Accounting for Future Health Events:

Including future costs and valuing health gains in health care decision-making

Meg Perry-Duxbury

Accounting for future health events:

Including future costs and valuing future health gains in health care decision-making

Meg Perry-Duxbury

ISBN: 978-94-6416-686-6
Cover design: Maya Knepflé, Spectric
Lay-out: Publiss | www.publiss.nl
Print: Ridderprint | www.ridderprint.nl

Accounting for Future Health Events

Including future costs and valuing future health gains in health care decision-making

Rekening houden met toekomstige gezondheidsgebeurtenissen

Opnemen van toekomstige kosten en waarderen van toekomstige gezondheidswinsten in besluitvorming over gezondheidszorg

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. FA. van der Duijn Schouten

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Thursday 23rd of September 2021 at 10:30 hours

by

Megan Sarah Perry-Duxbury
born in Oxford, United Kingdom

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	Prof.dr. W.B.F Brouwer
Copromotor:	Dr. P.H.M. van Baal
Other members:	Prof.dr.mr. S.M.A.A. Evers
	Prof.dr. M.J. Postma
	Prof.dr. M.P.M.H. Rutten-van Mölken

Table of Contents

CHAPTER 1

Introduction

Background

Globally, both health and health care spending are increasing rapidly. Between 2000 and 2016, the global average life expectancy increased by 5.5 years,[1] while global per capita health spending grew at a rate of 2.7% per year between 1995 and 2016, with global health spending reaching $8 trillion in 2016.[2] In addition, there is an ever growing supply of new health care technologies,[3] not all of which will be (cost-)effective and some of which may even cause damage to health.[4] This increasingly leads to questions regarding sustainability of health care systems and the cost-effectiveness of health care interventions. The importance of an optimal resource allocation towards and within the health care sector in that sense may be more pertinent than ever before.

Generally speaking, governments have goals of efficiency: whether health care resources are being used to get the most value for money,[5] and equity, which relates to the distribution of health, health care and health gains, for example reducing avoidable differences in health among groups of people with equal need for health, where such groups are often defined using demographic characteristics.[6] Resources, however, are scarce and thus the allocation of resources requires trade-offs. With regard to public spending, even when focusing solely on population health, trade-offs are made between and within the health care system and other departments, and between interventions that affect short- and/or long-term health. When deciding on where to allocate resources, we must recognise that current interventions may result in changes to future costs and benefits. Indeed, successful health interventions in the past are in part responsible for the current aging population, which highlights the importance of considering the full consequences of alternative actions.

Alongside (curative) technologies used within the health care sector, are public health systems that also aim to increase future quality of life or life expectancy, predominantly through prevention. For example, in 2015 the Netherlands spent €12.5 billion on prevention, most of which (76%) was aimed at health protection,[7] which is defined as the mitigation and prevention of infectious diseases and environmental threats.[8] The majority of spending on prevention occurred outside the official definition of health care expenditures, further highlighting how improvements to population health can occur outside the health care system.[9-11] Public health improvements are known

to have played a major role in the fast drop in mortality rates (occurring almost solely via a drop in infectious diseases) in the early twentieth century.[11] They also occurred in several other areas, however, including clean water technologies (such as filtration and chlorination), sanitation (such as the improvement in extensive sewage systems across urban areas), and improved labour conditions (as a result of, for example, the widespread use of hydraulic pumps and the corresponding reduction in exhaustion and injury in the work force). Preventative programmes, such as early warning systems for infectious disease, are currently high on the agenda owing to the COVID-19 crisis; the cost-effectiveness of these systems, however, is still to be evaluated.

Identifying cost-effectiveness

In health economics, whether an intervention is deemed cost-effective or not, in comparison to a relevant comparator, is often assessed using economic evaluation. Economic evaluation is the process of identifying, measuring and valuing the costs and consequences of at least two alternative activities and the subsequent comparative analysis of these activities.[12] The perspective taken by an economic evaluation is not necessarily the same across jurisdictions, with some jurisdictions (such as England and Wales) taking a health care perspective, where the aim is to maximize health from a fixed health care budget, and other jurisdictions (such as the Netherlands) taking a broader societal perspective, where the goal is to maximise social welfare, taking into account the consequences of health technologies on costs and benefits outside the health sector.[12,13] Depending on the perspective taken, the implied decision rule used to evaluate cost-effectiveness varies.

Equation 1.1 refers specifically to the cost-effectiveness decision rule used when taking the broader societal perspective[14]:

$$\frac{\Delta c_t}{\Delta Q} < v \tag{1.1}$$

Where ΔC_t refers to incremental total costs, and is a combination of health care costs, ΔC_h, and broader consumption costs - net of production gains, ΔC_c.

$$\Delta c_t = \Delta c_h + \Delta c_c \tag{1.2}$$

Equation 1.1 shows us that the difference in total costs between the intervention and comparator, ΔC_t, divided by the change in benefits, ΔQ, must be less than the consumption value of health, v. v is referred to as the cost-effectiveness threshold; (beyond this threshold interventions are considered too expensive to be reimbursed). The left-hand side of the equation is referred to as the incremental cost-effectiveness ratio (ICER) and is the outcome that economic evaluations typically produce.

Using the most common form of economic evaluation: cost-utility analysis (CUA), allows outcomes of economic evaluations of different programmes, also targeting different diseases, to be compared. This requires (health) benefits to be presented in such a way that they are comparable across programmes, which is commonly done in the form of quality adjusted life-years (QALYs).[12] QALYs are a summary measure of health that combine life-years gained and health related quality-of-life (HrQoL). For example, an incremental increase of 1 QALY can be due to an increase in 1 life-year gained (LYG) in perfect health (HrQoL = 1), or to an increase in 2 LYG in relatively poor health, where HrQoL is 0.5 on a scale of 0 to 1.

Not all decision-making bodies choose to use the societal perspective, as some prefer to focus on spending only within the health care system. The narrower health care perspective, in which only health care costs (ΔC_h) are considered, leads us to the following decision rule:

$$\frac{\Delta c_h}{\Delta Q} < k \tag{1.3}$$

Under this perspective, we require the ICER, $\frac{\Delta c_h}{\Delta Q}$, to be less than the cost-effectiveness ratio of the interventions displaced by investment into the new intervention being evaluated, where k represents this ratio (Eq. 1.3). k is another type of cost-effectiveness threshold, focused on the supply, rather than the demand, side of health care?

The two cost-effectiveness thresholds, v and k, are conceptually different and therefore estimated quite differently. v is a demand-side threshold and therefore requires information on the monetary (or, more precisely, the consumption) value of a QALY. This is typically elicited using stated preference techniques, in which a sample of individuals are asked their willingness-to-pay (WTP) for a change in health (usually their own). k on the other hand, is

a supply-side threshold, representing the value of displaced activities because of a new intervention, assuming health care spending is constrained by a health care budget. A threshold estimate for k would ideally be found using detailed 'league tables', ranking every intervention on its cost-effectiveness. Given that in reality few health care interventions are assessed for their cost-effectiveness and that we cannot confirm the required assumption that the least cost-effective interventions are always those that are displaced, we can estimate k by calculating the marginal gains generated by an increase in health care spending. k is then the average cost-effectiveness of interventions displaced or expanded at the margin.

While there has been criticism of the QALY,[15,16] it is generally seen as the best available measure to comparably cover a sufficient number of aspects of health. The main advantage of using the QALY to measure benefits of an intervention is that all changes in life-years and/or quality-of-life can be represented by the QALY, allowing for comparisons of treatments across diseases (e.g. the disease being treated, type of treatment, duration of treatment). One of the disadvantages of using the QALY as an outcome measure is that commonly used instruments for measuring and valuing health-related quality of life may not fully capture all benefits generated by health care interventions. Many interventions generate benefits beyond health and may be especially the case for interventions in the area of social care, palliative care or long-term care, where the emphasis may be more on generating wellbeing or preserving dignity than on gaining health. It is also true that interventions offering health protection (such as early warning systems for outbreaks of infectious diseases), improvements in wellbeing and in feelings of safety may be prominent outcomes that are potentially not captured in the QALY. If such broader outcomes are to be included in economic evaluations, it can be done through estimating their monetary value or through the use of broader instruments than common HRQoL instruments. Measures that aim to capture well-being and other outcomes, not covered by the QALY, have been developed in recent years.[17-19] For example, the ICECAP-A (for adults) and -O (for older people) measures specifically address feelings of security as one of its five attributes. Such measures may be used as a companion to or substitute for the QALY in economic evaluations where such feelings are relevant.[18,19]

Cost-effectiveness and the future

A challenge for decision-makers, with respect to health care spending, is that the interventions they choose to reimburse today can affect health, wealth and well-being in the future, directly or indirectly. Determining whether an intervention is cost-effective relative to a relevant threshold (be it defined in terms of k or v), requires that such future impacts are fully captured and valued in an economic evaluation. If this is not the case, economic evaluations may produce unreliable and biased estimates. Hence, we need to have a clear picture of both types of future costs that may occur as a result of health interventions as well as of ways to include them in economic evaluations.

Future costs

Consider an intervention such as bowel cancer screening, which detects bowel cancer in the early stages and, through the application of cancer treatments, may save the life of patients, thereby generating additional life-years. During these additional life-years (also referred to as 'life-years gained'), these patients will inevitably consume additional resources. If we want to include all costs and benefits in order to be able to make a comprehensive trade-off against the threshold, we need to know the costs of this additional consumption, and how to include it in economic evaluations.

In doing so, it is useful to distinguish different types of future costs. Take a patient who received treatment for bowel cancer early enough for it to be eradicated and assume that this intervention (compared to no treatment) generated five additional life-years. In these life years gained, the patient will not only consume follow-up care related to bowel cancer such as regular colonoscopies, known as future *related* medical costs, but also possibly consume care related to, for example, dementia developed during those additional life-years or perhaps an ear infection. These health care costs, which are only connected to the life-years gained and not the cancer itself, are referred to as future *unrelated* medical costs.

In addition to future related and unrelated medical expenditures, the patient will also have other consumption costs related to food, clothing, housing, holidays, and so forth. Such costs are referred to as future *non-medical*

consumption. These costs should be balanced with the value of the production of the patient during the life-years gained. It is also worth noting that the extent to which the welfare benefits of productivity and non-medical consumption are captured in the QALY is still unclear, which also raises questions regarding their appropriate inclusion in economic evaluations. [20,21]

Future (related and unrelated) medical costs may be deemed relevant from both a societal and the healthcare perspective, while future non-medical costs are only relevant when adopting a societal perspective, as these fall outside the health care budget. Nonetheless, the inclusion of various types of future costs (specifically future unrelated medical costs and future non-medical consumption) has been frequently debated.[22-24] While there are multiple views on the inclusion of future costs, this thesis takes the stance that all future costs and benefits relevant to the perspective being used should be considered in economic evaluation.[25-28]

Returning to the decision rule used when adopting a health care perspective, $\frac{\Delta c_h}{\Delta Q} < k$; if future costs are to be included in the ICER (i.e. the left hand side of the equation), then for the sake of consistency, these costs should also be included in any calculations of the cost-effectiveness threshold; otherwise, interventions that did not generate additional future costs (i.e. quality-of-life improving interventions) would be unfairly favoured against those that do (i.e. life-year increasing interventions). Given the way in which v is estimated, it may be that future costs are implicitly included in individuals' valuation of health gains. This is not the case for k however, because when estimating marginal gains in health care spending, future costs need to be deliberately included in the estimate.[29]

Future health benefits and infectious disease outbreaks

When evaluating preventative public health programmes, it is important to recognize that their benefits often occur outside the health care sector. For example, the favourable cost-effectiveness of influenza vaccines in children is in part due to the production losses that occur when parents stay home to care for their children.[30] When taking such a societal perspective, we need not only to estimate the appropriate ICER of an intervention, but also to estimate

the appropriate threshold value for health gains. The demand-side threshold, v, represents the consumption value of health (typically expressed in QALYs). One option for the estimate of v is to use the general WTP for a QALY, for instance obtained using surveys in which people value specific gains in own health[31]. The context in which health gains are gained may be more or less specifically described, which may also affect the value and composition of v.

As mentioned earlier, health interventions may lead to benefits that are not explicitly captured by the QALY, but which are still relevant for economic evaluation. For example, in the case of an early warning system to prevent pandemics, feeling safer is a benefit that would occur from such a system. This could occur through two mechanisms: (1) an early warning system is what would stimulate preventative action, which if successful increases public health and safety, and (2) even if there is no infectious disease outbreak, simply the knowledge of the early warning system existing may generate feelings of safety. It is arguable that these feelings of safety are a benefit that are captured neither by the QALY itself, nor by the elements normally included in the ICER. They may, however, be an element of (mental) well-being that should not be ignored. When eliciting the value of such a system, depending on how the valuation task is designed, people may also express the value of safety.

If the QALY is assumed to exclude non-marketed goods such as safety, then the value of these benefits (if deemed potentially relevant to the final decision) needs to be captured in another way in the economic evaluation. This could be done by eliciting a monetary value through revealed preference or stated preference approaches. In the former, an individual's consumption patterns are examined to elicit a value.[32] For example, we could look at health insurance payments to investigate how much a person values their health. When it comes to investigating a direct relationship between health and payments, this approach can be difficult or even unethical.[33] The stated preference approach requires individuals to (somehow) state their willingness to pay for an increase in health, e.g. via a questionnaire. Which method should ideally be used to elicit the value of feelings of safety, such as those generated by an early warning system for infectious diseases, is, as yet, unclear. It is also unclear how to include them in economic evaluations. This could be done by deriving an isolated estimate of the value of safety to be included in the ICER, or by deriving a value for health gains in a context relevant to the decision

problem, which may include the value of safety. Whilst it is possible that such an 'augmented' threshold would help to bring reimbursement decisions into line with more societal preferences, it would also risk a loss of comparability and consistency across settings.

Research objectives

In this thesis I demonstrate some ways in which we can account for future health events in health care decision-making, be it from a health care perspective or a societal perspective.

The information and background provided in earlier sections of the thesis leads to the following research questions:

- How can we estimate and standardize the inclusion of future costs in economic evaluations?
- What impact does including future costs have on whether interventions are deemed cost-effective?
- Which methods can be used to value feelings of (health) safety?
- How do people value health gains generated by programmes that prevent disease outbreaks?

Focusing on the health care perspective, I investigate how best to standardise the inclusion of future costs in both economic evaluation and in the process of estimating a supply-side threshold. Using data from both the Netherlands and the UK, this thesis adds to the current empirical literature by providing standardized methods for the inclusion of future costs. It also adds to the discussion on including future costs in threshold estimates. Focusing on the societal perspective, I look into how we can value broader health benefits and gains, specifically with regard to programmes for the prevention of infectious disease outbreaks. In doing so, this thesis not only sheds light on methodological issues surrounding the valuation of health benefits but also provides the first estimate of an investment threshold for an early warning system for infectious disease outbreaks.

Chapters 2, 3 and 4, focus on the first half of the research questions. All three chapters explore the methods surrounding how the costs of future disease

burden should be incorporated in the decision-making framework and shed light on the impact of including future costs both on the ICER (Chapters 2 & 3) and on estimates of the cost-effectiveness threshold (Chapter 4). Chapter 2 uses the Dutch context whilst Chapters 3 and 4 focus on England and Wales and their decision body NICE. Hence, in chapter 2, we are also able to investigate the impact of including future non-medical costs, as the Netherlands uses the societal perspective for economic evaluation.

Chapters 5 and 6 answer the second half of the research questions. Chapter 5 provides a literature review that aims to find the most common method used for eliciting a value for feelings of safety, and further discusses the implication of these results on health care decision-making, particularly in the context of preventative health interventions such as an early warning system for infectious disease outbreaks. Chapter 6 provides empirical estimates of what people are willing to pay for a health gain generated by the aforementioned early warning system. This also effectively provides a context-specific threshold estimate for public health interventions such as an early warning system for infectious diseases.

References

1. WHO. Life expectancy and healthy life expectancy. *The Global Health Observatory*. 2020.
2. Chang AY, Cowling K, Micah AE, et al. Past, present, and future of global health financing: A review of development assistance, government, out-of-pocket, and other private spending on health for 195 countries, 1995–2050. *The Lancet*. 2019;393(10187):2233-2260.
3. NICE. Technology appraisal data: appraisal recommendations. Updated November, 2020. Accessed November, 2020. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data/appraisal-recommendations.
4. Le Fanu J. Mass medicalisation is an iatrogenic catastrophe. *BMJ*. 2018;361:k2794.
5. Williams A. Priority setting in public and private health care: A guide through the ideological jungle. *J Health Econ*. 1988;7(2):173-183.
6. Braveman P, Gruskin S. Defining equity in health. *J Epidemiol Community Health*. 2003;57(4):254-258.
7. van Gils PF, Suijkerbuijk AW, Polder JJ, de Wit GA, Koopmanschap M. Nederlandse preventie-uitgaven onder de loep. *TSG: Tijdschrift voor Gezondheidswetenschappen*. 2020.
8. Rubin G. Health protection: Principles and practice. *International journal of epidemiology*. 2017;46(5):1722-1723.
9. de Bekker-Grob EW, Polder JJ, Mackenbach JP, Meerding WJ. Towards a comprehensive estimate of national spending on prevention. *BMC Public Health*. 2007;7(1):252.
10. Cutler D, Deaton A, Lleras-Muney A. The determinants of mortality. *Journal of economic perspectives*. 2006;20(3):97-120.
11. Cutler D, Miller G. The role of public health improvements in health advances: The twentieth-century united states. *Demography*. 2005;42(1):1-22.
12. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford University Press; 2015.
13. Garrison Jr LP, Pauly MV, Willke RJ, Neumann PJ. An overview of value, perspective, and decision context—a health economics approach: An ISPOR special task force report [2]. *Value in Health*. 2018;21(2):124-130.
14. Brouwer W, van Baal P, van Exel J, Versteegh MM. When is it too expensive? cost-effectiveness thresholds and health care decision-making. Eur J of Health Econ. 2019;20(2):175-180.
15. Neumann PJ, Greenberg D. Is the united states ready for QALYs? *Health Aff*. 2009;28(5):1366-1371.
16. Pettitt D, Raza S, Naughton B, et al. The limitations of QALY: A literature review. *Journal of Stem Cell Research and Therapy*. 2016;6(4).

