wenselijk is om de potentiële gezondheidseffecten te kwantificeren. Gekwantificeerde schattingen maken het mogelijk om belangrijke effecten te onderscheiden van verwaarloosbare. Bovendien maakt het optellen van alle positieve en negatieve gezondheidseffecten tot een netto effect het mogelijk om economische instrumenten te gebruiken, zoals kosten-effectiviteitsschatting. Hiervoor moeten de gezondheidseffecten wel onderling direct vergelijkbaar zijn. Hiervoor zijn samengestelde maten van volksgezondheid (SMVG) ontwikkeld. SMVG combineren informatie over sterfte en ziekte in een maat. Voorbeelden zijn voor gezondheid gecorrigeerde levensjaren ('DALYs' of 'QALYs') en -levensverwachting ('DALE')

Dit proefschrift onderzoekt eerst welke methoden tot op heden gebruikt zijn in GES, probeert vervolgens een methode uit die is gebaseerd op de

'potentiële impact fractie' en meer-dimensionale sterftetafels, en kijkt tenslotte naar de validiteit van voorspellingen in GES.

Welke methoden zijn gebruikt om gekwantificeerde voorspellingen te maken in studies naar de gezondheidsgevolgen van beleid buiten de gezondheids-sector?
De resultaten van de review van GES studies in **hoofdstuk 2** suggereren dat kwantificering betrekkelijk zeldzaam is in GES. Methoden zijn beschikbaar op het gebied van milieu en gezondheid, en in mindere mate verkeersongevallen, infectieziekten en gedragsfactoren.

Twee groepen methoden kunnen worden onderscheiden. De eerste groep omvat methoden die de invloed van beleid op blootstelling aan determinanten van gezondheid schatten. Deze gebruiken meestal data en metingen die direct beschikbaar zijn. Wiskundige modellen worden gebruikt voor verkeersgerelateerde factoren, milieufactoren en infectieziekten. Waar gegevens ontbreken worden schattingen van experts en aannamen van de auteurs gebruikt.

De tweede set methoden verbindt veranderingen in blootstelling aan gezondheidsuitkomsten. Hierbij worden epidemiologische en toxicologische risicoschattingen gebruikt, soms als onderdeel van wiskundige modellen. Kwantificering is grotendeels beperkt tot proximale determinanten, ofwel factoren die een rechtstreekse relatie met gezondheid hebben.

De validiteit van deze methoden is niet in alle gevallen duidelijk vastgesteld, en de mate van onzekerheid in de schattingen zelden in maat en getal uitgedrukt. De tijdshorizon en uitkomstenmaten verschilden per studie en samengestelde maten van volksgezondheid werden niet gebruikt. Tot op zekere hoogte vloeit de variatie in methoden en rapportage voort uit de verschillen tussen de beleidsbeslissingen die onderwerp van studie waren en uit de behoeften van beleidsmakers, maar het wijst ook op een zekere mate van onvolwassenheid van GES als onderzoeksveld en suggereert dat verdere ontwikkeling van de methodologie de kwaliteit kan verbeteren.

Kan de combinatie van de potentiële impact fractie en meer-dimensionale sterftetafels gebruikt worden om de veranderingen in samengestelde maten van volksgezondheid als gevolg van beleid buiten de gezondheidssector te voorspellen?
Het effect van verandering in blootstelling van populaties aan determinanten van gezondheid wordt berekend met gebruik van het concept 'potentiële impact fractie' (PIF). PIF is gedefinieerd als de proportionele verandering in incidentie (of sterfte) als functie van een verandering in blootstelling. PIF kan worden

ingebouwd in wiskundige simulatiemodellen. Voor het werk in dit proefschrift gebruiken we PIF in de 'proportionele meer-dimensionale sterftetafel' (MSLT). Dit is een relatief eenvoudig model dat uitkomsten kan geven zowel op het niveau van ziekte-specifieke maten (incidentie, prevalentie, sterfte) als op het niveau van samengestelde maten van volksgezondheid. Het wordt geimplementeerd in een spreadsheet. De toevoeging van 'bootstrapping' maakt het mogelijk de totale onzekerheid in de uitkomsten in te schatten die het gevolg is van onzekerheid in verschillende parameters in het model. Het model heeft twee dimensies, leeftijd en cohort óf periode. Het kan worden geinterpreteerd als een cohort (van één jaar breed) dat met de tijd ouder wordt, of als een hele populatie in een stabiele situatie wat betreft de blootstelling aan determinanten van gezondheid. In dit proefschrift kiezen we het laatste perspectief, dat het wel mogelijk maakt om in te schatten hoe groot gezondheidseffecten zullen zijn, maar niet wanneer ze zich precies zullen voordoen. Dynamische modellen combineren de twee perspectieven maar zijn ingewikkelder en vereisen meer gegevens. De keuze voor de meer-dimensionale sterftetafel is derhalve het resultaat van een afweging tussen precisie en eenvoud.

