Valuing Prevention Through Economic Evaluation
Some Considerations Regarding the Choice of Discount Model for Health Effects with Focus on Infectious Diseases

Jasper M. Bos,1 Philippe Beutels,2,3,4 Lieven Annemans5,6 and Maarten J. Postma1

1 Groningen University Institute for Drug Exploration, University of Groningen, Groningen, The Netherlands
2 Centre for the Evaluation of Vaccination, Epidemiology and Community Medicine, University of Antwerp, Antwerp, Belgium
3 Centre for Health Economics Research and Evaluation, University of Technology, Sydney, Australia
4 National Centre for Immunisation Research and Surveillance, Royal Alexandra Children’s Hospital, University of Sydney, Sydney, Australia
5 Department of Public Health, Ghent University, Ghent, Belgium
6 IMS Health, Health Economics and Outcomes Research, Brussels, Belgium

Abstract

In cost-effectiveness analysis, the valuing of costs and health effects over time remains a controversial issue. The debate mostly focuses on whether the discount rates for health and money should be equal and which discounting model and time preferences are most appropriate. In this paper we add to the debate by arguing that the assessment of effectiveness of a preventive intervention may influence the choice of the discounting procedure for health.

Health effects in cost-effectiveness analysis are commonly expressed in life-years gained, QALYs gained or lives saved. These denominators are only indirect and partial measures of the effects of a preventive intervention. The actual effect of the intervention is a reduction of the risk of mortality and morbidity in a given period of time. This risk reduction will not always coincide with the moment at which the impact on (quality-adjusted) life-years gained is made (i.e. at risk exposure), for example when preventing chronic disease with an asymptomatic stage. In this paper we show that truly acknowledging the origin of health benefits could have implications for the discounting procedure. We present a discounting model that adequately focuses on the reduction of risk. This model recognises the potential interpretation of risk reduction for infection as an economic good to be acquired with associated mortality reductions as later indirect effects. This implies that our suggested discounting model focuses on the moment(s) of risk reduction. A numerical example illustrates our approach. We discuss the associated potential implications for public health policy and discuss how the effects of the intervention can be additionally corrected for societal preferences.
Cost-effectiveness studies of medical interventions analyse the aggregated monetary and health effects over pre-specified periods of time. The proper valuing of costs and health effects arising at different points in time remains controversial in health economics. Discounting is performed in order to adjust future monetary and health effects for their differential timing.[11] It is common practice to discount future costs and effects at an equal rate, i.e. the social rate of time preference. Most commonly, this social rate of time preference is derived from the interest rate on long-term government bonds. Despite guidelines supporting this practice, there is an ongoing discussion on discounting, in particular with respect to health effects.[2-12]

Two main reasons exist for discounting, namely decreasing marginal utility of consumption (of both monetary and non-monetary commodities) and time preference.[5] Decreasing marginal utility is caused by future (economic) growth, through which the relative value of commodities decrease. For instance, an increase in salary of $100 now has a higher relative value than the same increase in the future if the salary has already doubled in the meantime. A second component of discounting is time preference. This preference for current instead of future utility has two main components; namely pure time preference and an uncertainty effect. Pure time preference reflects a form of impatience, whereas the uncertainty effect refers to the uncertainty about future life-years and the risk of not being able to consume them or the bare uncertainty about the existence of a future society.[1,5]

It is clear that both decreasing marginal utility and time preference exist for monetary costs and benefits. However, it seems unlikely that the marginal utility of health decreases at the same pace as that for money.[3,9] For instance, the marginal utility of health is unlikely to be significantly affected by the growth in life expectancy from a single intervention, simply because rises in life expectancy are usually small. Also, health and money are not logically one-to-one related and a constant relation between future health and economic growth is unlikely.[3,13] Thus, decreasing marginal utility of money should not form the de facto basis for discounting health and money at the same rate.