17. Hackert M. *Care for well-being: Development of the well-being of older people measure (WOOP)*. Dissertation. Erasmus University Rotterdam; 2020. Accessed November 15, 2020. https://repub.eur.nl/pub/130440
18. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: The ICECAP-A. *Quality of Life Research*. 2012;21(1):167-176.
19. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? *Soc Sci Med*. 2006;62(8):1891-1901.
20. Nyman JA. Should the consumption of survivors be included as a cost in cost–utility analysis? *Health Econ*. 2004;13(5):417-427.
21. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics*. 1997;16(1):33-64.
22. Meltzer D. Future costs in medical cost-effectiveness analysis. In: *The elgar companion to health economics, second edition*. 2nd ed. Cheltenham, England: Edward Elgar Publishing; 2013:481-489.
23. Gandjour A. Consumption costs and earnings during added years of life-a reply to nyman. *Health Econ*. 2006;15(3):315-317.
24. Richardson JR, Olsen JA. In defence of societal sovereignty: A comment on Nyman 'the inclusion of survivor consumption in CUA'. *Health Econ*. 2006;15(3):311-313.
25. de Vries LM, van Baal P., Brouwer W. Future costs in cost-effectiveness analyses: Past, present, future. *Pharmacoeconomics*. 2019;37:119-130.
26. Ratushnyak S, Hoogendoorn M, van Baal P. Cost-effectiveness of cancer screening: Health and costs in life years gained. *Am J Prev Med*. 2019;57(6):792-799.
27. van Baal P, Morton A, Meltzer D, Brouwer W. Future unrelated medical costs need to be considered in cost effectiveness analysis. *European Journal of Health Economics*. 2018:1-5.
28. van Baal P, Morton A, Brouwer W, Meltzer D, Davis S. Should cost effectiveness analyses for NICE always consider future unrelated medical costs? *Br Med J*. 2017;359:j5096-j5096.
29. van Baal P, Perry-Duxbury M, Bakx P, Versteegh M, van Doorslaer E, Brouwer W. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Econ*. 2019;28(1):87-100.
30. Nichol KL. Cost-effectiveness and socio-economic aspects of childhood influenza vaccination. *Vaccine*. 2011;29(43):7554-7558.
31. Bobinac A, van Exel J, Rutten F, Brouwer W. The value of a QALY: Individual willingness to pay for health gains under risk. *Pharmacoeconomics*. 2014;32(1):75-86.
32. Sen AK. Choice functions and revealed preference. *The Review of Economic Studies*. 1971;38(3):307-317.
33. Craig BM, Lancsar E, Mühlbacher AC, Brown DS, Ostermann J. Health preference research: An overview. *The Patient-Patient-Centered Outcomes Research*. 2017;10(4):507-510.

CHAPTER 2

Practical Guidance for Including Future Costs in Economic Evaluations in the Netherlands: introducing and applying PAID 3.0

With Klas Kellerborg*, Linda de Vries* & Pieter van Baal.

Published in: Value in Health. 2020. 23(11):1453-61.

* *Co-first author*

Abstract

A consensus has been reached in the Netherlands that future medical costs should be included in economic evaluations. Furthermore, internationally, there is the recognition that in countries that adopt a societal perspective estimates of future non-medical consumption are relevant for decision makers as much as production gains are.

The aims of this paper are twofold. Firstly, to update the tool 'Practical Application to Include future Disease costs (PAID 1.1) based on 2013 data, for the estimation of future unrelated medical costs and to introduce future non-medical consumption costs, further standardizing and facilitating the inclusion of future costs. Secondly, to demonstrate how to use the tool in practice; showing the impact of including future unrelated medical costs and future non-medical consumption in a case-study where a life is hypothetically saved at different ages and in two additional cases where published studies are updated by including future costs.

Using the latest published Cost of Illness (COI) data from the year 2017, we model future unrelated medical costs as a function of age, gender, and time to death — which varies per disease. The Household Survey from Centraal Bureau Statistiek is used to estimate future non-medical consumption by age. The updated Incremental Cost-Effectiveness Ratios (ICERs) from the case-studies show that including future costs can have a substantial effect on the ICER, possibly affecting choices made by decision makers.

This paper improves upon previous work and provides the first tool for the inclusion of future non-medical consumption in the Netherlands.

Introduction

While cost-utility analysis (CUA) is increasingly used to assess whether new interventions in healthcare yield sufficient value for money,[1] there are still several methodological issues that require attention. One such issue is the extent to which future costs should be included in CUA,[2,3] where future costs are costs that arise from extending individuals' lives and include all costs in the life-years gained (LYG) from an intervention. They are typically divided into medical (relevant for both societal and healthcare perspectives) and non-medical costs (only relevant for the societal perspective). Non-medical costs here refer to consumption (e.g. costs for housing and food) minus production (benefits from additional work in LYG). For medical costs, a distinction is made between *related* (e.g. costs for check-ups by a cardiologist after a heart-attack) and *unrelated* costs (e.g. costs for treating pneumonia after said heart-attack). Future related medical costs are typically included in CUA. However, including future unrelated medical costs has been frequently debated. Early in the debate, the extent to which future costs should be included was discussed using theoretical models aiming to optimize societal welfare. This led to multiple views on the topic,[4,5] the most compelling being that all future costs and benefits should be considered.[5] Later, the discussion was extended with the more practical view that since future unrelated medical consumption benefits are generally included, the costs thereof should be included to be consistent.[6] This argument was also used to state that future non-medical costs should *not* be included, arguing that the benefits thereof are not systematically included in the QALY.[7] However, there are different views on the extent to which the benefits from non-medical consumption and production are actually included[8-10] and there is so far no compelling (empirical) evidence regarding this.[3] The inclusion of future unrelated medical costs in CUA is now required in the Netherlands[11] and recommended in the United States.[12] While production in LYG is often considered part of productivity costs in CUA using a societal perspective, the inclusion of future non-medical consumption costs is only recommended in the United States.[12]

To facilitate the inclusion of future unrelated medical costs in the Netherlands, the Practical Application to Include future Disease costs (PAID 1.0) was introduced in 2011[13] and updated in 2016 (PAID 1.1). This tool provides age and gender specific average medical spending estimates, which can be specified to

exclude the costs of specific providers and diseases. Estimates are based on a conceptual model that combines various streams of literature. Costs by age are corrected for 'time-to-death' by estimating costs separately for survivors and decedents. 'Time to death' refers to the finding that health care costs are often higher in the last period of life.[14] Since older people are more likely to die, not correcting for this leads to an overestimation of the impact of age on medical expenditures[14] and ignores the fact that saving a life at a given age leads to the postponement of this high-cost last period of life.[15]Future related medical costs of specific diseases already included in the analysis can be excluded to prevent double counting.

This paper provides an extensive update of PAID, to PAID 3.0. First, it uses most recent available COI data (2017). Second, and the largest difference from PAID 1.1, future costs of non-medical consumption are included. We provide guidance on how to use PAID 3 supported by three case-studies. PAID 3.0 can be used free of charge via https://imta.shinyapps.io/PAID3/ and consists of a webapp made in Shiny in R.

Methods

As stated by Meltzer,[5] if the aim of economic evaluations is to maximize social welfare given available resources, all costs following from an intervention should be considered. This implies that both medical costs, related and unrelated, and non-medical costs should be included. The Incremental Cost-Effectiveness Ratio (ICER) including all costs can be written as follows:

$$ICER = \frac{\Delta\left[LY \times (RMC + PC)\right]}{\Delta\, QALY} + \frac{\Delta LY \times UMC}{\Delta\, QALY} + \frac{\Delta LY \times NMC}{\Delta\, QALY} \qquad (2.1)$$

Where:

- LY = life years
- RMC = related medical costs
- PC = productivity costs
- UMC = unrelated medical costs
- NMC = costs of non-medical consumption

Splitting the ICER equation into three ratios distinguishes the elements that are currently included in economic evaluation: related medical costs and productivity costs, from the additional costs that are not usually considered: future unrelated medical costs and future costs of non-medical consumption. Equation (1) also illustrates that differences in unrelated medical costs and future costs of non-medical consumption are purely the result of differences in survival. In our estimation of the ICER, in which future costs are included, we use per capita medical and non-medical consumption cost patterns by age as a starting point.

Lifetime costs of unrelated medical and non-medical consumption *LLY* x [*UMC* + *NMC*] for an individual aged *a* dying at age *n*, can be written as shown in equation 2.2:

$$LY \times [UMC + NMC] = \sum_{a}^{n-1} \sum_{i} sc_i(a,) + \sum_{i} dc_i(n) + \sum_{a}^{n} nmc(a) \qquad (2.2)$$

Where:

- a = age in years
- n = age at death
- dc = decedent costs (healthcare costs in last year of life)
- sc = survivor costs (healthcare costs in other years)
- nmc = average costs of non-medical consumption
- i = index of unrelated diseases

Unrelated medical costs

Rather than taking a bottom-up approach and predicting the risk of all unrelated diseases and connecting these to costs, we take a top-down approach and use total per capita healthcare costs by age and gender as a starting point for estimating unrelated medical costs. Using methods identical to those of van Baal and colleagues,[13] we first break down total healthcare costs by disease, enabling the exclusion of costs for diseases already included in the analysis. Although we explain these methods in the ensuing text, for a more detailed description we refer to the original paper by van Baal and colleagues.[13] Disease-

specific per capita healthcare costs were estimated using data from the Dutch COI from 2017.[16] Rather than using the System of Health Accounts (SHA)[17] perspective (used in PAID 1.1) we use the classification from the National Institute for Public Health and the Environment (RIVM). Although the SHA is internationally recognized, the RIVM definition includes more healthcare costs, such as international care. While average per capita spending hardly changed between 2013 and 2017, age and disease patterns have changed. For example, between 2013 and 2017, costs of psychological disorders increased (14% using 2017 prices) far more than costs in other disease categories such as diseases of the central nervous system (2% when using 2017 prices).

COI data are specified by gender and 21 age-classes, which we interpolated using cubic splines to obtain age-year-specific per capita expenditures and are calculated from population spending totals. The data are further attributed to 100 disease categories and 11 healthcare provider categories (overview in Appendix 2.A). These disease categories include 'Not disease related' and 'Not allocated', meaning that well-care is also included in our definition of unrelated medical costs. As healthcare costs are strongly determined by both age and time to death,[18] individual lifetime healthcare costs can be estimated as shown in the first two parts of equation 2.2. To obtain estimates for survivors and decedents, average per capita expenditures are decomposed into one part attributable to those dying and one part to those surviving at that particular age, assuming average costs are a weighted average of costs for survivors and decedents (age and gender indices are left out here for notational purposes):

$$ac_i = (1 - m) \times sc_i + m \times dc_i \tag{2.3}$$

Where:

- ac_i = average per capita healthcare expenditure for disease i
- m = mortality rate

Disease-specific costs for survivors and decedents can be estimated using equation 2.4, using mortality rates and the gender- and age-dependent ratio between costs for decedents and survivors (r_i):

$$dc_i = r_i \times sc_i \tag{2.4}$$

$$ac_i = sc_i + (r_i - 1) \times m \times sc_i$$

$$sc_i = \frac{ac_i}{1 + (r_i - 1) \times m}$$

Mortality rates from 2017 were obtained from Statistics Netherlands.[19] We used the same disease-specific ratios for costs between decedents and survivors for the hospital sector as used in previous versions of PAID. For ambulatory healthcare, drugs and appliances, and nursing and residential care, ratios from 1999 based on total expenditures were used.[20] To obtain disease-specific ratios for these providers, we exponentiated disease-specific hospital ratios by a scaling constant, describing the relation between costs for decedents and survivors between hospital care and other providers (see Appendix 2.C). For providers for which no ratios were available we assumed that costs for decedents were equal to costs for survivors, as it is predominantly in hospitals that differences in survivor and decedent costs are observed.[18,21]

Non-medical consumption

To estimate costs of non-medical consumption by age we used data from the cross-sectional Dutch Household Consumption survey from 2004 adjusted to 2017 price-levels using consumer price indices from Statistics Netherlands. In previous literature, economies of scale within households have been found to be important when estimating non-medical consumption,[22] implying lower per person consumption costs when household size is larger. For instance, spending on housing can be divided amongst more people when household size is larger, however the utility obtained from housing is likely to be the same whether someone lives on their own or not. This has important implications for estimating future costs of non-medical consumption, as preventing a death in a single-person household will result in more future non-medical consumption than preventing a death in a multi-person household.[23] To estimate costs of non-medical consumption for an average household by age, we fit two generalized additive models using penalized B-splines on age. The first model estimates annual consumption per household equivalent. Consumption per

household equivalent is calculated from household consumption using the OECD modified equivalence scale.[24] The OECD modified equivalence-scale assigns a weighting factor of .5 to each additional adult household member and 0.3 to each child in a multi-person household. The second model estimates the probability of a household having more than one adult; we are interested in making predictions for an average household. Using this equivalence scale implies that preventing a death in a single person household results in twice as much non-medical consumption as compared to a multi-person household with two adults. Details on these models and testing of assumptions can be found elsewhere[1]. The models are used to estimate average annual non-medical consumption by age of preventing a death in an average household as in Equation 2.5:

$$nmc(a) = [hh\ equiv(a) \times h(a) \times w] + [hh\ equiv(a) \times \big(1 - h(a)\big)] \qquad (2.5)$$

Where:

- h = probabilitity of household having >1 adult
- $hh\ equiv$ = annual non-medical consumption per household equivalent
- w = weight of deceased household member, .5 for and adult and .3 for a child

Case-studies

We demonstrate the impact of including future costs on the ICER via three case-studies. Benefits are discounted at 1.5% per year and costs at 4% per year, in adherence with Dutch guidelines.[11] For the first case-study a life is hypothetically saved at ages 0–100, while in the second and third case studies we replicate survival curves from previous studies. In the first case-study lifetables for estimating life-expectancy at all ages are used and combined with quality of life (QoL) data from Gheorghe and colleagues.[25]

For the second case-study, we replicated survival curves from a previously published cost-effectiveness study on oxaliplatin plus fluoropyrimidines versus fluoropyrimidines-only as adjuvant treatment of stage III colon

[1] Kellerborg K, Wouterse B, Brouwer W, Versteegh M, van Baal P. Including costs of non-medical consumption in economic evaluation: Estimation issues and distribution.

cancer,[26] wherein oxaliplatin showed an incremental QALY gain of 1.02 and 0.68 LYG, incremental costs of €9,961, and a corresponding ICER of €9,766. The sample consisted of patients previously diagnosed with stage 3 colon cancer who were randomized to either treatment or control groups. The median age of patients was 60 years. This study is then updated by including estimates of future medical costs, after excluding costs related to colon cancer, and including future non-medical consumption.

For the third case-study, we used the results from a clinical trial assessing survival of pembrolizumab monotherapy compared to platinum-based chemotherapy in a group of previously untreated patients with locally advanced or metastatic non-small-cell lungcancer.[27] The paper from which the survival curves are extracted, does not perform a CEA, and therefore there is no 'baseline' ICER or QALY gains. In this clinical trial the median age at baseline was 64 years of age and 71 percent of patients were male. This case-study demonstrates how to use PAID when survival is short. We recommend using estimates of living one year longer when studies have a relatively short time-horizon (< 5 years as rule of thumb), especially when survival between the new treatment and comparator are highly different in the first study-year. In that case, using decedent costs would create large differences in costs at baseline between the new treatment and the comparator for unrelated diseases. This is implausible as it implies a different past trajectory of costs for the same person before getting the treatment and conflicts with the definition of unrelated medical costs. Costs for living one year longer, $c(a,g)$, can be calculated as follows:

$$c(a,g) = sc(a,g) + dc(a+1,g) - dc(a,g) \tag{2.6}$$

Where:

- c = costs of living one year longer
- g = gender
- a = age in years

Furthermore, while the approach discussed above assumes independence between the healthcare intervention and cost of non-medical and unrelated medical consumption, we provide a framework allowing for a correlation between the intervention and unrelated medical costs — applied in the third

case study. We show the impact of adjusting PAID estimates of unrelated medical costs for this correlation, which is relevant when the studied population is expected to have a different health care use for unrelated diseases than the average population. Estimates can be adjusted using the framework as displayed in equation 2.7, where per capita costs are shown as the product of disease prevalence and per patient costs:

$$sc(a)_i = p(i|i) \times sc(a|i)_i \qquad (2.7)$$

$$dc(a)_i = m(a|i) \times dc(a|i)_i$$

Where:

- $p(i|a)$ = probability of disease i conditional on age a.
- $m(a|i)$ = mortality rate at age a conditional on having disease i.
- $sc(a|i)$ = survivor costs at age a conditional on having disease i.
- $dc(a|i)$ = decedent costs at age a conditional on having disease i.

Given the relationships displayed in equation 2.7 we adjusted unrelated costs to reflect higher prevalence and mortality for stroke among lung cancer patients.[28] We adjusted the unrelated costs for stroke by extracting the costs for stroke separately, multiplying stroke costs with the relative risk of stroke - 1.47 - as estimated by Chen and colleagues[28] and adding these back to the sum of unrelated medical costs, as shown in the equations below.

$$sc(a)_i = p(i|i) \times sc(a|i)_i \qquad (2.8)$$

$$dc(a)_i = m(a|i) \times dc(a|i)_i$$

Where:

- j = unrelated disease with higher costs (e.g. stroke).
- λ = multiplier

To demonstrate how to use PAID with survival data on an individual level we fitted two parametric survival models assuming a Weibull distribution to overall survival results presented in the Kaplan-Meier plot[27] from which we randomly drew individual survival times.

Results

Unrelated medical costs and non-medical costs

Panels A and B in Figure 2.1 show how average healthcare expenditures rise sharply after age 75 while per capita non-medical consumption shows a less strong age pattern but decrease at old age and peak at middle age (identical numbers for males and females since estimates are not gender-specific). These graphs show that up until around age 75, people have higher non-medical than healthcare consumption, whereafter healthcare exceeds non-medical consumption.