De PIF/MSLT methode vertaalt veranderingen in blootstelling naar veranderingen in gezondheid, en kan in principe worden toegepast op een brede verzameling determinanten. Een GES vereist echter ook methoden die de vraag beantwoorden hoe een beleidsbeslissing die blootstelling verandert. Voor dit deel van de analyse hebben we vooraf geen keuze gemaakt voor een bepaalde methode.

Om de mogelijkheden van deze methode te verkennen verrichtten we vier studies. **Hoofdstuk 3** richt zich op het EU-beleid op het gebied van groenten en fruit (GF). Het huidige beleid garandeert producenten een minimum-prijs door het subsidiëren van het van de markt halen van 'overproductie'. Het meeste van deze (prima eetbare) GF wordt vernietigd. We probeerden de gezondheidseffecten in te schatten van het beeindigen van deze subsidie, en namen de Nederlandse bevolking daarbij als voorbeeld. In het meest positieve geval zou de hoeveelheid GF die in de voorafgaande jaren van de markt werd gehaald, zijn geconsumeerd. Dit zou de gemiddelde consumptie van GF hebben doen toenemen met maximaal 1,80% (95% betrouwbaarheidsinterval: 1,12 tot 2,73). We hebben gegevens over de huidige consumptie gefit op een wiskundige (Weibull) verdeling en verhoogden het gemiddelde van die verdeling met 1,8%. De PIF/MSLT methode (met een bootstrap om de gecombineerde onzekerheid in de extra hoeveelheid groente- en fruit en het beschermende effect van GF te schatten) voorspelt dan een toename in de levensverwachting van 3,8 (95% onzekerheidsinterval: 2,2 tot 5,9) dagen

voor Nederlandse mannen en 2,6 (1,5 tot 4,2) dagen voor vrouwen. Dit komt overeen met een winst van ongeveer 1930 voor gezondheid gecorrigeerde levensjaren (DALY's) over de totale Nederlandse bevolking. De PIF/MSLT methode was nuttig, maar zonder betere methode voor het schatten van de effecten van beleidsverandering op GF consumptie konden we alleen een maximum-effect schatten. De meest waarschijnlijke omvang van het effect konden we niet inschatten.

Hoofdstuk 4 verkent het gebruik van de PIF/MSLT methode voor het inschatten van de gezondheidseffecten van gedragsveranderende interventies, met als voorbeeld computer-ondersteunde gezondheidsvoorlichting bedoeld om de consumptie van groente en fruit onder de Nederlandse bevolking te verhogen. Het effect van de interventie op GF consumptie werd verkregen van een eerdere evaluatiestudie. Het effect van consumptieverandering op de volksgezondheid werd gemodelleerd. De resultaten laten zien dat als de interventie de gehele volwassen Nederlandse bevolking zou bereiken en de effecten blijvend zouden zijn, het zou kunnen resulteren in een daling van de sterfte van 0,4% tot 0,7% en het behoud van 72 tot 115 levensjaren per 100,000 personen van 25 jaar of ouder. De toename in voor gezondheid gecorrigeerde levensverwachting (DALE) wordt geschat op 33 dagen voor mannen en 25 dagen voor vrouwen. Dit is echter een overdreven optimistisch scenario. Het werkelijke effect ligt waarschijnlijk tussen deze schatting en helemaal geen effect, en is voornamelijk afhankelijk van het bereik van de interventie en de duurzaamheid van de effecten. Omdat deze twee factoren lineair samenhangen met de gezondheidsuitkomsten kunnen eenvoudig schattingen worden afgeleid voor elke veronderstelling van bereik en duurzaamheid. Net als de studie naar het effect van het EU landbouwbeleid beperkt deze studie zich tot een maximum effect-scenario, en wederom wordt dit veroorzaakt door het ontbreken van informatie over de effecten van de interventie op blootstelling (in dit geval de consumptie van groente en fruit).