As stated, the social rate of time preference is derived from the market rate of interest for financial assets. The market rate of interest reflects the rate at which consumers are willing to trade present for future consumption. According to the consumer sovereignty principle, consumer preferences should dictate government policy.[17] So the social rate of time preference should equal the market rate of interest. This implies that time preference for all commodities, including health and money, is essentially the same. However, this reasoning is not undisputed. In particular, it has been argued that individuals exhibit different time preferences for different goods, implying that the rate of time preference for health might differ from that for money.[4] As a consequence, no single social rate of time preference would exist, but a set of time preferences for commodities of different nature. Empirical studies seem to confirm this by observing that individuals exhibit different time preferences for health and money.[8,14]

The recommendations of most guidelines to discount both health and money at the same rate are based on the arguments listed in the classic papers by Keeler and Cretin[15] and Weinstein and Stason.[16] These papers indicate that inconsistencies, such as the paralysing paradox and shifts in the cost-effectiveness ratio over time, could appear if different discount rates for costs and health effects are applied. In particular, for example, the paralysing paradox indicates that discounting health at a lower rate than money (or even not at all) results in a preference to delay healthcare programmes as delay improves the cost-effectiveness ratio in that situation. A number of authors share the opinion that these consistency arguments alone do not justify the use of identical or even similar discount rates.[2-6,9]

Therefore, in order to account for time preference it would seem appropriate to acknowledge that society exhibits different time preferences for different commodities and to select different discount rates for each. In particular, from the societal perspective, it has been suggested that the discount rate for health

© 2004 Adis Data Information BV. All rights reserved. Pharmacoeconomics 2004; 22 (18)
should be lower than that applied to money.[2] The use of differential discounting is not limited to health economics only, it has also been proposed, for instance, in the literature on environmental and climate economics. Hasselmann et al.[17] argued that the possible differences in societal time preference alone should justify that discount rates be regarded as independent and potentially varying between different commodities.

In this paper we add to the debate by arguing that the assessment of effectiveness of a preventive intervention influences the choice of the discounting procedure for health. We first examine the nature and timing of the health benefits resulting from preventive interventions and its relevance for the debate on discounting. Next, we will question the value of the discounted utility model for health benefits arising from a preventive intervention and exemplify these issues. In section 4, we will derive a discounting procedure for health based on the occurrence of risk reduction. In our discussion, we will outline the potential implications of using this methodology and suggest how societal preferences can be incorporated into a discounting procedure based on the concept of risk reduction.

1. Quantifying the Health Effects of a Preventive Intervention

An important issue in cost-effectiveness analysis is the measure that is used to quantify the effects of a medical intervention. Health effects are commonly expressed in a number of denominators, such as life-years gained, number of QALYs, deaths or cases averted. An important, often not recognised, issue is that all these denominators are indirect measures of the real effects of the intervention, being reductions of mortality and morbidity risks in the period of exposure to the disease. Eventually this reduction will be visible as a rise in life expectancy of the target population and can be translated into life-years gained or other denominators for cost-effectiveness analysis.

The occurrence of the reduction in the risk function of acquiring morbidity or mortality from a certain disease does not always coincide with the moment when the effects of the intervention are manifested in terms of increased life expectancy, nor do they necessarily coincide with the start of the intervention. For instance, hepatitis B virus (HBV) vaccination at a relatively young age may lead to the prevention of infection some time later in life. For non-vaccinated individuals, infection with HBV may lead to symptomatic infection (for the largest part also many years later). From that moment on there may be a reduction in the QOL and an increase in the risk of dying from ensuing chronic liver disease, again many years later. Hence, it is not surprising that the choice of discount rate and discount model are amongst the most influential analytical decisions in economic evaluation of HBV vaccination.[18,19]

Measuring the effects of the intervention as a risk reduction during the period of exposure from a given disease could have major implications for designing an adequate discounting procedure. If we consider, in economic terms, the reduction in risk of acquiring disease (and related morbidity or mortality) as a good that is used for a given period of time, it seems logical to discount health effects arising from this good according to the period of time in which the good is used. So, if we consider the timing of risk aversion the basis for our discounting procedure for health, the total value of the health effects gained by the intervention should be corrected only for the difference in timing between occurrence of risk reduction and the timing of the intervention. In short, this is the main rationale for the proposed approach to discounting in this paper.