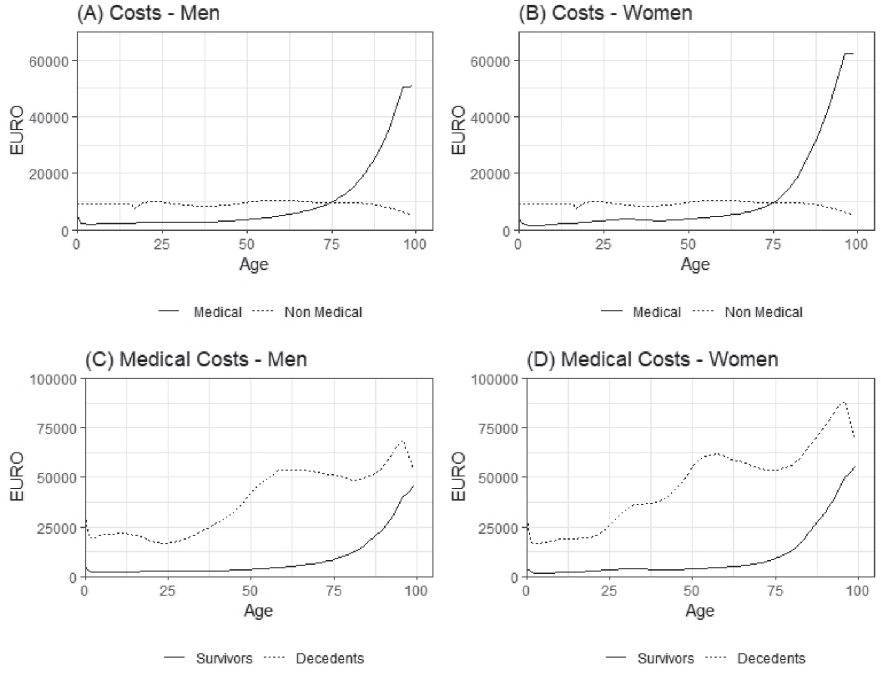


Figure 2.1. A & B — Average per capita medical costs and non-medical consumption by age. C&D — Medical costs, split into survivor and decedent costs by age.

Age-specific per capita medical costs for survivors and decedents are presented in graphs C and D, showing comparable patterns in spending by gender; although women's expenditures are higher, especially at older ages. These graphs show that differences between survivor and decedent costs are highest in the first year of life and between 50 and 75 years and become smaller at the highest ages. This can largely be attributed to causes of death and related periods of illness before dying at different ages. In the first year of life, death often follows a period with high use of medical care. The same holds for middle age. At the highest ages, survivors as well as decedents typically incur higher healthcare expenditures, narrowing the difference in costs.

Case-studies

For the first case-study we estimated the impact of including future costs on the ICER when death is prevented at a certain age (see Figure 2.2). It shows that the older people get, the more expensive it is to be saved.

The results of the second and third case-study are summarized in Table 2.1. Figures 2.3 and 2.4 show differences in costs and survival over time for the two case studies. Including future unrelated medical costs in case study 2, leads to an increase of €3,761 in the ICER; including non-medical consumption adds another €5,440 to the ICER.

Table 2.1 The impact of including future costs on the ICER for case-studies 2 and 3.

	Case-study 2 — € per QALY*	**Case-study 3 — € per life-year**	
		Unadjusted	Adjusted for stroke
Original ICER	9,580	N/A	N/A
Impact including unrelated medical costs on ICER	3,761 (13,341)	5,546	5,619
Impact including non-medical costs on ICER	5,440 (15,020)	9,126	9,126
Total impact on ICER	9,201 (18,781)	14,672	14,745

*Total ICER shown in brackets

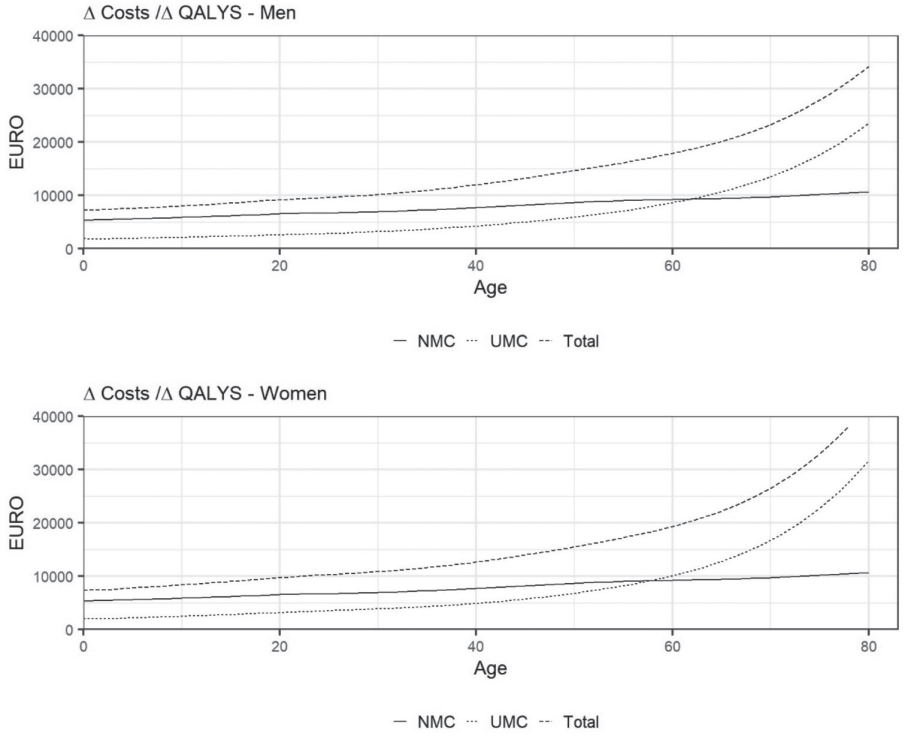


Figure 2.2 – Case-study 1. The hypothetical impact of including future unrelated medical costs (UMC) and future non-medical consumption (NMC) on the ICER when death is prevented (for free) at a certain age.

For the third case-study we estimated a mean survival of 25.1 months for the intervention group (Pembrolizumab) and 15.3 months for the comparator group (chemotherapy); Figure 2.4 (bottom) shows difference in survival. As stated above, in this study no baseline ICERs and QALYs were available. Therefore, only the impact of inclusion on the ICER can be estimated and impact is shown as cost per LYG. We estimated a discounted LYG of 0.77 for the intervention group compared to the comparator. Inclusion of future unrelated medical costs increased the ICER by €5,546, or €5,619 after adjustments for stroke incidence. Including future non-medical consumption further increased the ICER with €9,126. Note here that the impact on the ICER will be different when QALYs instead of life-years are used. If the LYG will be in less than perfect health, this will increase the impact on the ICER.

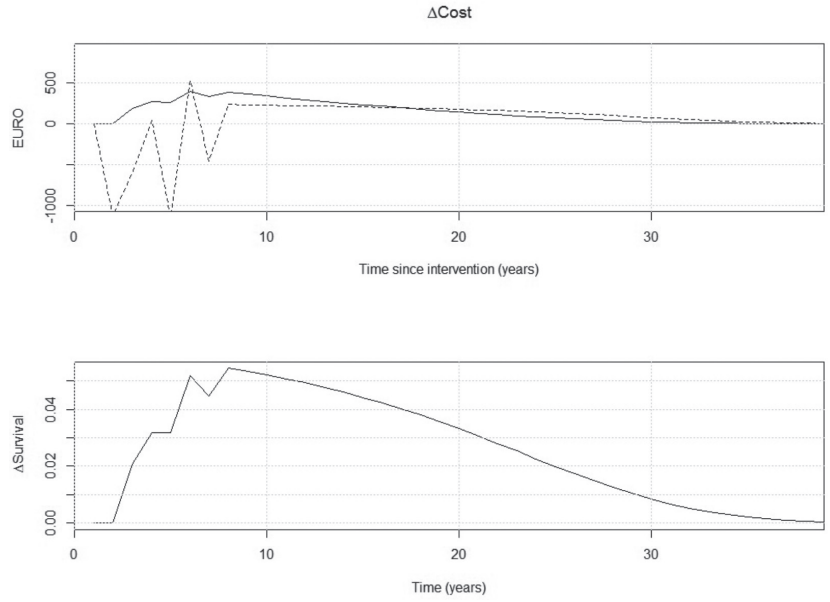


Figure 2.3 – Case study 2. The added costs for including unrelated consumption and non-medical consumption (top), and the difference in survival between intervention and comparator group (bottom).

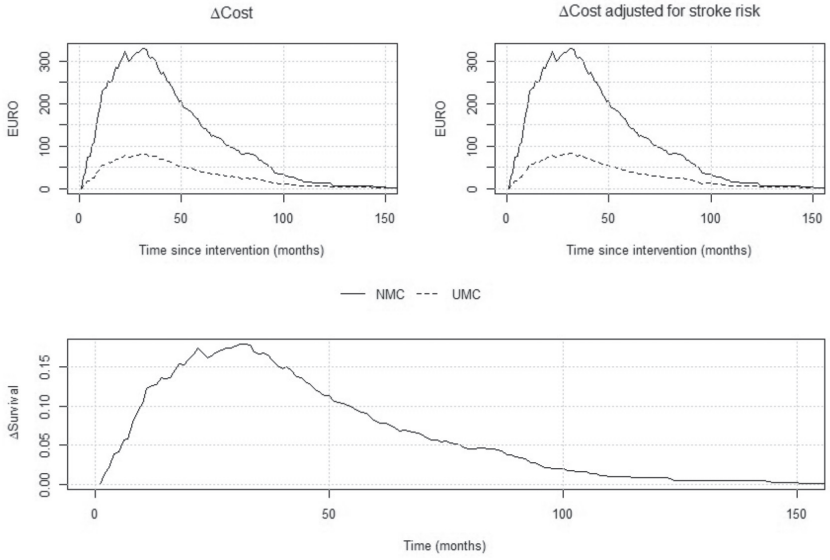


Figure 2.4 – Case study 3. The additional costs by time for the lung cancer intervention (top left), and the additional costs by time when adjusted for increase stroke risk (top right). Difference in survival between intervention and comparator group (bottom).

Discussion & Conclusion

In 2011, a practical tool to include future unrelated medical costs in a standardized manner was introduced.[13] In this paper we updated the tool with the most recent data on medical costs and included estimates for future non-medical consumption. Recent COI data was combined with mortality data and decedent-survivor cost ratios to provide disease-specific estimates of medical expenditures per capita in survivors and decedents. Related costs of an intervention are then excluded from total medical expenditure. Non-medical consumption was estimated, taking into account household economies of scale. This paper further demonstrated how to use the tool in practice, using case studies.

The first case-study refers to the situation of saving a life at a given age, with no intervention costs. It shows that the impact of including future costs becomes larger at higher ages, mainly due to rising healthcare expenditures with age, while in comparison to future medical costs, the impact of including non-medical consumption remains relatively stable over time. The consumption curve (Figure 2.1) follows a U-shape as seen in previous literature,[22,29] however when dividing these costs by QALY changes the curve flattens considerably. Another factor affecting the relative impact of including future costs at younger ages versus older ages, is that the more expensive (older) years, are discounted more highly when lives are saved at younger ages. Furthermore, the impact of including future non-medical consumption is larger than including future unrelated medical costs until approximately the age of 60. This may seem surprising when looking at Figure 2.1, which shows that, per capita, non-medical consumption is larger than medical consumption until approximately the age of 75. However, when estimating the impact of including future unrelated medical costs on saving a life at different ages, we consider time-to-death. As a result, high medical spending in the last year is postponed and additional medical spending is less than suggested by Figure 2.1.

In the second case-study a published evaluation comparing interventions for colon cancer is replicated. Including future unrelated medical costs increases the ICER by almost 40 percent and when all future costs are included the ICER more than doubles. In the Netherlands a cost-effectiveness threshold

ranging from €20,000 up to €80,000 per QALY gained is applied, where the height depends on the principle of proportional shortfall.[30,31] Using the iMTA Disease Burden Calculator,[32] we calculated a proportional shortfall for this case-study of 0.37, which implies that the relevant threshold in this case-study is €20,000.[30] Including future costs in this study could thus make this intervention not cost-effective as it pushes the ICER near the threshold It is important to note that an intervention being not cost-effective is not an undesirable outcome, but simply the result of correctly estimating the change in costs for an intervention.

In the third case-study, we demonstrate how to adjust for short time-horizons, and that PAID estimates can easily be applied to several forms of models. Furthermore, we show how to adjust estimates when costs for unrelated diseases in the studied population is suspected to differ from the general population. This is adjusted for here by using the increased risk of stroke among patients with lung cancer. In this case the difference between future unrelated medical costs, adjusted or unadjusted, is relatively small. However, if the costs of a disease for which the risk is increased were large and the additional risk substantial, the impact of such adjustment would be larger, as shown by Manns et al. in their paper on end-stage renal disease care.[33]

An important limitation is that there are no more recently estimated decedent-survivor cost ratios than those used here. Although more recent estimates of mean overall spending in the last year of life, compared to other years, show comparable numbers,[34] more detailed estimates may show different patterns. An update of these ratios would be useful for future research. A further limitation regarding decedent-survivor cost ratios is that we did not have estimates for all providers, and disease-specific estimates for three providers were derived by combining hospital estimates with provider-specific sector estimates. In a similar vein to this, the classification of costs amongst providers was different for 2017 COI data, and therefore fewer costs could be adjusted using these ratios. It is also worth noting that data from the household survey are relatively old; although data are adjusted to 2017 prices, changes in spending-patterns by age may not be captured. Furthermore, we estimated non-medical consumption by age and assumed no correlation between non-medical consumption and disease. While there is relatively little literature

covering this topic, there are some findings that suggest such a correlation. For example, it may be that medical consumption crowds out non-medical consumption for the severely ill.[35] However, the findings that non-medical consumption decreases from a certain age[22,29,36] may imply that as health decreases (as it does at older ages) so does non-medical consumption. Further research in this area is needed.

Finally, we do not address uncertainty in this paper. Uncertainty could stem from the two key elements of our estimates: survival and costs. While the original costs in this case are averages provided by CBS Netherlands and therefore with little surrounding uncertainty, there are still sources of uncertainty, such as decedent-survivor cost ratios; the larger the TTD effect (larger ratios), the smaller the impact of future costs on the ICER.[5]

In general, including future costs may have a systematic effect on reimbursement decisions as the 'upward' effect on the ICER changes differently by population and intervention. As the cost of extending life increases with age, this implies that the age at which an intervention is given will be of increased importance for the cost-effectiveness of an intervention. Another parameter that affects the magnitude of the impact of including future costs, and thus decisions is the ratio of life-years gained to QALYs gained for a particular intervention. It has been shown that the larger this ratio, the larger the impact of including future costs.[5]

In this paper no specific attention is paid to future related medical costs and future productivity as these are typically already included in economic evaluations and extensive guidance on how to estimate and include these costs is already available in the Netherlands.[37] However, when looking at the total impact of including future costs, production gained at working ages would presumably lead to those years being the least costly. This would, however, also depend on how productivity is measured. In the Netherlands, these costs are typically quantified using the friction costs method and thus limited to the friction period. Using the human capital approach or including informal and household production, would affect the impact of inclusion at different ages. The latter methods would imply higher negative costs (more productivity gains from living longer) and thereby lower ICERs. Another issue worth mentioning

is that, although there is agreement that including future unrelated medical costs would improve the internal consistency of the ICER, implying that costs are included when related benefits are included, how much QALYs capture the benefits from non-medical consumption (and also production) is currently unclear.[6] Furthermore, it is also unclear to what extent thresholds to which ICERs are compared include these benefits.[3] The impact of including future non-medical consumption and the comparison with existing thresholds should thus be interpreted with caution.

To conclude, this paper provided an update and extension of PAID and demonstrated through case-studies the application and impact of including future costs in economic evaluations. Updated ICERs show that including future costs, even if only the unrelated medical costs, can have a substantial effect on the ICER, which could affect decision makers' choices. For future research it would be interesting to see the estimates used in a variety of economic evaluations.

References

1. Garber AM, Sculpher MJ. Chapter 8: Cost effectiveness and payment policy. In: *Handbook of Health Economics*. Vol 2. Elsevier; 2011:471-497.
2. Rappange DR, van Baal P., van Exel J, Feenstra TL, Rutten F, Brouwer W. Unrelated medical costs in life-years gained: Should they be included in economic evaluations of healthcare interventions? *Pharmacoeconomics*. 2008;26(10):815-830.
3. de Vries LM, van Baal P, Brouwer W. Future costs in cost-effectiveness analyses: Past, present, future. *Pharmacoeconomics*. 2019;37:119-130.
4. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *J Health Econ*. 1997;16(1):1-31.
5. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ*. 1997;16(1):33-64.
6. Nyman JA. Should the consumption of survivors be included as a cost in cost-utility analysis? *Health Econ*. 2004;13(5):417-427.
7. Nyman JA. Measurement of QALYS and the welfare implications of survivor consumption and leisure forgone. *Health Econ*. 2011;20(1):56-67.
8. Meltzer D. Future costs in medical cost-effectiveness analysis. In: *The elgar companion to health economics, second edition*. 2nd ed. Edward Elgar Publishing; 2013:481-489.
9. Gandjour A. Consumption costs and earnings during added years of life — a reply to Nyman. *Health Econ*. 2006;15(3):315-317.
10. Richardson JRJ, Olsen JA. In defence of societal sovereignty: A comment on Nyman "the inclusion of survivor consumption in CUA". *Health Econ*. 2006;15(3):311-313.
11. Zorginstituut Nederland. *Guideline for Economic Evaluations in Healthcare*. Diemen, The Netherlands: Zorginstituut Nederland; 2016.
12. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. *JAMA*. 2016;316(10):1093-1103.
13. van Baal P, Wong A, Slobbe LCJ, Polder JJ, Brouwer W, De Wit GA. Standardizing the inclusion of indirect medical costs in economic evaluations. *Pharmacoeconomics*. 2011;29(3):175-187.
14. Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: A red herring? *Health Econ*. 1999;8(6):485-496.
15. Gandjour A, Lauterbach KW. Does prevention save costs?: Considering deferral of the expensive last year of life. *J Health Econ*. 2005;24(4):715-724.
16. Rijksinstituut voor Volksgezondheid en Milieu (RIVM). Volksgezondheidenzorg.info. https://www.volksgezondheidenzorg.info/. Accessed August 16, 2020.
17. World Health Organisation; European Commission, Organisation for Economic Co-operation and Development. *A System Of Health Accounts*. Paris, France: OECD; 2011.