Hoofdstuk 5 heeft als doel de toekomstige prevalentie van ernstig overgewicht (obesitas) en (matig) overgewicht in de Verenigde Staten te voorspellen, en in te schatten hoeveel de Amerikanen hun gedrag gemiddeld zouden moeten veranderen om de toename in het vóórkomen van overgewicht te stoppen. Het geeft ook een kader voor het inschatten van de effecten van populatie-gerichte interventies die pogen de prevalentie van overgewicht te verlagen. Momenteel is ongeveer 25% van de mannen en 32% van de vrouwen in de VS obees. We laten zien dat de verdeling van 'body mass index' (BMI) van de bevolking op verschillende tijdsmomenten redelijk lijkt op een lognormale verdeling, en dat die verandering door de tijd nagebootst kan worden

door het veranderen van het gemiddelde BMI. Aangenomen dat dit in de toekomst zo blijft en de huidige trends in het gemiddelde BMI zich voortzetten, zal in 2015 de prevalentie van obesitas meer dan 40% zijn voor mannen en 45% voor vrouwen. Berekening van de energiebalans geeft aan dat om de toename in de obesitas-prevalentie te stoppen, Amerikaanse mannen gemiddeld 2,2 minuten per dag extra zouden moeten lopen (95% onzekerheidsinterval: 1,8 tot 2,6) en vrouwen 3,2 (2,7 tot 3,6) minuten, onder voorwaarde dat dit niet wordt gecompenseerd door een verhoogde energie-consumptie. Hetzelfde resultaat wordt behaald als de gemiddelde dagelijkse energie-inname wordt verminderd met 9,1 (7,3 tot 10,9) kcal per dag voor mannen en 11,2 (9,6 tot 12,9) kcal/dag voor vrouwen – minder dan een blikje frisdrank per week.

In **hoofdstuk 6** kwantificeren we de potentiële effecten van maatregelen ter beperking van de blootstelling aan televisie-reclame voor voedingsmiddelen op overgewicht bij Nederlandse kinderen van 6 tot 12 jaar. Gebruikmakend van het in hoofdstuk 5 beschreven kader maakten we een simulatiemodel dat de gemiddelde blootstelling aan voedingsmiddelenreclame via energie-consumptie relateert aan lichaamsgewicht en de prevalentie van overgewicht en obesitas. Omdat de literatuur weinig houvast bood voerden we een Delphi studie uit waarin we experts om een inschatting vroegen van het effect van televisiereclame op de totale calorische inname, en voerden we een tweede analyse uit waarin gebruik werd gemaakt van de resultaten hiervan. We vonden dat een effectief verbod op voedingsmiddelenreclame de prevalentie van overgewicht en obesitas onder Nederlandse kinderen zou kunnen verminderen met 1,5 procentpunt, van ongeveer 17% tot 15,5%. Dit effect zou waarschijnlijk niet worden gehaald als een verbod zich beperkt tot adverteren via de televisie, omdat hetzelfde marketingbudget zou kunnen worden gebruikt om kinderen via andere kanalen te bereiken. Hoewel deze resultaten met enige voorzichtigheid moeten worden geinterpreteerd, suggereren ze dat het onwaarschijnlijk is dat een beperking van blootstelling aan televisie reclame als geisoleerde maatregel een grote invloed zou hebben op overgewicht bij kinderen.

De hierboven beschreven studies laten zien dat de potentiële impact fractie en de meer-dimensionale sterftetafel inderdaad kunnen worden gebruikt om veranderingen te schatten in samengestelde maten van volksgezondheid ten gevolge van beleidsmaatregelen buiten de gezondheidssector, in ieder geval voor de onderzochte beleidskwesties die voeding en overgewicht beïnvloeden. Hoewel de studies over overgewicht de meerdimensionale sterftetafel niet gebruikten was de benodigde modelstructuur hiervoor wel aanwezig en kon de analyse met wat extra inspanning en gegevens worden uitgebreid tot gezondheidsuitkomsten in termen van SMPH. Het was echter vaak moeilijk om

de effecten van beleid op blootstelling van populaties in te schatten. Bovendien probeerden we de methode alleen uit op beleidskwesties die een enkele determinant beïnvloeden waarvoor betrouwbare risicoschattingen beschikbaar zijn. Er zijn methoden voor het integreren van effecten via meerdere determinanten, maar mogelijk zijn goede risicoschattingen niet voor alle determinanten beschikbaar.