Let us again consider HBV vaccination of newborns. HBV infection is a disease that is often contracted via sexual intercourse, particularly in non-endemic countries. The result of vaccination is a reduction of mortality or morbidity due to a risk reduction of acquiring infection at the time of exposure to HBV, which will not occur before the age of onset of sexual activity. On average, this will be 1–2 decades after vaccination. The risk reduction arising from this intervention should be corrected for this time lapse of 1–2 decades. If we consider such an approach based on risk reduction, how should we
value the total reduction in mortality and morbidity resulting from the intervention?

2. The Utility Model for Health Gains

The most common method to value health effects (often life-years and/or quality gained) from an intervention is to value them by using a utility model based on expected utility theory. However, it has been argued that the descriptive value of expected utility is not adequate for the assessment of utilities in medical decision analysis.[20] Numerous examples exist of the fact that individuals’ preferences systematically violate expected utility theory (see below).[21,22] For a full review on utility models for health, we refer the reader to Miyamoto.[20]

Moreover, the most common utility model for health is characterised by linearity and additivity over time. These assumptions imply that the utility attached to a life-year gained is independent of the timing of occurrence and that the utility of survival is additive over time. In other words, the value of a sequence of outcomes corresponds to the sum of the discounted values of the component parts and the order of the sequence has no influence on the total utility.[23] This is certainly not in line with empirical research that highlights the importance of the order of the sequence of outcomes, for example with respect to increasing or decreasing pain experiences.[23,24]

Besides these well documented cases where individuals’ preferences are not adequately addressed by expected utility theory, several other normative and empirical arguments exist that contradict with the use of a discounting model for health consequences based on expected utility theory. The use of life-years gained as an outcome measure can never be seen without acknowledging the context, which is that they constitute a part of a life saved and should not be viewed as independent items. Life-years are not easily traded through time. One may adopt a healthy lifestyle to increase the number and quality of one’s own future life-years, but we will not be able to consciously trade a particular life-year for another one. The strong link between individual future life-years, where a life-year 10 years from now implies that the other 9 years are also consumed, is an indication that a discounting model for health outcomes based on expected utility theory (for the remainder of the paper we refer to this model as the discounted utility model for health) might not be adequate for the valuation of health outcomes over time.

Empirical research suggests that time preferences for health differ from those for consumer goods. One may question whether time preference for health is caused by the same mechanism as time preference for ordinary consumer goods. Moreover, discount rates for health are likely to differ for different types of health benefits, so a single rate for time preference for health may not exist.[9] In our view, time preference for health might be caused by a different mechanism to that for other consumer goods, as the single entity of a life-year cannot be viewed without implicitly acknowledging the rest of the life that it constitutes. Therefore, applying the expected utility theory to health outcomes expressed as life-years might fail to adequately grasp the peculiarities of health as an economic good.

3. Incorporating Societal Preferences in the Analysis

Previously, we proposed that the current utility model may not be suitable for health. However, not using this discounted utility model for health could give rise to practical problems. For instance, not discounting future life-years does not reflect the uncertainty about the bare existence of future society. Although this uncertainty is likely to be small for the near future, it might become an important factor when interventions with benefits in the distant future are evaluated. Also, not using the discounted utility model for health implies that the value of health is neutral in respect to age. This might not be in line with societal preferences, such as those related to equity. While equity is not related to discounting, preferences towards equity should also be incorporated in an analysis performed from the societal perspective. Society might value saving the life of an infant higher than that of an adult, simply because the former has not yet lived its share (this is often
referred to as the ‘fair innings argument’).[25] Furthermore, priority might be given to saving someone in immediate need over prevention.

Disease-specific preferences could also exist through distorted perceptions. For instance, the societal preference for preventing meningococcal C infections in children might be greater than the preference for avoiding chickenpox in children, while the total potential burden of disease of the latter is likely to be similar, and possibly even larger. Partly, these preferences will be caused by severity or nature of the disease. Another form of preference might be towards the prevention of diseases that lead to a diminished QOL in the future. If the aim is to adequately value the risk reduction of an intervention, all types of preference should be elicited and incorporated in the analysis. This applies to time-related, equity-based and other preferences. Clearly, at this moment, this is far from being achievable in a consistent and representative way.