18. Wong A, van Baal P, Boshuizen HC, Polder JJ. Exploring the influence of proximity to death on disease-specific hospital expenditures: A carpaccio of red herrings. *Health Econ*. 2011;20(4):379-400.
19. Central Bureau for Statistics (CBS). Statline. https://www.cbs.nl/en-gb/ourservices/open-data/statline-as-open-data/statline. Accessed August, 2020.
20. Polder JJ, Barendregt JJ, van Oers H. Health care costs in the last year of life — the Dutch experience. *Soc Sci Med*. 2006;63(7):1720-1731.
21. De Meijer C, Koopmanschap M, d'Uva TB, Van Doorslaer E. Determinants of long-term care spending: Age, time to death or disability? *J Health Econ*. 2011;30(2):425-438.
22. Alessie R, De Ree J. Explaining the hump in life cycle consumption profiles. *Economist (Leiden)*. 2009;157(1):107-120.
23. Nelson JA. Household economies of scale in consumption: Theory and evidence. *Econometrica: Journal of the Econometric Society*. 1988:1301-1314.
24. Hagenaars AJ, De Vos K, Asghar Zaidi M. *Poverty statistics in the late 1980s: Research based on micro-data*. Office for Official Publications of the European Communities; 1994.
25. Gheorghe M, Brouwer W, van Baal P. Did the health of the dutch population improve between 2001 and 2008? investigating age-and gender-specific trends in quality of life. *The European Journal of Health Economics*. 2015;16(8):801-811.
26. van Gils CW, de Groot S, Redekop WK, Koopman M, Punt CJ, Uyl-de Groot CA. Real-world cost-effectiveness of oxaliplatin in stage III colon cancer: A synthesis of clinical trial and daily practice evidence. *Pharmacoeconomics*. 2013;31(8):703-718.
27. Mok TS, Wu Y, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): A randomised, open-label, controlled, phase 3 trial. *Lancet*. 2019;393(10183):1819-1830.
28. Chen P, Muo C, Lee Y, Yu Y, Sung F. Lung cancer and incidence of stroke: A population-based cohort study. *Stroke*. 2011;42(11):3034-3039.
29. Fernández-Villaverde J, Krueger D. Consumption over the life cycle: Facts from consumer expenditure survey data. *Rev Econ Stat*. 2007;89(3):552-565.
30. Reckers-Droog VT, van Exel J, Brouwer W. Looking back and moving forward: On the application of proportional shortfall in healthcare priority setting in the Netherlands. *Health Policy*. 2018;122(6):621-629.
31. Brouwer W, van Baal P, van Exel J, Versteegh MM. When is it too expensive? cost-effectiveness thresholds and health care decision-making. *Eur J Health Econ*. 2019;20(2):175-180.
32. Versteegh M, Ramos IC, Buyukkaramikli NC, Ansaripour A, Reckers-Droog V, Brouwer WB. Severity-adjusted probability of being cost effective. *Pharmacoeconomics*. 2019;37(9):1155-1163.
33. Manns B, Meltzer D, Taub K, Donaldson C. Illustrating the impact of including future costs in economic evaluations: An application to end□stage renal disease care. *Health Econ*. 2003;12(11):949-958.

34. Bakx P, O'Donnell O, Van Doorslaer E. Spending on health care in the Netherlands: Not going so dutch. *Fiscal Studies*. 2016;37(3-4):593-625.

35. Zaidi A, Burchardt T. Comparing incomes when needs differ: Equivalization for the extra costs of disability in the UK. *Rev Income Wealth*. 2005;51(1):89-114.

36. Gourinchas P-O, Parker JA. Consumption over the life cycle. *Econometrica*. 2002;70(1):47-89.

37. Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Tan S. *Costing manual: Methodology of costing research and reference prices for economic evaluations in healthcare*. Diemen, The Netherlands: Zorginstituut Nederland; 2015.

Appendix

Appendix 2.A - Healthcare providers

Table 2.A.1 Summary of healthcare provider categories in PAID 3.0 (based on the categories distinguished in the Dutch Costs of Illness study)

Cost of Illness VTV *(Volksgezondheid* Toekomstverkenning*)* healthcare provider categories	**% of total costs in 2017**	**Data used to attribute average costs per disease to last year of life and other years**
Hospitals (HC)	30.3	Hospital records linkage
Nursing and residential care facilities (LTC)	20.5	Hospital records scaled to insurance claims
Providers of ambulatory healthcare (GP)	10.8	Hospital records scaled to insurance claims
Retail sale and other providers of medical goods (Med)	9.0	Hospital records scaled to insurance claims
Provision and administration of public health programmes*	1.9	Not applicable**
General health administration and insurance*	4.4	Not applicable**
Other healthcare*	3.3	Not applicable**
Welfare*	0.5	Not applicable**
Ambulance and transport*	0.6	Not applicable**
Disabled care*	11.3	Not applicable**
Mental healthcare*	7.4	Not applicable**

* *These healthcare providers are grouped together and referred to as 'other healthcare providers'*

** *Costs for 'other healthcare providers' depend only on age and gender for PAID 3.*

Appendix 2.B - Disease categories

Table 2.B.1: Summary of disease categories in PAID 3.0 (based on the categories distinguished in the Dutch Costs of Illness study)

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
	Infectious and parasitic disease				1.72
1	Intestinal infectious diseases	A00-A09	001-009	r001 (001-008) r002 (009)	12.50
2	Tuberculosis	A10-A19, B90	010-018, 137	r003 (010-018, 137)	1.14
3	Meningitis	A39, A87, G00-G03	036, 047, 320-322	r006 (remainder of 001-139, except 0340, 0993, 0994, 135, 1361)	2.27
4	Septicemia	A40-A41	038	r004 (038)	5.68
5	HIV/AIDS	B20	042-044	r005 (042-044 or 2795, 2796)	15.91
6	Sexually transmitted diseases	A60, A50-A58 A63, B00, B07, B08	054, 078, 090-099	r006 (remainder of 001-139, except 0340, 0993, 0994, 135, 1361)	4.55
7	Hepatitis	B15-B19, K77	070, 573.1	r006 (remainder of 001-139, except 0340, 0993, 0994, 135, 1361) r072 (570, 571.4-573)	13.64

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
8	Other infectious diseases	A20-A46, A35, A42, A48, A68-A69, A70-A71, A75, A77-A85, A87-A88, A90, A92, A93, A95, A98, B01-B06, B08, B09, B26-B27, B30, B33, B50-B57, B60, B91, B95-B99, Z11, Z20, Z23, Z41, Z51, Z79	019-035, 037, 039-041, 045-046, 048-053, 055-069, 071-077, 079-089, 100-136, 138-139, v01-v07, v73-v75	r006 (remainder of 001-139, except 0340, 0993, 0994, 135, 1361) r130 (remainder of V01-V82)	43.18
	Neoplasms				6.69
9	Oesophagus cancer	C15	150	r015 (remainder of 140-208)	1.46
10	Stomach cancer	C16	151	r015 (remainder of 140-208)	0.58
11	Colorectal cancer	C18-C21	153-154	r007 (153, 154)	10.20
12	Pancreas cancer	C25	157	r015 (remainder of 140-208)	1.75
13	Lung cancer	C33-C34	162	r008 (162)	7.87
14	Breast cancer	C50	174	r010 (174,175)**	14.87
15	Cervical cancer	C53-C55	180	r011 (179,180,182)**	2.33
16	Ovary cancer	C56-C57	183	r012 (183)**	1.46
17	Prostate cancer	C61	185	r013 (185)*	6.71
18	Bladder and kidney cancer	C64-C68	188-189	r014 (188) r015 (remainder of 140-208)	4.66

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
19	Non-Hodgkin's disease	C82- C83	200, 202	r015 (remainder of 140-208)	3.79
20	Other lymphoid cancer and leukemia	C81, C90-C95	201, 203-208	r015 (remainder of 140-208)	12.14
21	Other cancers	C00-C14, C17, C22-24, C26-C32, C38-C43, C50, C69-C80, C7A, Z12	140-149, 152, 155-156, 158-161, 163-172, 175-178, 190-199, 209, v76	r015 (remainder of 140-208) r010 (174,175)** r130 (remainder of V01-V82)	21.57
22	Other benign neoplasms	C44, D03, D10-D23, D30-D36	173, 210-216, 223-239	r009 (172,173) r016 (230-234) r017 (2113,2114) r019 (remainder of 210-239)	10.79
	Endocrine, nutritional and metabolic diseases				2.85
23	Diabetes mellitus including diabetic complications	E10-E11, E0842, E0942, E1042, E1142, E1342, E113, N048, N08, N038, N058	250, 357.2, 362.0, 581.8, 582.8, 583.8	r022 (250) r034 (remainder of 320-359) r036 (remainder of 360-379) r090 (580-5834, 5838, 5839, 5900-5902, 5908, 5909, 591, 5933-5935, 5937, 5996) r099 (remainder of 580-629 except 5997)	63.70

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
24	Other endocrine, nutritional and metabolic diseases	E009, E01, E04-E09, E15, E21-E22, E24, E27-E29, E30-E32, E34, E40, E41, E43-E46, E50-E51, E53-E56, E65-E67, E70, E74, D80, Z13	240-249, 251-279, V77	r023 (remainder of 240-278) r130 (remainder of V01-V82)	36.30
	Diseases of the blood and the blood-forming organs				0.51
25	Diseases of the blood and blood-forming organs	D50, D51, D56-D59, D61, D63-D75, Z13	280-289, V78	r020 (280-285) r021 (135, 2790-2793, 2798, 286-288, 2890, 2894-2899) r130 (remainder of V01-V82)	100.00
	Mental and behavioural disorders				28.60
26	Dementia	F01-F05, F329	290, 311	r024 (2900-2902, 2904-2909, 2941) r028 (296, 2980, 3004, 3011, 311)	35.99
27	Schizophrenia	F20	295	r027 (295, 2970-2973, 2978-2979, 2983-2989)	1.64
28	Depression	F30, F341	296, 300.4	r028 (296, 2980, 3004, 3011, 311)	4.50
29	Anxiety	F40-F42, F449, F488, F43, F438	300.0, 300.10-300.15, 300.2-300.3, 300.5, 308, 309.8	r029 (remainder of 290-319)	3.07

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
30	Personality disorders	F431, F6811, F688, F60	300.16-300.19, 301	r028 (296, 2980, 3004, 3011, 311) r029 (remainder of 290-319)	2.73
31	Dependency on alcohol and drugs	F10-F16, F18-F19,	291-292, 303-305	r025 (291, 303, 3050) r026 (292, 2940, 304, 3051-3059)	3.27
32	Other mental disorders	F02-F06, F07, F4320, F4321, F45, F481-F489, F54, F64-F66, F81, F84, F90, F93, F95, Z134	293-294, 299, 300.6-300.9, 302, 306-307, 309.0-309.7, 309.9, 310, 312-316, v79	r024 (2900-2902, 2904-2909, 2941) r026 (292, 2940, 304, 3051-3059) r029 (remainder of 290-319) r130 (remainder of V01-V82)	16.02
33	Mental retardation, including Down's syndrome	F70-F73, F79, Q909	317-319, 758.0	r029 (remainder of 290-319) r110 (740-759)	32.86
	Diseases of the nervous system				6.71
34	Parkinson's disease	G20-G21	332	r034 (remainder of 320-359)	3.49
35	Multiple sclerosis	G35	340	r031 (340)	3.49
36	Epilepsy	G40	345	r032 (345)	5.23
37	Cataract	H26	366	r035 (366)	5.23
38	Disorders of accommodation and refraction	H52	367	r036 (remainder of 360-379)	15.41
39	Blindness and low vision	H54	369	r036 (remainder of 360-379)	7.56
40	Conjunctivitis	H00-H02	373-374	r036 (remainder of 360-379)	1.74

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
41	Other diseases of the eye and adnexa	H04-H05, H10, H15-H17, H20, H30, H40, H44, H47, H50, H53,	360-361, 362.1-362.9, 363-365, 368, 370-372, 375-379	r036 (remainder of 360-379)	6.98
42	Ear disorders	H60-H95	380-389	r037 (380-389)	22.97
43	Other diseases of the nervous system and sense organs	G04-G19, G22-G34, G36-G39, G40-G99, Z135	323-331, 333-339, 341-344, 346-356, 357.0-357.1, 357.3-357.9, 358-359, v80	r030 (3310) r034 (remainder of 320-359) r130 (remainder of V01-V82)	27.91
	Diseases of the circulatory system				11.66
44	Hypertension	I10-I15	401-405	r038 (401-405)	6.35
45	Coronary heart disease	I21-125	410-414	r039 (413; ICD-9-CM: 4111, 413) r040 (410) r041 (411-412, 414; ICD-9-CM: 4110, 4118, 412, 414)	22.24
46	Heart failure	I50-I51	428-429	r044 (428) r048 (2891-2893, remainder of 390-459 except 435, 446 and 4590)	8.03
47	Other heart disease, including pulmonary circulation	I30-I49	390-398, 415-427	r048 (2891-2893, remainder of 390-459 except 435, 446 and 4590) r042 (415-417) r043 (426, 427)	19.40

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
48	Stroke	I60-I69	430-438	r033 (435) r045 (430-434, 436-438)	14.38
49	Diseases of arteries	I70-I79	440-448	r046 (440) r048 (2891-2893, remainder of 390-459 except 435, 446 and 4590)	9.36
50	Other circulatory diseases	I80-199	451-459	r048 (2891-2893, remainder of 390-459 except 435, 446 and 4590)	20.23
	Diseases of the respiratory system				3.39
51	Acute upper respiratory infections	J00-J06	460-466	r049 (0340, 460-465, 487; ICD-9-CM: 340, 460-465, 487, 488) r051 (466 (acute lower respiratory infections other than acute bronchitis, acute bronchiolitis and pneumonia were not separated in ICD-9, no J22 equivalent))	11.49
52	Pneumonia and influenza	J09-J18	480-487	r050 (480-486) r049 (0340, 460-465, 487; ICD-9-CM: 340, 460-465, 487, 488)	16.67
53	Asthma and chronic obstructive pulmonary disease (COPD)	J40-J47	490-496	r054 (490-492, 494, 496) r055 (493) r056 (remainder of 460-519)	14.37

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
54	Other respiratory diseases	J30-J39, J60-J99	467-479, 488-489, 497-519	r049 (0340, 460-465, 487; ICD-9-CM: 340, 460-465, 487, 488) r053 (470-473, 475-478) r056 (remainder of 460-519)	57.47
	Diseases of the digestive system				6.84
55	Other diseases of teeth, jaw and salivary glands	K00, K030-K039, K04, M26, K080, K082-K089, K09-K14	520, 521.1-521.9, 522, 524, 525.0, 525.2-525.9, 526-529	r057 (520-525) r058 (526-529)	62.11
56	Gastroduodenal ulcers	K25-K28	531-534	r060 (531-534)	0.57
57	Appendicitis	K35-K38	540-543	r062 (540-543)	1.42
58	Abdominal hernia	K40-K46	550-553	r063 (550) r064 (551-553)	2.85
59	Inflammatory intestinal disease	K50-K52	555-556	r065(555, 556)	6.84
60	Other intestinal diseases	K55-K64	557-569	r066 (558) r067 (560) r068 (562) r069 (565, 566, 5690-5694) r070 (557, 564, 5695, 5698, 5699)	1.71

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
61	Chronic liver disease and cirrhosis	K70	571	r071 (5710-5713) r072 (570, 5714-573)	1.42
62	Other liver diseases	K72, K75, K763-K769, K77	570, 572, 573.0, 573.2-573.9	r072 (570, 5714-573) r076 (remainder of 520-579)	0.00
63	Gallbladder diseases	K80-K83	574-576	r073 (574) r074 (575, 576)	3.70
64	Other diseases of the digestive system	K20, K29-K31, K86-K90	530, 535-537, 577-579	r059 (530) r061 (535-537) r075 (577) r076 (remainder of 520-579)	19.37
65	Diseases of the genitourinary system				3.16
66	Nephritis and nephropathy	N00-N01, N032-N039, N043-N044, N049, N059, N17-N19	580, 581.0-581.7, 581.9, 582.0-582.7, 582.9, 583.0-583.7, 583.9, 584-589	r090 (580-5834, 5838, 5839, 5900-5902, 5908, 5909, 591, 5933-5935, 5937, 5996) r091 (5836, 5837, 584-586) r093 (0994, 587-589, 5903, 5930-5932, 5936, 5938, 5939, 595- 597, 5980, 5981, 5988, 5989, 5990-5995, 5998, 5999, 6256)	27.78

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
67	Acute renal and urinary infections	N11, N30, N34, N390	590, 595, 597, 599.0	r090 (580-5834, 5838, 5839, 5900-5902, 5908, 5909, 591, 5933-5935, 5937, 5996) r093 (0994, 587-589, 5903, 5930-5932, 5936, 5938, 5939, 595- 597, 5980, 5981, 5988, 5989,5990-5995, 5998, 5999, 6256)	8.64
68	Other renal and urinary diseases	N13-521, N32, N35, N360-N369	591-594, 596, 598, 599.1-599.9	r090 (580-5834, 5838, 5839, 5900-5902, 5908, 5909, 591, 5933-5935, 5937, 5996) r092 (592, 594, 7880) r093 (0994, 587-589, 5903, 5930-5932, 5936, 5938, 5939, 595- 597, 5980, 5981, 5988, 5989, 5990-5995, 5998, 5999, 6256)	28.40
69	Hyperplasia of prostate	N40	600	r094 (600)*	4.94
70	Other disorders of male genital organs	N41-N51	601-608	No estimates	4.94
71	Disorders of female genital organs	N60-N92, N94	610-627, 629	No estimates	18.52
72	Female infertility	N97, Z31	628, v26	No estimates	6.79
	Pregnancy, childbirth and the puerperium				2.07
73	Pregnancy	O00-O48, Z34	630-648, V22-V23	No estimates	34.91