Het belang van de mogelijkheid de PIF/MSLT methode te gebruiken is dat het gezondheidsveranderingen uitdrukt in termen van samengestelde maten van volksgezondheid. Dit maakt gezondheidseffecten vergelijkbaar tussen beleidsopties en beleidsgebieden.

Hoe kan de validiteit van de voorspellingen in GES worden vastgesteld?

In **hoofdstuk 7** beargumenteren we dat een GES studie als valide kan worden beschouwd als de formele validiteit, de plausibiliteit en de predictieve validiteit in orde zijn. Formele validiteit richt zich op het bepalen in hoeverre de argumentatie voldoet aan de regels van de logica en uitkomt op een conclusie die wel waar moet zijn als de onderliggende veronderstellingen waar zijn. Plausibiliteit is de mate waarin een onafhankelijke beoordelaar het theoretische raamwerk beschouwt als begijpelijk, toepasbaar, plausibel en compleet. Predictieve validiteit is de mate waarin de voorspellingen zijn bevestigd door feiten. Zowel formele validiteit als plausibiliteit kunnen meestal worden vastgesteld, maar het vaststellen van predictieve validiteit is moeilijker. Het vereist uitkomstevaluatie van GES, en dit is zelden te realiseren omdat het lang duurt voordat effecten zich openbaren, door migratie, meetproblemen en het ontbreken van gegevens en goede indicatoren, en tenslotte doordat voorspellingen de navolgende gebeurtenissen kunnen beinvloeden (‘Oedipus-effect’). Aangezien predictieve validiteit meestal niet kan worden vastgesteld in GES, moeten we genoegen nemen met het testen van formele validiteit en plausibiliteit. Dat lijkt misschien een magere basis voor conclusies, maar bij het maken van beleid is het een geaccepteerd gegeven dat het beter is beslissingen te nemen op basis van de beste beschikbare gegevens dan te wachten op de best denkbare.

Discussie

In **hoofdstuk 8** worden de bevindingen opgesomd en tegen het licht gehouden. We trekken de conclusie dat de mogelijkheden voor het kwantificeren van gezondheidseffecten in GES afhangen van de staat van het relevante

onderzoeksgebied en de beschikbaarheid van data aan de ene kant, en van de beleidscondities aan de andere kant. De politieke omgeving bepaalt hoeveel tijd en geld er beschikbaar is, welke beleidsopties ter discussie staan, en heeft invloed op wat als 'bewijs' kan worden beschouwd. Dit proefschrift richt zich echter op de wetenschappelijke kwesties. We hebben gezien dat kwantificering van gezondheidseffecten niet vaak voorkomt in GES. Het moeilijkst lijkt daarbij het inschatten van de effecten van beleid op determinanten van gezondheid. De benodigde methoden en gegevens liggen vaak buiten het gebied van de gezondheidswetenschappen, wat noopt tot intersectorale samenwerking. Standaardisering van methoden is slechts mogelijk binnen een beleidsgebied. Toch zal voor sommige combinaties van beleid en determinanten de vereiste methodologie bestaan, en soms zullen ook de benodigde gegevens beschikbaar zijn. Waar dit het geval is en beleidseffecten tot het niveau van blootstelling aan determinanten zijn gekwantificeerd, kan de analyse tot gezondheidseffecten worden doorgetrokken met behulp van de combinatie van de potentiële impact fractie en meer-dimensionale sterftetafels of soortgelijke methoden. Gebruikmakend van een 'comparative risk assessment'-achtig kader kunnen de methoden en gegevens worden gestandaardiseerd voor verschillende determinanten en beleidsgebieden, en de gezondheidseffecten worden uitgedrukt in samengestelde maten van volksgezondheid. Hierdoor kunnen de effecten worden vergeleken tussen beleidsopties en tussen maatregelen op verschillende beleidsgebieden. De mate van zekerheid van de voorspelde gezondheidseffecten zou ook moeten worden uitgedrukt in maat en getal. Voor de onzekerheid in de parameters van modellen kan dit worden gedaan door 'bootstrapping'. Bijna onvermijdelijk moeten er in de loop van het predictie-proces aannamen worden gedaan die niet comfortabel worden ondersteund door empirische bewijzen. Dit is acceptabel zolang die aannamen duidelijk worden vermeld, en het vergroot het belang van een grondige sensitiviteitsanalyse en het inschatten van de totale mate van onzekerheid. Kwantitatieve GES zou kunnen leren van economische evaluatie, met name wat betreft waardeoordelen over de gezondheidseffecten (discounting, sociale ongelijkheid, methoden voor besluitvorming). De validiteit van de voorspellingen kan worden ingeschat door een kritische beschouwing van de formele validiteit, de plausibiliteit en de predictieve validiteit.