Elucidation and incorporation of such different types of preferences could be established by using frameworks from philosophy, for instance Cohen[9] suggested the use of a form of the healthcare budget allocation framework by Kamm. Kamm’s framework for the allocation of scarce resources, such as organs needed for transplantation, distinguishes between for instance preferences according to age and need.[26] For an overview of other alternative frameworks, we refer the reader to the article by Petrou and Wolstenholme.[27] A different point of view would be to elicit preferences for risk reductions by using the willingness-to-pay (WTP) methodology and expressing the efficiency of the intervention in a cost-benefit analysis, as is common practice in environmental economics. The use of a dedicated framework for the valuation of the risk reduction has the advantage that all components of preference, and even process utility could be accounted for. In the following section of this paper we first present a discounting model for health based on the occurrence of risk reduction and later on we discuss possible implications for public health policy.

4. Discounting Health Gains: an Application

In section 1 we showed that the effects of a preventive intervention might be expressed as a risk reduction on acquiring disease during the period of exposure. This has implications for the discounting procedure, and the health effects resulting from such a risk reduction should be corrected for the time between the intervention and the occurrence (‘consumption’) of the risk reduction. The value attached to this risk reduction (essentially a reduced risk of dying later on from the disease) may be expressed as a total value attached to lives saved at a certain age, for instance using the remaining life expectancy for this valuation.

Since consumer preferences for the purchase of prevention are likely to follow similar patterns as for the purchase of consumer goods,[28] we might use the social rate of time preference to correct for the time at which the risk reduction occurs. However, as indicated in section 3, all societal preferences (time-related and others) should be incorporated into the analysis. This might be accomplished by using a correction function to weigh the value attached to the risk reduction with the societal preference towards the risk reduction.

So, the total value attached to risk reduction would be a composite of remaining life expectancy attached to surviving beyond a certain age and a correction factor for societal preferences. This procedure is essentially an extension of the method used by Jönsson et al.[29] and advocated by Johannesson.[9]

We will now illustrate this method using a hypothetical birth cohort which is vaccinated at time t = 0 against an infection. We compare the cost effectiveness of intervention B (the vaccine) to the current situation of no vaccination against the disease (intervention A) [refer to all equations in figure 1].

According to standard methodology, the incremental cost-effectiveness ratio (iCER) of intervention B over A is shown in equation 1.

Both costs and effects are discounted using the social rate of time preference r; the value for r is usually between 1% and 8%, most often 3–5%. The
Fig. 1. Equations 1–5. \( C_A(t) \) = costs of intervention A at time \( t \) [do nothing]; \( C_B(t) \) = costs of intervention B at time \( t \) [vaccination at \( t = 0 \)]; \( E_B(t) \) and \( E_A(t) \) = effects of B and A at time \( t \) [to be expressed in life-years]; \( iCER \) = incremental cost-effectiveness ratio; \( N \) = population size; \( p \) = risk reduction; \( r \) = the social rate of time preference; \( RLE(t) \) = remaining life expectancy at age \( t \); \( R(t) \) = initial risk of infection-related mortality at time \( t \); \( S(t) \) = probability to survive to age \( t \); \( t_i \) = timing of occurrence of risk reduction. A detailed hypothetical example of this discounting procedure is given in appendix I.

reduction in the risk of infection of intervention B versus A is expressed here as a function \( p \) (for simplicity we assume here that this is constant through time). This risk reduction of infection (and the associated mortality) \( p \), for the birth cohort \( N \) is timing between intervention and the occurrence of risk reduction \( ti \) with the social rate of time preference \( r \). In mathematical terms this is shown in equation 2.