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
74	Childbirth	O60-O84, Z76, Z37, Z38	650-669, V20, V27, V30-V39	No estimates	37.74
75	Puerperium	O85-O92, Z39	670-676, V24	No estimates	20.75
76	Contraception	Z30	V25	No estimates	7.55
	Diseases of the skin and subcutaneous tissue				1.64
77	Eczema	L22-L25	691-692	r078 (690-693, 6943, 696-6983, 6988, 6989)	13.10
78	Decubitus	L89	707	r079 (remainder of 680-709)	13.10
79	Other diseases of the skin and subcutaneous tissue	L02-L21, L27, L29, L40, L43, L44, L50-L51, L53, L60, L65-L66, L70, L74, L81-L98	680-690, 693-706, 708-709	r077 (680-686) r078 (690-693, 6943, 696-6983, 6988, 6989) r079 (remainder of 680-709)	75.00
	Diseases of the musculoskeletal system and connective tissue				7.47
80	Rheumatoid arthritis	M05-M08	714	r083 (0993, 711-716, 718, 719, 7271, 7284)	9.92

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
81	Osteoarthrosis	M15	715	r080 (Not a concept in ICD-9 at four-digit level. Can only be defined by using the optional fifth digit 5 to 715, i.e. 715.15, 715.25, 715.35 and 715.95) r081 (Not a concept in ICD-9 at four-digit level. Can only be defined by using the optional fifth digit 6 to 715, i.e. 715.16, 715.26, 715.36 and 715.96) r083 (0993, 711-716, 718, 719, 7271, 7284)	
82	Dorsopathy	M40-M54	720-724	r085 (720, 721, 7230, 7235, 7240, 737) r086 (7220-7227, 7229) r087 (7231, 7234, 7236, 7241-7243, 7245)	14.36
83	Osteoporosis	M810, M844	733.0-733.1	r089 (remainder of 710-739)	1.83
84	Internal derangement of the knee	M23	717	r082 (717)	5.74
85	Unspecified musculoskeletal diseases or conditions	M35, M75, M60, M61, M65, M79	725-729	r083 (0993, 711-716, 718, 719, 7271, 7284) r084 (1361, 2794, 446, 710, 725, 7285) r086 (7220-7227, 7229) r088 (726, 7270, 7272-7279) r089 (7280-7283, 7286-7289, 729)	37.08

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
86	Other diseases of the musculoskeletal system	M00, M12-M14, M20-M21, M24-M25, M32-M35, M40-M42, M85- M86, M88-M89, M91-M92, M95, M99	710-713, 716, 718-719, 730-732, 733.2-733.9, 734-739	r083 (0993, 711-716, 718, 719, 7271, 7284) r084 (1361, 2794, 446, 710, 725, 7285) r089 (remainder of 710-739)	31.33
	Congenital malformations				0.58
87	Congenital anomalies of nervous system	Q00-Q05	740-742	No estimate	3.33
88	Congenital anomalies of circulatory system	Q20-Q25	745-747	No estimate	26.67
89	Other congenital anomalies, excluding Down's syndrome	Q11, Q16, Q30, Q35, Q38, Q41-Q43, Q50, Q60, Q67-Q97, Z36	743-744, 748-757, 758.1-758.9, 759, v28	No estimate	70.00
	Certain conditions originating in the perinatal period				0.23
90	Disorders relating to premature birth	P07	765	No estimate	75.00
91	Other conditions originating in the perinatal period	P00-P04, P08-P15, P22-P28, P50-P90	760-763, 766-767, 769-770, 772-779	No estimate	16.67

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified				1.91
92	Symptoms, signs and ill-defined conditions	R40-R99	780-799	No estimate	100
	Injury, poison and certain other consequences of external causes				3.28
93	Skull-brain injury	S02, S04- S06	800-801, 803-804, 850-854, 950-951	No estimates	12.50
94	Fractures of upper extremities	S42-S52	810-819	No estimate	7.14
95	Hip fracture	S72	820-821	No estimate	16.07
96	Other lower extremity fracture	S82, S92	822-829	No estimate	14.29
97	Superficial injury	S00, S05, S09-S10, S20, S30, S40, S60, S70, S80, S90, T07	910-924	No estimates	2.38

Disease category number	Cost of Illness System of Health Account Disease categories	International Statistical Classification of Diseases and Related Health Problems (ICD) - 10 codes	International Statistical Classification of Diseases and Related Health Problems (ICD) - 9 codes	Matched disease-specific ratios estimated using International Shortlist for Hospital Morbidity Tabulation (ISHMT) classification (ICD-9 codes between brackets)	% of total costs in 2017 (disease subcategories show share of costs within header category)
98	Other injury	S01, S03, S07-S08, S11-S19, S21-S29, S31-S39, S41, S53-S59, S61-69, S71, S73-S79, S81, S83-S89, S91, S91-S99	802, 805-809, 830-849, 855-909, 925-949, 952-999	No estimates	48.21
	Not allocated/ Not disease related				10.68
99	Not allocated	Z01, Z03, Z09, Z13, Z43, Z45, Z48- Z51, Z65, Z76-Z79, Z80-Z84, Z85-Z88, Z91	V10-V19, V21, V40-V57, V58.0-V58.4, V58.6-V58.9, V63-V64, V66-V68, V71-V72, V81-V82	No estimates	91.97
100	Not disease-related	Z00, Z02, Z52, Z56, Z59, Z60, Z69, Z71, Z74, Z75, Z76	V59-V62, V65, V70	No estimates	8.03

**disease-specific ratio only estimated for men*
***disease-specific ratio only estimated for women*
No estimate: no disease-specific ratio found for both men and women

Appendix 2.C - Derivation scaling factor ratios

To obtain disease-specific ratios for these providers, we exponentiated the disease-specific hospital ratios by a scaling constant describing the relation between costs for decedents and survivors between hospital care and the other providers (equation C.1). The log scale, instead of multiplying by a constant, is chosen for scaling to prevent that negative ratios would become positive (or vice versa).

$$r_{i,j>1} = r_{i,j=1}^{x_j>1} \tag{2.C.1}$$

Where:

- j = index denoting the healthcare provider.
- j = 1 implies hospital care.
- $r_{i,j>1}$ = ratio for disease i for healthcare provider j other than hospital care.
- $x_{j>1}$ = scaling constant for healthcare provider j other than hospital care.

Equation C.1 implies that age- and sex-specific distributions of ratios are proportional on the log scale for each healthcare provider. Using equation C.1 for a baseline disease (i=1), this can be rewritten as equation C.2:

$$r_{i=1,j>1} = r_{i=1,j=1}^{x_{j>1}} \rightarrow \log(r_{i=1,j>1}) = x_{j>1}\log(r_{i=1,j=1}) \rightarrow x_{j>1} = \frac{\log(r_{i=1,j>1})}{\log(r_{i=1,i=1})} \tag{2.C.2}$$

We assume that the scaling factor $x_{j>1}$ is equal for all diseases, which leads to equation C.3:

$$x_{j>1} = \frac{\log(r_{i=1,j>1})}{\log(r_{i=1,j=1})} = \frac{\log(r_{i>1,j>1})}{\log(r_{i>1,j=1})} \; for\ all\ values\ of\ i \tag{2.C.3}$$

The scaling factor was found by minimizing the distance between total survivor costs using the estimated ratios for total expenditures and total survivor costs as the sum of disease-specific survivor costs (equation C.4):

$$\left(sc_{tot,j>1} - \sum_i \frac{ac_{i,j>1}}{1 + (r_{i,j=1}^{x} - 1) \times m}\right)^2 \tag{2.C.4}$$

CHAPTER 3

Cured today, ill tomorrow: a method for including future unrelated medical costs in economic evaluation in England & Wales

With Miqdad Asaria, James Lomas & Pieter van Baal.
Published in: Value in Health. 2020. 23(8):1027-1033.

Abstract

In many countries, future unrelated medical costs occurring during life-years gained are excluded from economic evaluation, while benefits of unrelated medical care are implicitly included leading to life-extending interventions being disproportionately favoured over quality-of-life interventions. This paper provides a standardized framework for the inclusion of future unrelated medical costs and demonstrates how this framework can be applied in England and Wales.

Data sources are combined to construct estimates of per capita NHS spending by age, gender and time to death, and a framework is developed for adjusting these estimates for costs of related diseases. Using survival curves from three empirical examples, we demonstrate how our estimates for unrelated NHS spending can be used to include unrelated medical costs in CEA and the impact depending on age, life-years gained and baseline costs of the target group.

Our results show that including future unrelated medical costs is feasible and standardisable. Empirical examples show that this inclusion leads to an increase in the ICER of between seven and thirteen percent.

This paper contributes to the methodology debate over unrelated costs and how to systematically include them in economic evaluation. Results show that it is both important and possible to include future unrelated medical costs.

Introduction

Population ageing and its relationship with healthcare has not escaped attention in the research community.[1] A concern regarding the treatment of ageing in economic evaluation is that 'extending' life leads to additional consumption of healthcare.[2] A patient who receives a medical intervention providing them with additional life-years will continue consuming healthcare in their life-years gained. For example, a patient who is treated for a heart attack and survives may, during their life-years gained, get cancer. The costs in the life-years gained that are directly related to the disease being treated, for example, cardiovascular disease, are referred to as future *related* medical costs. Future *unrelated* medical costs, such as cancer treatment costs in the life-years gained, are a consequence of the life-extending nature of the treatment.[3] Studies show that increasing hospital survival leads to an increase in emergency admissions in patients whose lives were saved.[4] It is likely that this increase in admissions leads to an increase in medical costs.

The inclusion of future unrelated medical costs is a topic of debate in health technology assessment, with the US and the Netherlands recommending[5] or requiring[6] the inclusion of future unrelated medical costs in economic evaluation. Furthermore, researchers have previously argued that future unrelated medical costs should be stipulated to be included in guidelines for England and Wales, provided by the National Institute for Clinical Excellence (NICE),[7,8] an amendment to NICE's current guidelines, which state that any costs considered unrelated to the condition or technology of interest should be excluded.[9]

There are several arguments for the inclusion of future unrelated medical costs. First, to the extent that unrelated treatments are a firm commitment made by the healthcare system (may be less applicable in countries with less stable and comprehensive benefits packages, such as LMICs)[10] and given a fixed healthcare budget, extending life and thereby increasing future unrelated medical costs leads to health opportunity costs by leaving less budget for others in added life-years. By excluding future unrelated medical costs, the opportunity cost of these life-extending interventions is underestimated.[11] Second, excluding future unrelated costs generated by life-extending interventions, indirectly

introduces a bias against quality-of-life improving interventions, which do not add *future* costs. Third, estimates of quality-of-life and life expectancy are typically obtained from people receiving unrelated care; excluding costs and including benefits of unrelated future medical care is inconsistent.[3]

There is a pragmatic argument against the inclusion of future costs that is worth discussing here: the argument that future costs for all diseases would need to be separately modelled, thus the estimation of these costs is too complex to be carried out for every economic evaluation.[12-14], There are, however, methods facilitating the estimation of future unrelated medical costs[2,15], which have been applied in several countries, including England & Wales.[16-19] What all these methods have in common is that, rather than predicting the risk of all unrelated diseases and connecting these predictions to costs, they take per capita costs by age and gender, that comprise all medical spending, as a starting point. Per capita costs are then multiplied by survival curves to estimate incremental future unrelated medical costs. To avoid double-counting costs of related diseases, some studies have adjusted these per capita costs for related diseases.[17,20] A further issue concerning ageing and economic evaluation is that much of the increase in healthcare costs attributed to ageing can be attributed to someone being in their last year of life. This is referred to as 'time to death' (TTD)[21] and is most visible in hospital inpatient care, given the high cost of many inpatient treatments.[22,23] Previous studies have also considered that health spending is centred in the last phase of life, concluding that future unrelated medical costs are overestimated if one ignores TTD.[15,18]

Methods

Conceptual Model

In economic evaluations for NICE, an ICER is calculated to provide a measure of an intervention's cost-effectiveness against the threshold, *k*. The ICER in its basic form is written as $\frac{\Delta Costs}{\Delta QALYS}$, where the change in costs refers to a change only in related medical costs. As established in the introduction, however, interventions that extend survival implicitly generate future unrelated medical costs in the additional life-years gained. Therefore, the decision rule for cost-effectiveness, from a healthcare perspective, can be written as:

$$\frac{\Delta[\,L \times (C_r + C_u)]}{\Delta[L \times Q\,]} < k \tag{3.1}$$

L stands for life-years, Q for quality of life, and C_r and C_u for related medical costs and unrelated medical costs. k represents the cost-effectiveness threshold. Given that unrelated medical costs conditional on survival are independent of the intervention ($\Delta C_u = 0$), Equation 3.1 is rewritten as:

$$\frac{\Delta(L \times C_r) + \Delta L \times C_u}{\Delta[L \times Q\,]} < k \tag{3.2}$$

The difference in unrelated costs between intervention and comparator is solely dependent on the difference in life-years.

The variable of interest is the incremental future unrelated medical costs ΔL x C_u, which is denoted as Δlhc_u. The simplest way of estimating Δlhc_u is to use age-specific per capita health care spending and to multiply these with survivor curves in the treatment and comparator scenarios:

$$\Delta lhc_u = \sum_a l'(a) \times ac(a) - \sum_a l(a) \times ac(a) \tag{3.3}$$

Where *l'(a)* and *l(a)* denote the probability of surviving to age a in the treatment and comparator scenario respectively. $ac(a)$ indicates total annual health spending per capita at age *a*. This method has been proposed by Meltzer [2] and has the advantage that it is simple and data requirements are modest. However, if lifetime related costs are already included, then using Equation 3.3 leads to double counting of related costs. To overcome this, per capita health spending should be corrected so that only per capita costs of unrelated diseases (denoted $ac_u(a)$) are included (Eq. 3.4).

$$\Delta lhc_u = \sum_a l'(a) \times ac_u(a) - \sum_a l(a) \times ac_u(a) \tag{3.4}$$

Standardizing estimates

In order to remove the double counted related healthcare costs, per capita unrelated costs can be calculated in a standardized manner, using information on the related costs included in the original evaluation. To do this, total per capita costs can be treated simply as the sum of per capita related and unrelated costs:

$$ac_u(a) = ac(a) - ac_r(a) \tag{3.5}$$

Per capita related costs are often not directly available. However, they can be seen as the product of disease prevalence of disease *r* (denoted $p(a,r)$), and costs per patient for disease *r* related costs (denoted $ac_r(a \mid r)$).

$$ac_r(a) = p(a,r) \times ac_r(a \mid r) \tag{3.6}$$

Equation 6 provides a framework to adjust average costs per capita for costs of usual care for related diseases that are often included in an economic evaluation but also are part of *ac(a)*.

Related costs are anticipated to be small when evaluating most interventions, given the relatively small number of people with each disease in a population. Exceptions are particularly likely in some public health interventions.[17,24] Note that when end-of-life costs are provided in an economic evaluation these can also be used to adjust per capita costs as per patient costs for disease *r* are a weighted average of end-of-life costs and costs for those who are not in their last year of life:

$$ac_r(a \mid r) = [1 - m(a \mid r)] \times sc_r(a \mid r) + m(a \mid r) \times dc_r(a \mid r) \tag{3.7}$$

Here, $m(a \mid r)$ denotes the mortality rate at age *a* conditional on having disease *r*, $dc_r(a \mid r)$ denotes end-of-life/decedent costs for disease *r* conditional on having disease *r* and $sc_r(a \mid r)$ represents survivor costs, conditional on having the disease.

It is always a possibility that the participants in the intervention trial are not average consumers of health care, for example due to co-morbidities. Some diseases are known to be causally related, and thus it is expected that average health care costs for those with co-morbidities would be higher than those of an 'average' individual. In these cases, unrelated costs can be updated by obtaining co-morbidity specific costs that are not defined as related costs, separately for survivors and decedents if possible, and then adding these co-morbidity costs to the unrelated cost.

It can also be beneficial to adjust for TTD, by disaggregating individual future unrelated medical costs, which is labelled as lhc_u, into the sum of survivor,

$sc_u(a)$ and decedent, $dc_u(n)$, unrelated medical costs, where survivor costs are costs at each age, *a,* excluding the age at which the individual dies and decedent costs are costs incurred in the last year of life (Eq. 3.8). b is the age at which the intervention is implemented, and n is the age at which the individual dies.

$$lhc_u = \sum_{a=b}^{n-1} sc_u(a) + dc_u(n) \tag{3.8}$$

Average unrelated medical costs by age therefore need to be split into survival costs, $sc(a)$, and decedent costs, $dc(a)$. This is shown in Equations 3.9 to 3.11. Average medical costs, $ac(a)$, are a weighted average of decedent and survivor costs in a certain year. Total survivor and decedent costs from the provided average costs are calculated using mortality rates, m, and the ratio of medical costs between those dying and surviving, $\Phi(a)$. This decedent-survivor cost-ratio is taken from previous literature, in which health care expenditure panel data is combined with TTD and age information to estimate these ratios[23]. Given equations 3.9 and 3.10, which provide the decomposition of $ac(a)$ and the definition of Φ respectively, we can derive $sc(a)$, thereby facilitating the calculation of the aforementioned weighted average.

$$ac(a) = (1 - m(a)) \times sc(a) + m(a) \times dc(a) \tag{3.9}$$

$$dc(a) = \phi(a) \times sc(a) \tag{3.10}$$

$$sc(a) = \frac{ac(a)}{1+(\phi(a)-1)\times m(a)} \tag{3.11}$$

Data

For present purposes, this paper takes a healthcare perspective, aligned with NICE's brief. NICE is charged to appraise cost-effective use, covering NHS procedures and Personal and Social Services (PSS).[25] Given that PSS does not cover all long-term care options, long-term care data is not included. Average per capita healthcare spending data estimated by Asaria *et al.* (2017) is used,[26] who used administrative Hospital Episode Statistics (HES) data[27] from 2011 along with aggregate data on the number of GP visits in a year. These per capita data are available for gender and each age up until '85+' — an average for all

ages above 84. Costs are available for three sectors: inpatient care, outpatient care, and GP and pharmaceutical spending (Figure 3.1). The data are further smoothed using cubic splines. For mortality data, 2011 statistics from the Office of National Statistics (ONS) are used,[28] which provides population and cause of death figures for England and Wales by age and gender.