De belangrijkste conclusies en aanbevelingen uit dit proefschrift luiden als volgt:

- Criteria voor het uitvoeren en rapporteren van voorspellingen in GES zouden moeten worden geformuleerd om vergelijking en het beoordelen

van de validiteit te vergemakkelijken. Dit proefschrift, met name de hoofdstukken 2 en 7, beoogt daaraan een bijdrage te leveren.

- Methoden voor het kwantificeren van de effecten van beleid op blootstelling aan determinanten van gezondheid zouden moeten worden ontwikkeld en gestandaardiseerd per beleids- of wetenschappelijk gebied, in samenwerking met experts op betrokken velden buiten de gezondheidswetenschappen.

- Voor gebruik in GES zouden 'kant-en-klare' simulatiemodellen moeten worden ontwikkeld waarmee de effecten van een veranderde blootstelling kunnen worden vertaald in veranderingen in gezondheidsstatus. Deze modellen kunnen worden gebaseerd op een combinatie van de potentiële impact fractie en de meer-dimensionale sterftetafel in de een of andere vorm. De gegevens waarmee het model wordt gevuld zouden volgens een vaste systematiek moeten worden verzameld, idealiter op een schaal die aansluit bij de beleidsprocessen die van advies worden voorzien. Op het Nederlandse nationale niveau biedt de gegevensverzamelijk van het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) een goede basis.

- Interdisciplinair onderzoek naar de effecten van beleidsbeslissingen in het verleden kunnen toekomstige GES studies van bewijsmateriaal voorzien.

- Beleidsinstanties zouden hun maatregelen moeten screenen op mogelijke veranderbare gezondheidseffecten (of zorgen dat anderen dit doen), opdracht geven tot GES studies waar dit geindiceerd is, en het gebruik van die studies evalueren. Dit zal de ontwikkeling van methoden stimuleren en ervoor zorgen dat die methoden relevant zijn voor beleid.

Dankwoord

Als schrijver van een proefschrift voelde ik me een soort quasi-zelfstandig ondernemer. Ik heb genoten van de vrijheid mijn tijd zelf in te delen, van de vrijheid het onderzoek vorm te geven, en de vrijheid om van het interessante werk van anderen kennis te nemen. Ik heb het bij MGZ, en daarbuiten, prima naar mijn zin gehad. Er zijn veel mensen die ik daarvoor graag wil bedanken. Ik noem er een aantal in het bijzonder.

Johan Mackenbach, ik heb genoten van je ongeëvenaarde analytische scherpte en je vermogen de grote lijnen snel te ontwaren. Je wist de begeleiding van ons projectje te combineren met het leiden van een afdeling, en bleef daarbij betrokken en open voor discussie (mits met goede argumenten: voorbereiding essentieel). En die discussie was er genoeg, want het was niet altijd gemakkelijk de richting van het onderzoek te bepalen. Dankzij jou ligt er nu toch dit proefschrift.

Jan Barendregt, je modelleerwerk was natuurlijk de basis van het project, maar je expertise en inbreng beperkten zich daar zeker niet toe. Uitstapjes naar de economie en wetenschapsfilosofie waren ook geen probleem. Natuurlijk was het niet zo aardig van je om halverwege naar Australie te verdwijnen, zelfs al bleef je bij het project betrokken. Maar dat is inmiddels rechtgetrokken, onze werkplekken liggen nu weer minder dan tien meter van elkaar. Over de jaren ben ik toch een beetje een volgeling van je geworden, dat mag duidelijk zijn.

Ed van Beeck, heel veel dank voor de begeleiding bij het laatste stuk van het traject. Bij al je kundigheid, kennis en werkdruk houd je toch een geweldige bescheidenheid en een oprechte belangstelling voor anderen. Voor mij ben je een model-academicus.

Marleen, het was leuk om samen op te werken met een initiatiefrijk persoon als jij. Zo konden we ervaren dat intersectoraal samenwerken best lastig kan zijn, zeker in combinatie met het maken van een proefschrift, en onze frustraties daarover delen. De bestuurskundige inbreng van jou, Tom en Kim voegde een belangrijke en interessante dimensie aan het werk toe. Ik heb genoten van onze discussies. En het het loopt ook allemaal goed af: we promoveren op dezelfde dag. Van harte!