In line with our argument presented in the previous sections, this outcome can be discounted for the timing between intervention and the occurrence of risk reduction \( ti \) with the social rate of time preference \( r \). This risk reduction of infection and the associated mortality \( p \) might be estimated from standard life tables, ideally excluding the risk from dying with the infection under consideration. Consequently, this is multiplied by the initial risk of infection-related mortality \( R(t) \).

Equation 1:

\[
iCER = \sum_{t=0}^{\infty} \frac{[C_B(t) - C_A(t)]} {S(t)} \cdot (1 + r)^{-t}
\]

Equation 2:

\[
iCER = \frac{\sum_{t=0}^{\infty} [C_B(t) - C_A(t)] \cdot (1 + r)^{-t}} {\sum_{t=0}^{\infty} [N \cdot S(t) \cdot R(t) \cdot p \cdot \sum_{t=0}^{\infty} (1 + r)^{-t}] \cdot (1 + r)^{-t}}
\]

Equation 3:

\[
iCER = \frac{\sum_{t=0}^{\infty} [C_B(t) - C_A(t)] \cdot (1 + r)^{-t}} {\sum_{t=0}^{\infty} [N \cdot S(t) \cdot R(t) \cdot p \cdot \sum_{t=0}^{\infty} (1 + r)^{-t}] \cdot w(t) \cdot (1 + r)^{-t}}
\]

Equation 4:

\[
iCER = \frac{100 \cdot 150} {100 \cdot (0.01 \cdot 0.7 \cdot \sum_{16} (1 + 0.03)^{-t})} = 1100
\]

Equation 5:

\[
iCER = \frac{100 \cdot 150} {100 \cdot (0.01 \cdot 0.7 \cdot 65 \cdot (1 + 0.03)^{-5})} = 382
\]
This expression could be modified to contain societal preferences by making the total value of the risk reduction, as expressed in the remaining life expectancy, a function of societal preference. In equation 3 a basic representation of such a function is given where \( w(t) \) is a function representing the societal value of the risk reduction. In this function, preferences not related to timing, such as those related to age, need, QOL or disease-specific preferences should be represented.

5. Discussion

As mentioned in this paper, a significant body of literature exists that questions the rationale for discounting health and monetary issues at an equal rate. The debate on discounting is indicative of one of the fundamental problems in cost-effectiveness analysis, namely relating health effects and money without acknowledging that both might be different in nature. Clearly, health and money are related in some cases, but this does not change the limited extent of this relationship. Although we should not dismiss the relevance of the inconsistency argument or the paralysing paradox, a number of authors have shown that these arguments largely depend on assumptions made about the nature of decision making and the relationship between health and money.[3,9]

In this paper, we have argued that not only do these assumptions influence the discount rate for health effects, but that the measure of the health outcomes dictates how they should be discounted. Expressing the effects of an intervention in the utility gained from single life-years implies that these individual life-years should be corrected for their moment of occurrence since they are consumed at fixed points in the future. Hence, a saved life would have to be transformed into a stream of single life-years, which are discounted back to net present value (i.e. the discounted utility model for health). When assuming this model, the analyst is forced to use similar discount rates for health and money in order to avoid inconsistencies as indicated by Keeler and Cretin[15] and Weinstein and Stason.[16] As argued in this paper, it can be questioned whether this model is an accurate description of reality.

The alternative method described in this paper acknowledges that the main effect from preventive interventions is a reduction of the risk of acquiring a certain disease during the period of exposure. Also, by treating the gained life-years as a non-separable stream, we take into account that life-years are just a statistical construct and can never be seen without the life that it constitutes. By correcting the timing between the start of the intervention and the occurrence of risk reduction by using the social rate of time preference, the typical inconsistencies can be avoided. However, the correction of the health effects for social preferences might be difficult to operationalise. As suggested by Cohen,[8] it might be worthwhile to investigate frameworks from philosophy, but this is definitely an area where more research is needed.

In our opinion, WTP methodology could provide an alternative method to elicit preferences. This then leads to the use of a cost-benefit framework in which all aspects of the risk reduction are valued. The explicit monetary valuation of the risk reduction, as in cost-benefit analysis, provides the clear advantage that there are fewer controversies about discounting. The key problem with putting the alternative discount procedure into practice is how to obtain reliable estimates for the value of relative risk reduction at each timepoint and the probability of surviving to the age of averted mortality \( S(t) \).