Decedent-survivor cost-ratios estimated by Howdon *et al.* (2018)[23] are used to adjust for TTD. The authors used HES data from years 2005/6 to 2011/2. Ratios are available for inpatients age 50 and upwards. It is assumed that ages below 50 take the ratio provided for age 50. For outpatient and GP/pharmaceutical expenditure the decedent-survivor ratio is assumed to be 1:1.

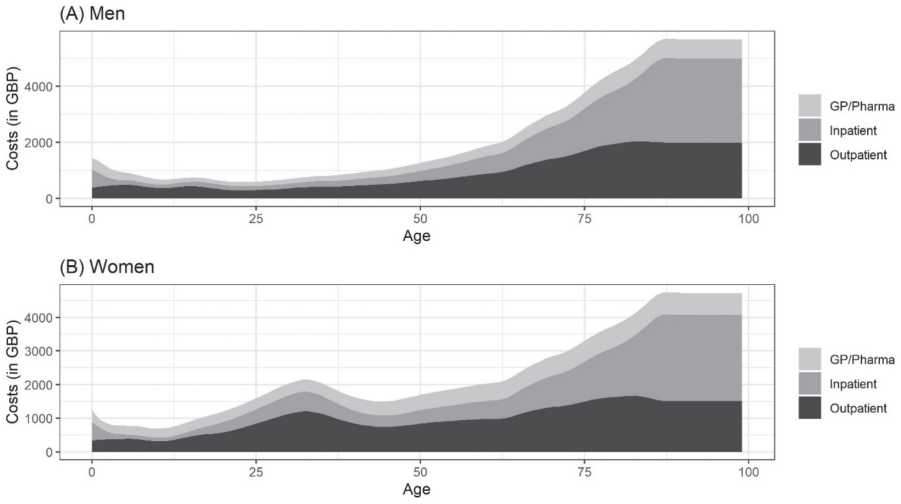


Figure 3.1 – Average Medical Costs by Sector. Costs are adjusted for 2018 price levels.

Cases

Before delving into disease-specific cases, future unrelated medical costs for the 'average' person are estimated, using a hypothetical intervention for which there are no intervention costs and only future costs — for example saving someone from a car accident. This is applied to all ages and demonstrates the cost-effectiveness ratio of saving a life. Here, saving someone's life has no future related medical costs. Average future unrelated medical costs are

summed for each age and divide by QALYs gained. To calculate QALYs we multiply survival by quality-of-life estimates, from Heijink et al. (2011).[29] They predict EQ-5D scores, by gender and age using UK value sets. For all cases, discount rates of 3.5% for both outcomes and costs are used, as recommended by NICE.

The first case based on existing research is Osimertinib — a medication used to treat non-small-cell lung carcinomas.[30] The study by Bertranou *et al.* (2018) compares Osimertinib to platinum-based doublet chemotherapy (PDC) in patients age 62 and above. It was recommended by NICE in 2016, with a 1.54 QALY gain and a £41,705 per QALY ICER. The second case is the use of Midostaurin; a multi-targeted protein kinase. In the study by Tremblay *et al.* (2018),[31] Midostaurin with the standard of care (SOC) is compared to SOC for newly diagnosed acute myeloid leukemia (AML) adult patients, aged 48 years and above. There were life year gains of 1.67 and QALY gains of 1.47. It was recommended for reimbursement by NICE in 2018 with ICERs of £30,263 per life year and £34,327 per QALY. The third case is the use of transcatheter aortic valve implantation (TAVI) compared to medical management (MM). Van Baal et al. (2016) estimate survival curves from Watt *et al.* (2012) who found an ICER of £16,100.[32] They look at patients over the age of 80 and find a QALY gain of 1.24. TAVI was also recommended by NICE in 2017.

For the above cases the original studies' survival curves were extracted. Comparator future unrelated medical costs were subtracted from intervention future unrelated medical costs. Unrelated costs are combined with survival curves, assuming a starting age of 62 years for the Osimertinib case,[30] 48 years for the Midostaurin case, and 80 years for the TAVI case, adjusting for related costs (mentioned in the original literature), and TTD. By dividing this difference in costs by the difference in QALYs we are left with the increase in the ICER.

In the Osimertinib case, specific costs of end-of-life care are provided; Equation 3.7 can be used to estimate average related costs. In the TAVI case the co-morbidity of diabetes mellitus (DM) is adjusted for. It has been shown that 57% of patients who cannot undergo surgery for aortic stenosis suffered from prohibitive co-morbidities. Studies have found that approximately 36%

of those who have received TAVI have DM,[33] and that average costs for DM in the UK are approximately £3,500.[34] Using this information DM specific costs are calculated for this population (by multiplying average costs by prevalence) before adding them to unrelated medical costs. We assume that average UK costs are transferable to England and Wales.

Costs per patient provided in our cases [30,31] and prevalence data for England and Wales from the UK Prevalence Project (2015),[35] are used for cancer prevalence, while the NHS Health Survey for England 2017 is used for cardiovascular disease prevalence.[36] Population mortality rates for both of the cancers in our cases were accessed from Cancer Research UK (2016).[37,38] For cardiovascular disease, 2014 mortality rates from the British Heart Foundation Cardiovascular Mortality Statistics are used.[39]

Results

In this section, the case of saving a life is dealt with first. The upper graphs in Figure 3.2 show average future medical costs by age and gender, independent of disease. Average costs and estimated decedent and survivor costs are displayed separately for men and women. The figure shows that decedent costs are higher than survivor costs at all ages, and that future medical costs increase with age. Furthermore, survivor future medical costs deviate from average future medical costs from age 80 onwards i.e. when mortality rates substantially increase. The lower graphs show the change in future unrelated medical costs divided by the change in LYGs and QALYs when a life is saved for free at age a, for both genders. For example, saving a life at birth leads to an ICER of £1,300 per QALY while saving a life at age 80 gives an ICER of £8,000 per QALY. These graphs also show that adjusting for TTD has little impact on the ICERs.

The differences between all three cases' intervention and comparator for both survival rates (top) and future unrelated medical costs (bottom) are shown in Figure 3.3. For the Osimertinib case, there is a dramatic difference in survival in the first years between patients who received the intervention and those who received the comparator, peaking at approximately 0.6. In the Midostaurin case, the differences in survival are much smaller (~ 0.075) at the beginning between intervention and comparator and, the decline in these differences is

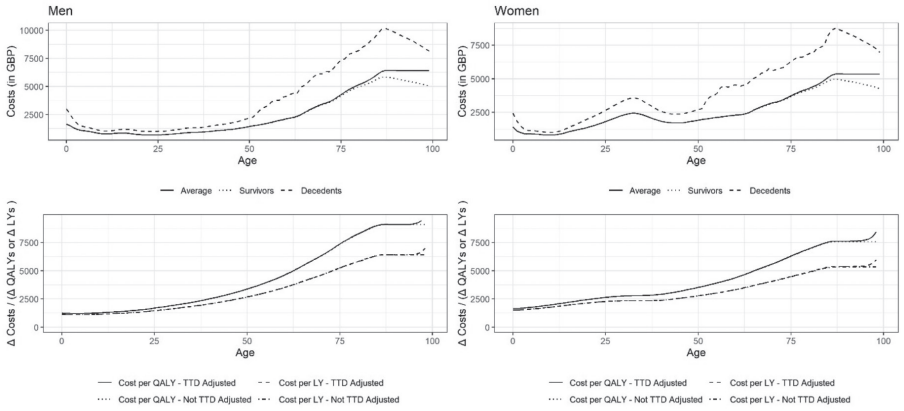


Figure 3.2 – Average, Survivor & Decedent Individual Medical Costs (top) & Saving a Life at Age a (bottom). Costs and outcomes in the lower graphs were discounted according to NICE guidelines — 3.5% discount rate for both costs and outcomes. Costs are adjusted for 2018 price levels.

more drawn out. The difference in survival between TAVI and MM is still substantial, peaking at approximately 0.3 at age 83. Looking at the lower two graphs; the difference in future unrelated medical costs between treatment and comparator, either adjusted or unadjusted for TTD. For all studies, the fact that survival in the treatment group is higher in the first years after treatment means that decedent costs are postponed by several years. This is shown clearly in Figure 3.3, where unrelated costs are larger for the comparator in the early years for TTD adjusted estimates; lower survival means higher expected decedent costs in the early years after treatment.

Table 3.1 shows the difference between ICERS including future medical costs where estimates are shown adjusted and unadjusted for TTD and double counting (i.e. excluding population average disease specific costs from the estimate of unrelated costs), once again discounted according to NICE guidelines. The estimates are shown along with the reported change in LYG and QALYs and the ratio between these two variables, as this is a further indicator of how large the impact of including future costs will be.[2] There is indeed an increase in all case ICERs. When looking at the results when adjusted for TTD and double counting, the ICER comparing Osimertinib with PDC increased by £5,112 (12%), the ICER for Midostaurin and SOC versus SOC increases by £3,167 (8%), and ICER for TAVI versus MM increases by £6,345 (37%). In all

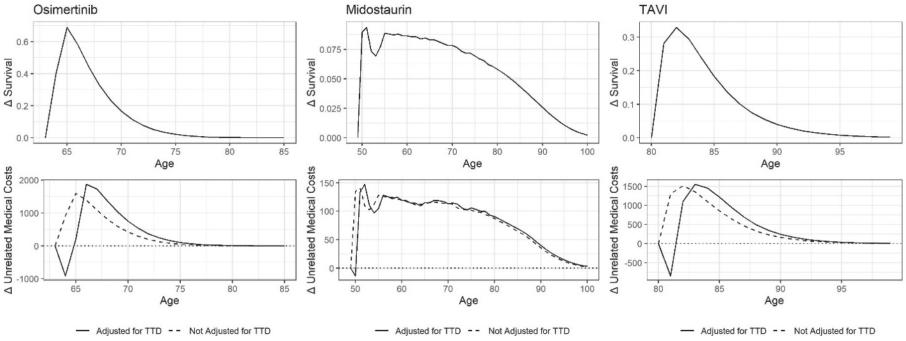


Figure 3.3 - Difference in Survival (top) and Unrelated Medical Costs (bottom) for all three cases. Costs and outcomes were discounted according to NICE guidelines — 3.5% discount rate for both costs and outcomes. Costs are adjusted for 2018 price levels.

cases the difference in the ICERs resulting from adjusting for double counting is modest. Table 3.1 also shows that adjusting for TTD changes the ICER by between roughly £1,000, and £2,507 in our cases. Furthermore, by adjusting for co-morbidities, TAVI costs increase by approximately an additional £1,200, in comparison to only adjusting for TTD.

Table 3.1 Difference in ICERs, Adjusted for TTD & Related Costs

Intervention	Osimertinib vs PDC	Midostaurin vs SOC	TAVI vs MM
Age at start of intervention	62 years	48 years	80 years
Change in LE (ΔL)	3.12	1.67	1.8
Change in QALY ($\Delta\lceil L \times Q \rceil$)	1.54	1.47	1.24
Change in LE/Change in QALY ($\frac{\Delta L}{\Delta[L \times Q]}$)	2.03	1.14	1.45
Reported ΔCost/ ΔQALY ($\frac{\Delta[L \times C_r]}{\Delta[L \times Q]}$)	£42,956	£38,033	£16,905
ΔCost / ΔQALY: Including future costs	£48,442 (5,486)	£41,434 (3,401)	£24,736 (7,831)
ΔCost/ ΔQALY: Including future costs, adjusted for TTD	£47,191 (4,235)	£40,760 (2,727)	£22,379 (5,474)
ΔCost/ ΔQALY: Including future costs, adjusted for TTD & Comorbidity*	~	~	£23,578 (6,673)

Intervention	Osimertinib vs PDC	Midostaurin vs SOC	TAVI vs MM
ΔCost/ ΔQALY: Including future costs	£48,418	£41,270 (3,237)	£24,076
adjusted for double counting	(5,463)		(7,171)
ΔCost/ ΔQALY: Including future costs,	£47,225	£40,594	£22,308
adjusted for double counting & TTD	(4,269)	(2,561)	(5,403)
ΔCost/ ΔQALY: Including future	~	~	£23,507
unrelated costs, adjusted for double			(6,602)
counting, TTD & Comorbidity*			

Note: Difference between actual and reported ICER shown in brackets. Costs and outcomes were discounted according to NICE guidelines — 3.5% discount rate for both costs and outcomes. Costs, including original ICERs, are adjusted for 2018 price levels.

** Only diabetes mellitus taken as a co-morbidity.*

Discussion

This paper has a dual purpose: firstly, to show that the inclusion of unrelated healthcare costs can have potentially significant policy-relevant implications for healthcare systems requiring a systems perspective and, secondly, to demonstrate the feasibility of a method of including them. In addition, given that economic evaluations are conducted for a large variety of medical interventions, it is beneficial to have a standardized approach to including unrelated future medical costs. This paper has provided such an approach, along with a complementary online tool (http://imta.shinyapps.io/PAIDUK). It shows the importance of future unrelated medical costs being included in economic evaluation and the impact of adjusting the calculations in order to take TTD and double counting into account.

By estimating the change in the ICER due to hypothetically saving a life at each age (Figure 3.2) we see that including future unrelated medical costs in economic evaluation leads to increases in the ICER. For example, if we were to save the life of a man (woman) at age 75, the increase in the ICER due to unrelated future costs would be around £7,500 (£6,250) per QALY, and that these changes to the ICER increase with age. These results mitigate the worry that including future unrelated medical costs in economic evaluation is particularly disadvantageous for diseases in children as we find increases in the ICER resulting from including these costs are lowest at the younger ages.

The results show that adjusting for double counting has a modest impact on our results because the interventions examined affect relatively small subsets of the population. This adjustment will be more important for public health interventions impacting larger populations. Adjusting for TTD had a substantial impact on the ICER in our case studies, with the larger effects showing in older populations, where death is relatively more expensive.

When comparing the three cases, there are a few further results worth noting: First, it appears that the older the target group, the larger the impact of including future unrelated medical costs. Given that costs are highest at older ages, increased survival in older target groups leads to comparatively higher differences in future unrelated medical costs between treatment and intervention. Second, in interventions with a target group with higher future medical costs than the population average, adjusting for relevant comorbidities leads to substantial increases in the ICER. This is unsurprising, given that additional (costly) comorbidities will cause unrelated medical costs to increase. Third, the ratio of change in LYG to change in QALYs is a further indicator of the impact of including future costs — the larger the ratio, the larger the impact. In other words, interventions where QALY gains were primarily driven by life extensions, were more affected by including these costs than interventions were QALY gains were driven by quality-of-life improvements.

Overall, the inclusion of future unrelated medical costs appears to have a considerable impact on the ICER. Given that reimbursement decisions are not based solely on cost-effectiveness but on a myriad of factors, we cannot say with certainty that increases in the ICER would influence specific reimbursement decisions. However, an increase of between seven and thirty percent in the ICER could be enough to change reimbursement decisions. The fact that increases in the ICER are not of the same magnitude between the cases used, shows that including future unrelated medical costs may lead to a shift in the hierarchy of which interventions are viewed as most cost-effective, mitigating bias towards life-extending interventions. Our results are presented in an online tool, in which our estimates of future medical costs can be accessed and adjusted for specific interventions, with options to adjust for TTD and double counting of related costs.

There are limitations to our approach. First, assuming that average medical costs are the same for every person within an age and gender group. While this can be somewhat rectified by subtracting related costs, there is the possibility that some patient groups will have different unrelated future costs, for example due to being too weak for certain treatments. Second, the data used has some restrictions, for example decedent-survivor ratios only being available from age 50 onwards. Furthermore, these are average ratios, covering all inpatient expenditure. Wong et al. (2010) show how drastically these ratios can differ from disease to disease in the Netherlands — for example ratios at age 50 for lung cancer and diabetes are approximately 1,000 and 7 respectively. The framework provided suggests adjusting for related costs before TTD, independent of whether related end-of-life costs are available, thereby assuming that the ratio of decedent-survivor costs is the same for both average and related costs. Third, it has come to our attention that medical expenditure data for England and Wales, when compared to similar data for the Netherlands, are low. Given that England and Wales and the Netherlands spend comparable proportions of their GDP on healthcare, it can be assumed that this is due to the collection of the data (bottom-up versus top-down) and long-term care not being included in our estimations. Our results for average unrelated future medical costs for England and Wales are in line with similar work by Briggs et al. (2018), suggesting that these differences are country-specific rather than solely attributable to our study. Fourth, 2011 data is used as a starting point for our costs; assuming that current spending patterns remain constant over time. Finally, we do not explicitly address the uncertainty around our estimates, which could stem from either survival gains or unit costs. Going back to the conceptual model presented in the methods section, specifically Equation 1; the addition of unrelated medical costs per QALY to the ICER can be written as $\frac{\Delta[L \times C_u]}{\Delta[L \times Q]}$. Due to life-years ($L$) being in both the numerator and denominator, uncertainty surrounding survival 'cancels out'. QALYs are provided by the cases used, and therefore our main source of uncertainty is in the unit costs themselves. As the original average costs are calculated from population-wide data, uncertainty is of relatively little concern here. However, there are still sources of uncertainty, specifically the age pattern of costs and decedent-cost ratios. Estimating these ratios for England and Wales is beyond the scope of this paper, however these are relevant and interesting avenues for future research.

As this is the first work to present a standardized option for the inclusion of future unrelated costs for England and Wales, there is much future research to be considered. It may be beneficial to test the assumption that during life-years gained, unrelated medical costs are equal to per capita average medical costs, using disease-specific patient data. Furthermore, previous literature[40] has provided an estimate of the NICE cost-effectiveness threshold, using supply-side data. They find marginal medical expenditure per QALY and suggest this as a threshold for NICE. It would be worth estimating the impact of the inclusion of future medical costs on this estimate, as excluding them would lead to inconsistency between ICER and threshold estimates.