Het werk aan dit proefschrift voerde me regelmatig naar de lommerrijke percelen van het RIVM. Ik vond er een prachtige natuurlijke en menselijke omgeving. Henriette, Hans, Fons en andere VTV-ers: bedankt voor de leuke weken data-spitten en nuttige gesprekken, waaraan ik zeker zoveel kennissen heb overgehouden als kennis. Rudolf, dank je voor de feedback en de cijfers. Het was erg bemoedigend te zien hoe modellen kunnen uitgroeien tot een beleidsadviserend instrument als het Chronische Ziekten Model. Wanda, het was prettig om een stukje samen op te gaan bij mijn pogingen tot beleidsadvisering

over overgewicht. Lea, Gerard, Manon, en Jantine, ik heb met genoegen met jullie van gedachten gewisseld over wat dat GES nou is of zou moeten zijn tijdens jullie GES bijeenkomsten en internationale congressen.
Hendriek en Jet (en natuurlijk Johan), het was erg stimulerend en leerzaam met jullie een gezamenlijk idee uit te werken tot een Europees onderzoeksvoorstel, en dat nog gefinancierd te krijgen ook. Het kostte veel te veel tijd, maar uiteindelijk hebben we ook de laatste bureaucratische horden weten te nemen. En toen vertrok ik naar de andere kant van de wereld. Maar aan de plezierige vooruitzichten in Nederland heeft het niet gelegen.
Hans Brug, het was een genoegen met je samen te werken en ik verheug me erop je weer te zien tijdens de promotieplechtigheid, nu eens in een andere toga. Jaap Seidell, dank voor alle input over overgewicht, ondanks je vele verplichtingen.
Josine van den Bogaard, dank je voor je praktische kijk vanuit de GGD. Zonder je meedenken over een lokale casus om ons instrumentarium op te testen was het nooit gelukt die te vinden. Met helaas ook niet, maar de zoektocht leverde toch een mooi kijkje in Rotterdam op, naast het inzicht dat (ons soort?) onderzoek niet zomaar in elk beleidsproces inpasbaar is. Ook bedankt voor de feedback op ons handboek dat geen handboek was.
Frank Bunte, dank voor het kijkje in de wondere wereld van de econometrische modellen. Jammer dat het koppelen van de modellen er niet van kwam. Dat blijft een uitdaging voor toekomstige onderzoekers.
En dan zijn er nog Margje Jansen van TNO en de risico's van groenten en fruit, Gerda Feunekes van Unilever die ons op het spoor van het Convenant Overgewicht zette, en Karin van Gorp, Carin Cuijpers, Ria Westendorp en anderen van het ministerie van VWS die bereid waren onze plannen, resultaten en adviezen aan te horen en te toetsen op bruikbaarheid in de praktijk.
Dichter bij huis was ik blij met Oscar, Michelle, Egil en, op het nippertje, Farizah als kamergenoten. Wilma, Helene, Esther, Debby, Egil en vele anderen, ik zou graag weer eens van die bizarre lunchgesprekken voeren. En gelukkig waren er Kees, Peter en ooit Ton die de computers onder controle hielden, en Else, Sonja, Jolanda, Mirela, Yvonne en ook Ilse die de bureaucratie bestierden. Wilma Nusselder, veel succes als laaste van de TAM-ers. We houden contact! Tot slot Egil, geweldig dat je me bij wilt staan bij de verdediging van dit proefschrift, twee dagen voor je bruiloft. Alvast gefeliciteerd!
Wilma, Dik, Gerard, en Sake, bedankt dat jullie me hebben willen betrekken in de 'health impact assessment' van het Afrikaanse onchocerciasis-bestrijdingsprogramma. Het was een uitdaging die we gezamenlijk aangingen, zonodig tot middernacht of later. Het onderzoek was nieuw en interessant, en het voerde me niet alleen naar zingende nonnen in Ouagadougou en warme stranden