Although this alternative method for discounting health gains may better reflect the distinctive nature of health, some remarks need to be made. By specifically acknowledging that benefits of an intervention occur as a risk reduction, the effects of interventions with a large time lapse between intervention and the appearance of the effect in terms of lives gained are valued higher than when both health and monetary effects are discounted at an equal rate. On average, for both HBV and HIV a fairly long time lag exists between exposure to the disease and death due to infection. Interventions against these diseases would be valued as more attractive using a discounting procedure based on occurrence of risk reduction. Considering interventions in infectious disease that additionally succeed in diminishing transmission of
the causative agent in subgroups of the population that didn’t receive the intervention (i.e. herd immunity), we note that risk reductions also occur in these groups. Ideally, these indirect risk reductions should also be incorporated in the analysis. Importantly, the application of this alternative discounting method for interventions against diseases with a short time lag would lead to similar results as the standard method. The results will be identical if the time lag is <1 year, and an annual discount rate is applied.

Clearly, in order for the application of this discounting method to be in line with societal preferences, both time-related and others, these would first have to be elicited. Therefore, more research is necessary on the perception of health effects and societal preferences. Furthermore, even if we manage to elicit representative societal preferences on all the above aspects, the question remains whether decision makers would accept economic evaluation as a purely prescriptive tool (such as cost-benefit analysis is meant to be). Experience of some of the authors with economic evaluation in decision-making processes indicates that decision makers might be unable or unwilling to set priorities on the basis of economic evaluations without imposing their own (or hopefully those of the population they represent) preferences in some form.

6. Conclusions

In this paper, we have shown that truly acknowledging the origin of health benefits could have implications for the discounting procedure. We have proposed an alternative discounting model that focuses on the moment(s) of risk reduction. This model recognizes the potential interpretation of risk reduction for infection as an economic good to be acquired with associated mortality reductions as later indirect effects. When this model is used to assess interventions such as prevention of infection where there is a long time lapse between the intervention and the realisation of the effect (in terms of life-years saved), the intervention will be valued higher than with a traditional discounting method that discounts health and monetary effects at an equal rate.

Acknowledgements

The authors wish to thank Bernhard van den Berg for his comments on earlier versions of this paper. Jasper Bos was also employed by The Netherlands Vaccine Institute, Bilthoven, The Netherlands while constructing this article. The authors have no conflicts of interest that are directly relevant to the content of this article.

Appendix 1: A comparison of standard methodology versus a discounting procedure focused on the occurrence of risk reduction

Consider a hypothetical cohort of 100 patients who receive intervention B at time t = 0. Intervention B is compared with the ‘do nothing’ scenario A. The present value of the incremental costs of intervention B versus A is €150 per patient. The relative risk reduction is 70%. Without the intervention, a patient has an initial probability R of acquiring the infection of 1%. If the patient does not develop the condition, he/she lives until 80 years of age RLE = 65 if patient dies at 15 years.

Without the intervention, risk of developing the condition starts at t = 5 years of age, leading to mortality eventually at t = 15 years of age. So, the effect of the intervention is basically a risk reduction at t = 5. In this example, we assume the probability of surviving to the time of averted mortality S(t) to be 100%. Using the standard discounting procedure (see equation 1) with a discount rate of 3%, leads to the results as shown in equation 4.

Using a discounting procedure based on the occurrence of risk reduction (see equation 2) acknowledges that the effect of this intervention does not take place at time of mortality (t = 15) but at the moment of the occurrence of the risk reduction (at t = 5 years). See equation 5.

Using the discount procedure based on the occurrence of risk reduction leads to a relevantly more favourable cost-effective ratio in this example.

References


Correspondence and offprints: Jasper M. Bos, A Deussinglaan 1, 9713AV Groningen, Groningen, The Netherlands. E-mail: j.bos@farm.rug.nl

© 2004 Adis Data Information BV. All rights reserved.