To conclude, this paper provides an important methodological contribution by outlining how future unrelated medical costs can be included in health technology assessment. It also demonstrates how these methods apply for England and Wales and provides an online tool for doing so in practice.

References

1. de Meijer C, Wouterse B, Polder J, Koopmanschap M. The effect of population aging on health expenditure growth: A critical review. *Eur J Ageing*. 2013;10(4):353-361.
2. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ*. 1997;16(1):33-64.
3. de Vries LM, van Baal P, Brouwer W. Future costs in cost-effectiveness analyses: Past, present, future. *Pharmacoeconomics*. 2019;37:119-130.
4. Laudicella M, Martin S, Li Donni P, Smith PC. Do reduced hospital mortality rates lead to increased utilization of inpatient emergency care? A population-based cohort study. *Health Serv Res*. 2018;53(4):2324-2345.
5. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. *JAMA*. 2016;316(10):1093-1103.
6. Versteegh M, Knies S, Brouwer W. From good to better: New dutch guidelines for economic evaluations in healthcare. *Pharmacoeconomics*. 2016;34:1071-1074.
7. van Baal P, Morton A, Brouwer W, Meltzer D, Davis S. Should cost effectiveness analyses for NICE always consider future unrelated medical costs? *Br Med J*. 2017;359:j5096-j5096.
8. van Baal P, Morton A, Meltzer D, Brouwer W. Future unrelated medical costs need to be considered in cost effectiveness analysis. *Eur J Health Econ*. 2019:20(1):1-5.
9. Morton A, Adler AI, Bell D, et al. Unrelated future costs and unrelated future benefits: Reflections on NICE guide to the methods of technology appraisal. *Health Econ*. 2016;25(8):933-938.
10. Culyer AJ. Cost, context, and decisions in health economics and health technology assessment. *Int J Technol Assess Health Care*. 2018;34(5):434-441.
11. van Baal P, Meltzer D, Brouwer W. Future costs, fixed healthcare budgets, and the decision rules of cost-effectiveness analysis. *Health Econ*. 2016;25(2):237-248.
12. van Lier LI, Bosmans JE, van Hout HP, et al. Consensus-based cross-european recommendations for the identification, measurement and valuation of costs in health economic evaluations: A European Delphi study. *Eur J Health Econ*. 2018;19(7):993-1008.
13. Morris S, Devlin N, Parkin D. *Economic analysis in health care*. John Wiley & Sons; 2007.
14. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA*. 1996;276(15):1253-1258.
15. van Baal P, Wong A, Slobbe LCJ, Polder JJ, Brouwer W, de Wit GA. Standardizing the inclusion of indirect medical costs in economic evaluations. *Pharmacoeconomics*. 2011;29(3):175-187.

16. Briggs ADM, Scarborough P, Wolstenholme J. Estimating comparable English healthcare costs for multiple diseases and unrelated future costs for use in health and public health economic modelling. *Plos One*. 2018;13(5):e0197257-e0197257.
17. van Baal P, Hoogendoorn M, Fischer A. Preventing dementia by promoting physical activity and the long-term impact on health and social care expenditures. *Prev Med*. 2016;85:78-83.
18. Gandjour A, Lauterbach KW. Does prevention save costs?: Considering deferral of the expensive last year of life. *J Health Econ*. 2005;24(4):715-724.
19. Blakely T, Atkinson J, Kvizhinadze G, Nghiem N, McLeod H, Wilson N. Health system costs by sex, age and proximity to death, and implications for estimation of future expenditure. *N Z Med J*. 2014;127(1393).
20. van Baal P, Perry-Duxbury M, Bakx P, Versteegh M, van Doorslaer E, Brouwer W. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Econ*. 2019;28(1):87-100.
21. Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: A red herring? *Health Econ*. 1999;8(6):485-496.
22. Wong A, van Baal P, Boshuizen HC, Polder JJ. Exploring the influence of proximity to death on disease-specific hospital expenditures: A carpaccio of red herrings. *Health Econ*. 2011;20(4):379-400.
23. Howdon D, Rice N. Health care expenditures, age, proximity to death and morbidity: Implications for an ageing population. *J Health Econ*. 2018;57:60-74.
24. van Baal P, Feenstra TL, Polder JJ, Hoogenveen RT, Brouwer W. Economic evaluation and the postponement of health care costs. *Health Econ*. 2011;20(4):432-445.
25. McCabe C, Claxton K, Culyer AJ. The NICE cost effectiveness threshold: What it is and what it means. *Pharmacoeconomics*. 2008;2(9):733-44.
26. Asaria M. Health care costs in the English NHS: Reference tables for average annual NHS spend by age, sex and deprivation group. 2017:1-19.
27. Asaria M, Grasic K, Walker S. Using linked electronic health records to estimate healthcare costs: Key challenges and opportunities. *Pharmacoeconomics*. 2016;34(2):155-160.
28. Office for National Statistics (United Kingdom). Deaths by single year of age tables - UK. 2018.
29. Heijink R, Van Baal P, Oppe M, Koolman X, Westert G. Decomposing cross-country differences in quality adjusted life expectancy: The impact of value sets. *Population Health Metrics*. 2011;9(1):17.
30. Bertranou E, Bodnar C, Dansk V, Greystoke A, Large S, Dyer M. Cost-effectiveness of osimertinib in the UK for advanced EGFR-T790M non-small cell lung cancer. *J Med Econ*. 2018;21(2):113-121.
31. Tremblay G, Dolph M, Patel S, Brandt P, Forsythe A. Cost-effectiveness analysis for midostaurin versus standard of care in acute myeloid leukemia in the United Kingdom. *Cost Eff and Resour Alloc*. 2018;16(1):33-33.
32. Watt M, Mealing S, Eaton J, et al. Cost-effectiveness of transcatheter aortic valve replacement in patients ineligible for conventional aortic valve replacement. *Heart*. 2012;98(5):370-376.

33. Puls M, Bleckmann A, Jacobshagen C, et al. Diabetes increases short- and long-term mortality after transcatheter aortic valve implantation (TAVI). *Dtsch Med Wochenschr*. 2014;139(16):822-828.

34. Kanavos P, van den Aardweg S, Schurer W. *Diabetes expenditure, burden of disease and management in 5 EU countries*. LSE Health and Social Care; 2012.

35. 20-year cancer prevalence in the UK. Public Health & Macmillan. Accessed September 1, 2019. http://www.ncin.org.uk/about_ncin/segmentation.

36. Health survey for England 2017. Accessed September 1, 2019. National Health Service. https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2017.

37. Acute myeloid leukaemia (AML) statistics. Cancer Research UK. Accessed September 1, 2019. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml#heading-Zero.

38. Lung cancer statistics. Cancer Research UK. Accessed September 1, 2019. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer.

39. Cardiovascular disease statistics 2014. British Heart Foundation. Accessed September 1, 2019. https://www.bhf.org.uk/informationsupport/publications/statistics/cardiovasculardisease-statistics-2014.

40. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the national institute for health and care excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14):1-503.

CHAPTER 4

The relevance of including future health care costs in cost-effectiveness threshold calculations

With Meg Perry-Duxbury, James Lomas, Miqdad Asaria & Pieter van Baal.

Submitted

Abstract

The supply-side threshold for the UK NHS has been empirically estimated as the marginal returns to health care spending on health outcomes. These estimates implicitly exclude future health care costs, which is inconsistent with the objective of making the most efficient use of health care resources. This paper illustrates how empirical estimates of the threshold within health care can be adjusted to account for future health care costs. Using cause-deleted life tables we illustrate how such estimates can be adjusted. While the effect of including future health care costs can have substantial effects on ICERs of specific life-extending interventions, we find that including future costs has relatively little impact (an increase of £743 per QALY) on the threshold estimate. This implies that for some life-extending interventions the impact of including future costs on whether an intervention is deemed cost-effective may be considerable.

Introduction

A key criterion for deciding whether to reimburse a health care intervention is whether the estimated incremental cost-effective ratio (ICER) lies below the relevant cost-effectiveness threshold.[1] In England and Wales, the National Institute for Health and Care Excellence (NICE) makes explicit use of such a threshold.[2] NICE's 2013 guidelines state that: '*A technology can be considered to be cost effective if its health benefits are greater than the opportunity costs of programmes displaced to fund the new technology, in the context of a fixed NHS budget*'.[3] However, the current threshold used by NICE, which ranges from £20,000 to £30,000 per QALY, reflects values implied by previous decisions made by NICE rather than an explicit consideration of the evidence regarding the potential health foregone caused by implementing new health technologies.[4]

In theory, estimating the health foregone of displaced care, also referred to as the supply-side or *k*-threshold[5] when expressed per life year or QALY, requires us to know precisely which activities would be displaced at the margin and estimate the health benefits forgone in displacing these activities. These thresholds could be estimated through (detailed) 'league tables', which rank interventions based on cost-effectiveness. However, this would require knowledge regarding the cost-effectiveness of all currently funded interventions. Additionally, we would need to assume that a new intervention would be funded by displacing the least cost-effective interventions on the list, which may not be what happens in practice. An alternative, empirically practical, approach is to estimate the cost per QALY of health care expenditure at the margin,[5-9] which provides an estimate of the average cost-effectiveness of unspecified interventions (potentially) displaced or expanded at the margin.

Interventions displaced or expanded at the margin may be quality-of-life-improving, life-extending, or both. When an intervention has a life-extending element the inclusion of future health care costs is necessary for a consistent estimate of cost effectiveness and an efficient allocation of the health care budget.[10-14] For example, when a patient who is treated for an otherwise fatal heart-attack gains additional life-years, incurs additional related health care costs (e.g. cardiologist check-ups), but also incurs additional unrelated health care costs (e.g. cancer or dementia treatment). This is true not only for

new treatments, but also for existing (and potentially replaced) treatments. Hence, estimates of the supply-side threshold should include both related and unrelated future health care costs. Existing analyses estimating these supply-side thresholds have ignored both related and unrelated future costs.[7] In this paper we aim to improve upon previous work by describing how such empirical estimates of the supply side threshold can be adjusted to account for future health care costs and how this impacts the cost-effectiveness results of interventions when future costs are also included in economic evaluation.

Methods

Conceptual Model

Assuming the aim of health maximization and a fixed health care budget, an intervention can be implemented if its incremental cost-effectiveness ratio (ICER) is below the supply-side threshold, k. Both the ICER and k can be viewed as a ratio of marginal costs to marginal benefits. For the ICER, the incremental costs and benefits are triggered by a specific intervention for a known patient group (denoted with an asterisk *), while for k they are triggered by the range and mix of interventions for various patient groups that are displaced at the margin. We can write the decision rule of the ICER needing to be less than k as:

$$\frac{\Delta Costs^*}{\Delta QALYs^*} < \frac{\Delta Costs}{\Delta QALYs} \tag{4.1}$$

Interventions that improve survival, implicitly generate both future related and unrelated medical costs in life years gained.[10,15-17] By making a distinction between length of life and health spending conditional on being alive, we can rewrite the cost-effectiveness decision rule as:

$$\frac{\Delta[\, L^* \times (c_r^* + c_u^*)]}{\Delta QALYS^*} < \frac{\Delta[\, L \times (c_r + c_u)]}{\Delta QALYs} \tag{4.2}$$

In which L denotes life-years and c_r and c_u related and unrelated medical costs per year, respectively. Given that unrelated medical costs conditional on survival are independent of the intervention ($\Delta c_u = 0$), Equation 4.2 can be rewritten as:

$$\frac{\Delta(L^{*}\times c_{r}^{*})+\Delta L^{*}\times c_{u}^{*}}{\Delta QALYs^{*}} < \frac{\Delta(L\times c_{r})+\Delta L\times c_{u}}{\Delta QALYs} \quad (4.3)$$

The difference in unrelated costs between intervention and comparator is solely dependent on the difference in life-years. Excluding future unrelated costs from estimations of k and the ICER implicitly assumes that the incremental unrelated health care expenditures are zero or net out to zero, which seems unrealistic. Note also, in this context, that consistently excluding unrelated costs from both sides of equation 4.3 leads to an inefficient use of the healthcare budget and thus to health losses.[12]

Updating the threshold

Claxton et al.[7] calculate the cost-effectiveness threshold using estimates of the effect of an annual one percent increase in the total health care budget on annual mortality rates. Expenditure and outcome elasticities were estimated using an instrumental variables approach. These elasticities were used to calculate the cost of a life-year in each programme. The calculated life-years gained from the increase in the health budget were then adjusted for quality of life, taking account of gender, age and disease. The increase in total NHS spending, divided by the sum of all QALY changes across all Programme Budgeting Categories (PBCs), was presented as the relevant threshold. As Claxton et al. only included changes in health spending in the first year, while estimating lifetime health benefits, all future costs were missing from their estimates. Consequently, we aimed to update their estimates to account for future costs:

$$k^{\dagger} = k + \frac{\Delta L\times(c_{r}+c_{u})}{\Delta QALYs} \quad (4.4)$$

Where k denotes the Claxton et al. threshold estimate and $k^{\dagger}$ the threshold including future costs of both related and unrelated diseases. The life table calculations for estimating future health care costs were carried out for each separate PBC, where future costs are derived from previous work by Perry-Duxbury et al.[16] The increase in total health spending is distributed over the different PBC using the expenditure elasticities. The disease-specific impacts of this increase on cause-specific mortality (used in the cause deleted life tables to estimate life-years gained), were calculated using the outcome elasticities

estimated by Claxton et al. The comparator being a counterfactual in which the budget did not increase. Note that outcome elasticities were not available for all disease categories, mostly because of little to no mortality effects in the specific PBC. For these PBCs, future health care costs do not impact the threshold. To translate estimates expressed in pounds per life year gained to pounds per QALY gained, we used the PBC specific ratios of incremental costs per life-year to costs per QALY from Claxton et al., who used PBC specific quality of life decrements by age and gender. We translate life-years to QALYs using the aforementioned ratios (Table 4.1).

Table 4.1 Cost per LY/Cost per QALY Ratios from Claxton et al. used to translate life-years

	Cost per LY	Cost per QALY	Cost per LY / Costs per QALY
Infectious diseases	61,425	20,829	2.95
Cancer	11,931	16,997	0.70
Endocrine	38,122	3,124	12.20
Neurological	92,282	5,480	16.84
Circulatory	6,544	7,038	0.93
Respiratory	28,528	1,998	14.28
Gastrointestinal	12,983	7,293	1.78
Genitourinary	141,746	43,813	3.24
Maternity & Neonates	1,608,817	2,969,208	0.54

Note: Disease-specific QoL used. Only for disease categories for which there is an observed mortality effect.

Overall threshold estimates were calculated using weights based on the percentage share of health effects provided by Claxton et al. The equations and further explanation of the life-tables approach are presented below.

Estimating future costs and life-years

In both the situations in which the budget increases for PBC which includes health for related diseases r and the counterfactual the lifetime expected discounted costs, $C(b,r)$, and life-years, $L(b,r)$, for a person age b, can be estimated using the formulas below.

$$C(b,r) = \sum_{a=b}^{100}(1.035^{a-b} \times c(a) \times \prod_{b}^{a} e^{-m(a,r)-m(a,u)}) \quad (4.5)$$

$$L(b,r) = \sum_{a=b}^{100}(1.035^{a-b} \times \prod_{b}^{a} e^{-m(a,r)-m(a,u)}) \quad (4.6)$$

Where $C(b,r)$ represents remaining lifetime health care costs for someone age b treated in PBC with related diseases r and $L(b,r)$ represents remaining life-years for someone age b treated in PBC with related diseases r. $c(a)$ represents yearly per capita costs by age and $m(a,r)$ and $m(a,u)$ represent related and unrelated mortality rates by age, where a is the index for age.

We assume that yearly per capita costs depend on age and time to death.

$$c(a) = [m(a,r) + m(a,u)] \times c(a|m = 1) + [1 - m(a,r) - m(a,u)] \times c(a|m = 0) \quad (4.7)$$

Here $c(a|m = 1)$ denotes yearly costs at age a conditional on dying at age a while $c(a|m = 0)$ denotes yearly costs for those who do not die at age a. These are derived from previous work by Perry-Duxbury et al. (2020). Note that we do not make a distinction between related and unrelated future costs as neither were included in the Claxton et al. study. When expenditure increases, related mortality rates, $m(a,r)$, are multiplied by outcome elasticities (which are negative) and 1.01 - reflecting the one percent increase in expenditure (Eq. 4.8)[7]. The lifetime costs and life years from the budget increase, $C'(b)$ and $L'(b)$ are calculated by replacing disease related mortality with, $m'(b)_r$, where ' indicates the situation in which expenditure increases. To be consistent with Claxton et al. we assume a one-year change in mortality.

$$m'(b,r) = m(b,r) \times [1 + 0.01 \times o(r)] \quad (4.8)$$

Where $o(r)$ represents the outcome (mortality) elasticity for PBC containing spending for disease r.