bij Dar, maar hield me ook langer bij MGZ. Ik ben benieuwd naar de resultaten van de vervolgstudie!
Hoewel ze er nog weinig van gezien hebben, is dit proefschrift natuurlijk mede mogelijk gemaakt door mijn ouders. Pa, nou begin ik toch verdorie weer iets meer op je te lijken. Verdere vrienden en familie hebben met het proefschrift weinig te maken (en met mij de deze jaren ook niet zoveel als ik wel zou willen), maar ik wil ze toch bedanken voor de kleur die ze aan mijn leven geven. En Eva, dank je voor je steun bij de verdediging van dit werkstuk.
Zo rond het verschijnen van mijn eerste artikel heb ik eens gezegd ik dat ik beter was in het maken van kinderen dan wetenschappelijke publicaties. Dat denk ik nog steeds, maar ik ben blij dat de publicaties het nu in numeriek opzicht winnen. Hugo, Ernst, Koen, ik kan geen genoeg van jullie krijgen. Linda, je maakt me elke dag blij met je aanwezigheid.

Curriculum Vitae

Lennert Veerman was born on 8 January 1971 in Amsterdam. In 1989 he obtained his secondary school diploma at the Hermann Wesselink College in Amstelveen. He spent a year on various jobs in Paris and Amsterdam before taking up his medical studies at the Free University Amsterdam. During his studies he was active in the International Federation of Medical Students' Associations (IFMSA), a voluntary organisation of medical students worldwide, which he presided in 1995-96. He spent part of his internships at a district hospital in Lesotho, southern Africa, and completed the clinical part of his medical studies cum laude in 1999. After a few months of clinical work at the department of internal medicine of the Spaarne hospital in Haarlem, he started his career in Public Health. From 1999 to 2002 he worked as a public health physician in the care for asylum seekers in the region of Amsterdam. During this period he completed four courses at the Faculty of Medical Anthropology at the University of Amsterdam, lectured on the medical care for refugees at the Pharos Institute and started the course at the Netherlands School for Public and Occupational Health (NSPOH) that would lead to his registration as a public health physician in 2004. From 2002 to 2007 Lennert worked at the department of Public Health, Erasmus MC, which resulted in this thesis and in the completion of a Masters of Public Health at the Netherlands Institute for Health Sciences (NIHES) in 2004. In April 2007 he took up a position as research fellow at the Centre for Burden of Disease and Cost-effectiveness of the School of Population Health, University of Queensland, Brisbane, Australia. Lennert is married to Linda Robertus. They have three sons: Hugo (2003), Ernst (2004) and Koen (2006).

Publications

Veerman JL, Mackenbach JP. [The importance of the National Public Health Compass for public health] Ned Tijdschr Geneeskd. 2005 Jan 29;149(5):224-5. Dutch.

Veerman JL, Barendregt JJ, Mackenbach JP. Quantitative health impact assessment: current practice and future directions. J Epidemiol Community Health. 2005 May;59(5):361-70.

Veerman JL, Barendregt JJ, Mackenbach JP. The European Common Agricultural Policy on fruits and vegetables: exploring potential health gain of reform. European Journal of Public Health, 2006 Feb;16(1):31-5.

Veerman JL, Barendregt JJ, Mackenbach JP, Brug J. Using epidemiological models to estimate the health effects of diet behaviour change: the example of tailored fruit and vegetable promotion. Public Health Nutrition, 2006;9(4):415-20.

Veerman JL, Bekker MPM, Mackenbach JP. Advocacy and HIA: a challenging combination. Soz Praeventiv Med, 2006;51:151-2.

Veerman JL, Barendregt JJ, Van Beeck EF, Mackenbach JP. Validity in health impact assessment. J Epidemiol Community Health 2007 Apr;61(4):362-6

Veerman JL, Barendregt JJ, Van Beeck EF, Seidell JC, Mackenbach JP. Stemming the obesity epidemic: a tantalizing prospect. Obesity. (in press)

Mackenbach JP, Veerman JL, Barendregt JJ, Kunst A. Health inequalities in HIA. In: Health Impact Assessment. Eds: Kemm J, Parry J, Palmer P. Oxford, OUP 2004. Pp. 25-38.

Veerman JL, Van Beeck EF, Barendregt JJ, Mackenbach JP. Childhood obesity: the potential effectiveness of limiting food advertising on television. (manuscript)

The Netherlands Organisation for Health Research and Development financially supported the research described in this thesis (ZonMw grant number 2100.0083).

This thesis was printed with financial support of the J.E. Jurriaanse Stichting, the Erasmus University, and the Department of Public Health, Erasmus MC, Rotterdam. The Erasmus University Trustfund financially supported the participation in several congresses.