Estimating disease specific and NHS-wide thresholds

We use PBC specific ratios of costs per life-year to costs per QALY gained from Claxton et al. to translate costs per life year gained to costs per QALY gained. Thus, QALYs for the situation in which expenditure increases and the counterfactual (denoted $Q'(b)^*$ and $Q(b)^*$) are calculated by multiplying the respective remaining life-years by these ratios. We then estimate both PBC

specific thresholds, $k(r)^{\dagger}$, and a subsequent threshold for all NHS spending, $k^{\dagger}$, which is arguably the most policy-relevant estimate.

$$k(r)^{\dagger} = k(r) + \frac{\sum_{b=0}^{100} pop(b) \times [\left(C'(b,r) - C(b,r)\right) \times p(b,r)]}{\sum_{b=0}^{100} pop(b) \times [\left(Q'(b,r) - Q(b,r)\right) \times p(b,r)]} \tag{4.9}$$

$$k^{\dagger} = \sum_{r \in D} (k(r)^{\dagger} \times w(r)) \tag{4.10}$$

In the above equations, k represents the Claxton et al. threshold estimate while $k^{\dagger}$ represents threshold estimates including future costs. $p(b,r)$ is the percentage of patients in age group b treated in PBC with related diseases r, D represents the set of all PBCs, and $w(r)$ represents the weighting taken as percentage share of total health effects for a PBC with related diseases r.[7]

Results

The impact of including future costs on PBC specific threshold estimates varies quite a bit. For example, neurological and endocrine diseases sit at the low end of the spectrum, while cancer has far more future health care costs attributed to it. The first reason for this is the difference between QALYs and LYs gained within a PBC, which are provided in Table 4.1. PBCs for neurological and endocrine disorders have relatively high ratios of cost per LY to costs per QALY, meaning that an increase in the health care budget leads to more quality-of-life gains than life-year gains. The opposite is seen in diseases such as cancer and circulatory disease where most health gains are due to increases in length of life A second reason is that the older the patient population within a PBC, the more future costs they will incur. This is due to medical consumption increasing with age and discounting — the higher the age at which one receives an intervention, the less the corresponding future costs are discounted. Thus, it is of little surprise that diseases such as cancer and circulatory disease have higher adjusted thresholds, while diseases that tend to affect the young, like infectious diseases, only see future health care cost additions of £510.

The final row of Table 4.2 shows the difference in the cost-effectiveness threshold when calculated for the full NHS budget — a one percent increase in the NHS budget divided by the sum of all life years or QALYs. This difference

Table 4.2 The addition of future costs to PBC specific and overall threshold estimates (£'s)

A. Cost per LY				B. Cost per QALY (PBC specific QoL†)			
	Original Estimate	Addition of future costs	Final Estimate		Original Estimate	Addition of future costs	Final Estimate
Infectious Diseases	61,425	1,502	62,927	Infectious Diseases	20,829	510	21,339
Cancer	11,931	2,508	14,439	Cancer	16,997	3,574	20,571
Blood	-	-	0	Blood	9,419	-	9,419
Endocrine	38,122	2,205	40,327	Endocrine	3,124	181	3,305
Mental Health	-	-	0	Mental Health	18,744	-	18,744
Neurological	92,282	1,816	94,098	Neurological	5,480	108	5,588
Vision	-	-	0	Vision	45,788	-	45,788
Hearing	-	-	0	Hearing	6,239	-	6,239
Circulatory	6,544	2,957	9,501	Circulatory	7,038	3,181	10,219
Respiratory	28,528	2,711	31,239	Respiratory	1,998	190	2,188
Gastrointestinal	12,983	2,111	15,094	Gastrointestinal	7,293	1,186	8,479
Skin	-	-	0	Skin	101,042	-	101,042
Musculoskeletal	-	-	0	Musculoskeletal	15,628	-	15,628
Genitourinary	141,746	3,058	144,804	Genitourinary	43,813	945	44,758
Maternity and Neonates	1,608,817	953	1,609,914	Maternity and Neonates	2,969,208	1,758	2,970,966
Total‡	25,214	2,673	28,403	Total‡	12,936	743	13,679

Notes: costs and outcomes are both discounted by 3.5% as recommended by NICE Guidelines. † QoL adjusted for PBC-specific decrements by age and gender. ‡ percentage share of total health effects is used as weighting for all total estimates. Table A uses share of change in Net Years of Life Lost, Table B uses share of change in disease-specific QALYs.

is a weighted average of the PBC specific threshold estimates, leading to an addition of £743 to the overall threshold. The impact of future health care costs on the overall threshold estimate is limited because much of the share of total health effects is attributed to respiratory disease (30%), neurological disease (14%) and circulatory disease (14%), with only the latter of the three PBCs having high future costs per QALY. Additionally, some shares of total health effects are attributed to PBCs with no mortality effects, such as mental health (12%).

Discussion

We find that the impact of including future costs on threshold estimates is limited, which is due to the fact that, in the Claxton et al. paper, the marginal returns to health spending were, to a large extent, driven by improvements in quality of life as opposed to increases in length of life. Given that ICERs for specific interventions might be affected more heavily by the inclusion of future costs, these findings have important implications for decisions based on cost effectiveness. This implies that interventions which are life improving rather than life-extending, and interventions on younger patients are less likely to be reimbursed when analyses, unadjusted for future costs, are conducted. For instance, in our earlier work[16] we found that including future costs in three economic evaluation cases led to increases in the ICER of between £3,200 and £7,200. These increases in ICERs are much greater than the increase in the threshold.

We estimate adjusted PBC specific thresholds as well as an adjusted overall threshold. The updated overall threshold was £13,679 per QALY, £743 higher than the original estimate of £12,936. PBC specific estimates of the adjustment ranged from £108 per QALY, to £3,574 per QALY when using disease and age specific quality of life estimates. PBCs with the highest threshold varied depending on the denominator of the estimate. For example, the PBC with the largest addition to the threshold was genitourinary disease (£3,058) when dividing costs by *life-years* gained and cancer (£3,574) when dividing by QALYs. These differences are solely due to the ratios between cost per life-year and cost per QALY taken from Claxton et al., some of which seem somewhat surprising. For example, regarding respiratory diseases, the ratio suggests that, given a one percent increase in the budget, increases in QALYs

are fourteen-fold the increases in life-years. It seems worthwhile for future studies to further investigate these influential relationships.

The overall threshold is arguably the most relevant estimate for decision-makers, as resource allocation decisions in the NHS are made centrally across the entire healthcare budget rather on a per disease basis. [18] [2] The overall adjusted threshold we estimate (£13,679), is higher than that of the unadjusted threshold estimate (£12,936) (Claxton et al.) but substantially lower than the threshold range used in practice by NICE.[3]

There are limitations to our approach. First, in this demonstrative case study, we only update the point-estimate of the empirical NHS threshold. This is due to most of the inputs for our calculations, such as population health, disease prevalence and mortality being stable. Possible sources of uncertainty are the (disease-specific) outcome elasticities estimated by Claxton et al., however investigating this goes beyond the scope of our paper. Second, we were only able to update the PBC specific thresholds for PBCs for which mortality elasticities were available from the existing empirical work, and thus our overall threshold estimate may be an underestimation, if the increase in expenditure did indeed affect mortality of PBCs for which elasticities were not available. That being said, the majority of QALYs gained from the excluded PBCs were due to increases in quality of life, thus the effect is likely to be modest. Finally, we do not divide the impact of the increase in health spending on future costs into related and unrelated costs. If an increase in the budget turns out to increase quality of life and/or lead to savings in related costs, then our adjusted threshold value would be an overestimate.

As NICE and other HTA bodies around the world review the methods for conducting economic evaluations, our evidence suggests that it is time for them to consider using an empirically estimated supply side threshold and include future health care costs not only in cost-effectiveness analyses but also in the threshold that the results of these analyses are judged against. Such an approach would help to fully capture the threshold associated with new health technologies and to maximise the health gains that can be achieved from any given level of health spend. This is particularly relevant at a time when health budgets are being increasingly stretched by the coronavirus pandemic.

[2] An exception to this rule may be the NHS Cancer Drugs Fund which pays for cancer drugs that would otherwise be rejected by NICE for being too expensive.

References

1. Brouwer W, van Baal P, van Exel J, Versteegh MM. When is it too expensive? cost-effectiveness thresholds and health care decision-making. *Eur J of Health Econ*. 2019;20(2):175-180.
2. Schwarzer R, Rochau U, Saverno K, et al. Systematic overview of cost–effectiveness thresholds in ten countries across four continents. *J Comp Eff Res*. 2015;4(5):485-504.
3. NICE UK. *Guide to the methods of technology appraisal*. NICE. 2013.
4. Appleby J, Devlin N, Parkin D, Buxton M, Chalkidou K. Searching for cost effectiveness thresholds in the NHS. *Health Policy*. 2009;91(3):239-245.
5. van Baal P, Perry-Duxbury M, Bakx P, Versteegh M, van Doorslaer E, Brouwer W. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Econ*. 2019;28(1):87-100.
6. Vallejo-Torres L, García-Lorenzo B, Castilla I, et al. On the estimation of the cost-effectiveness threshold: Why, what, how? *Value in Health*. 2016;19(5):558-566.
7. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the national institute for health and care excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14):1-503.
8. Martin S, Rice N, Smith PC. Comparing costs and outcomes across programmes of health care. *Health Econ*. 2012;21(3):316-337.
9. Martin S, Rice N, Smith PC. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *J Health Econ*. 2008;27(4):826-842.
10. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics*. 1997;16(1):33-64.
11. Rappange DR, van Baal P, van Exel J, Feenstra TL, Rutten F, Brouwer W. Unrelated medical costs in life-years gained: Should they be included in economic evaluations of healthcare interventions? *Pharmacoeconomics*. 2008;26(10):815-830.
12. van Baal P, Meltzer D, Brouwer W. Future costs, fixed healthcare budgets, and the decision rules of cost-effectiveness analysis. *Health Econ*. 2016;25(2):237-248.
13. Morton A, Adler AI, Bell D, et al. Unrelated future costs and unrelated future benefits: Reflections on NICE guide to the methods of technology appraisal. *Health Econ*. 2016;25(8):933-938.
14. van Baal P, Morton A, Brouwer W, Meltzer D, Davis S. Should cost effectiveness analyses for NICE always consider future unrelated medical costs? *Br Med J*. 2017;359:j5096-j5096.
15. de Vries LM, van Baal P, Brouwer W. Future costs in cost-effectiveness analyses: Past, present, future. *Pharmacoeconomics*. 2019;37:119-130.

16. Perry-Duxbury M, Asaria M, Lomas J, van Baal P. Cured today, ill tomorrow: Including future unrelated medical costs in economic evaluations for England & Wales. *Value in Health*. 2020;23(8):1027-1033.
17. Kellerborg K, Perry-Duxbury M, de Vries LM, van Baal P. Practical guidance for including future costs in economic evaluations in the Netherlands: Introducing and applying PAID 3.0. *Value in Health*. 2020.
18. Maynard A, Bloor K. The economics of the NHS cancer drugs fund. *Applied health economics and health policy*. 2011;9(3):137-138.

CHAPTER 5

The value of safety and its relationship with the evaluation of health interventions: a literature review

With Job van Exel & Werner Brouwer.
Published in: The European Journal of Health Economics. 2019;20(7):1041-61.

Abstract

Improving (feelings of) safety is an important goal of many health systems, especially in the context of recurrent threats of pandemics, and natural disasters. Measures to improve safety should be cost-effective, raising the issue of how to value safety. This is a complex task due to the intangible nature of safety. We aim to synthesize the current empirical literature on valuing of safety to gain insights into current methodological practices. After a thorough literature search in two databases for papers from the fields of life sciences, social sciences, physical sciences and health sciences that empirically measure the value of increasing safety, 33 papers were found and summarized. The focus of the research was to investigate the methodologies used. Attention was also paid to theoretical papers and the methodological issues they present, and the relationship between safety and three categories of covariate results: individual characteristics, individual relationship with risk, and study design. The field of research in which the most papers were found was environmental economics, followed by transportation and health. There appeared to be two main methods for valuing safety: Contingent valuation and Discrete Choice Experiments, within which there were also differences—for example the use of open or dichotomous choice questions. Overall this paper finds that there still appears to be a long way to go before consensus can be attained about a standardised methodology for valuing safety. Safety valuation research would benefit from learning from previous experience and the development of more standardised methods

Introduction

Many of today's societies are governed by rules, regulations and protocols, many of which are designed with the aim of keeping citizens 'safe'. Safety can be defined as 'the condition of being protected from or unlikely to cause danger, risk or injury'.[1] With recurrent news about threats of global warming, terrorist attacks, pandemics and natural disasters, it is no surprise that safety is a significant concern for citizens, companies, and governments. All wish to minimize the possibility of death, damage, illness or injury. However, a question that is increasingly relevant in these same societies is whether policies that aim to increase the safety of citizens, not only in the health sector, but also in for instance the transport or environmental sector, provide good value for money. After all, public money can be spent only once and investments in increased safety displace other (worthwhile) investments. In order to evaluate the efficiency of these policies, safety needs to be assigned a value. Due to safety being an intangible, non-monetary good, economists tend to consider risk- or uncertainty-reduction instead of 'safety',[2-4] with risk-reduction being the most tangible and therefore the most applied definition in the literature. This being said, there is no 'golden standard' for safety valuation. Early approaches were based on life insurance premiums, which were then replaced, initially by human capital methods, and more recently by stated preference methods.[5] This ongoing shift in approaches shows that valuing safety is a field in which methods are frequently evolving.

Research into the topic of valuing safety is scarce, scattered across scientific fields, and no review of safety valuation literature is currently available. However, (the value of) safety is likely to become increasingly important in health (economics) and beyond. Large scale surveillance systems to prevent or mitigate the consequences of pandemics by early detection of outbreaks and early determination of their causes are an example of improving safety. Other examples with direct health consequences are improved safety by stricter regulations for food production, hospital procedures or air pollution. In evaluating such measures and policies the value of safety may be a crucial element, but little is known as to how to best capture it.

Therefore, the aim of this paper is to present a review of the existing literature; synthesizing the methodologies used in empirical research papers that value

safety. The reviewed papers come from different scientific fields, including environmental economics, transport economics, food safety, crime, and health economics — indicating that the results presented in this paper may be beneficial to any future research that requires safety valuation. As the direct outcomes from these various fields are incomparable (e.g. the value of reduced risk of flooding versus the value of reduced risk of train accidents), the focus of this study is on the methodology of valuation and the characteristics of respondents, context and study design associated with elicited values of safety, as these are the most comparable aspects of the papers. Subsequently, we will emphasize the implications for valuing safety in the context of health.

The main aim of this paper is to give a review of the methods used in *empirical* research on safety. Such empirical research should be embedded in theoretical research on valuing safety, and also the interpretation of empirical studies ideally is informed by such theoretical insights. Therefore, the structure of this paper is as follows. First, section 2 discusses the theoretical background to the valuation of safety. Thereafter, in section 3, the methods of the literature search are discussed, followed by the findings of the research (section 4). Finally, we discuss the results with a special focus on lessons for valuing safety in health.

Theoretical Background

One of the ways to compare alternative policies or interventions is by applying a (form of) cost-benefit analysis (CBA), in which the costs and benefits of the alternatives in question are compared between and within said alternatives.[6] In order to compare the benefits from interventions that differ in outcome — for example an improvement in road safety versus an improvement in city air quality — these benefits must be expressed in a comparable metric, traditionally often in monetary terms. Even in health care, where other outcome measures are sometimes used, like Quality-Adjusted Life-Years to express health outcomes in cost-utility analysis, other costs and benefits are typically expressed in monetary terms.

When taking an often-advocated societal perspective in the evaluation,[7] all costs and benefits need to be included in the evaluation regardless of where or when they fall in society. If some of the benefits (not included in QALYs)

involve non-marketed goods, these goods need to be included and hence valued. The two main approaches of assigning monetary value to non-market goods are revealed and stated preference. The revealed preference approach uses observed prices and choices to derive the value of a given outcome, while the stated preference approach elicits preferences from hypothetical choices, for instance through surveys or choice experiments, to measure how an individual values the chosen non-market good.[8] Using stated preferences is more common in valuing non-market goods, as it is hard to find real world observations from which revealed preferences can be derived univocally. The most common types of stated preference studies used to value non-market goods are contingent valuation (CV) studies and discrete choice experiments (DCE). CV studies directly ask individuals their value, in terms of willingness to pay (WTP), for some non-market good, given a certain hypothetical scenario,[9] whereas DCEs also use a hypothetical scenario, but ask respondents to choose between options with several different attributes in order to indirectly extract their valuation.[10]

In any valuation, three aspects are crucial: (i) what is being valued, (ii) how it is being valued and (iii) who is valuing the good on offer. These three aspects are briefly addressed below.

In terms of *what* is being valued, in the instance of safety valuation, 'safety' is very complex to define, and therefore is can be easier to think of an improvement of safety being a reduction of risk of some adverse event occurring, a reduction of uncertainty or the reduction of impact of a specific incident which is perceived to be unsafe. However, even with a more tangible definition of safety, several issues still arise when trying to value it. A first issue relates to safety itself and it is that being protected has an objective and a subjective element. An example of the difference can be found in situation where objective crime figures are going down, but subjective feelings of safety do not improve. From a utilitarian perspective, one may claim that there can be value in both improving objective safety (fewer victims, less damage) and subjective safety (a stronger feeling of safety may lead to higher utility). Therefore, improving only subjective but not objective safety may still produce benefits and value. Most empirical studies deal with valuing 'objective risks', but it needs noting that what exactly is being valued matters.

This is also true for the type of 'event' that individuals are kept safe from. Of course, one would expect, ceteris paribus, improved safety from death to be valued more highly than improved safety from a mild illness. In some cases, these differences may be less obvious and differ between respondents. For example, individuals may 'dread' certain situations more than they dread others. To illustrate this with the example of avoiding deaths, people may fear certain types of death more than others. For instance, they may fear immediate deaths more than a 'more gradual' process of dying. Similarly, people may be more willing to pay for safety from 'bad deaths', such as murder and drowning,[11] than from other types of deaths. This is relevant to consider in interpreting (the heterogeneity of) results. Whether or not such differences affect final results of an economic evaluation also depends on aspects like baseline risks,[11] but for the valuation exercise these differences emphasise the importance of being clear about what is being valued.

Similarly, and relevant in the context of safety in health and other domains, is the concept of a *catastrophe*. Some safety measures are aimed at preventing large scale impacts, such as pandemics of deathly diseases or floods of large areas of some country or region. Such contexts of a valuation exercise may invoke responses reflecting that 'large concentrated losses are over-counted relative to dispersed losses'[12] — for example a plane crash in comparison to a number of car accidents leading to similar health losses. In a catastrophe, when risk reduction is only described in terms of a reduction in victims, this may undervalue the impact on the feeling of safety in other people. Such contexts show the interconnectedness of objective and subjective safety and it is important to understand and, if possible, distinguish these in the context of valuing safety. Especially catastrophes may have far-reaching spill-over effects and therefore studies valuing reduction in risk of an outcome that may be perceived as a catastrophe may need to include additional information or measures.[12]

In terms of *how* safety is being valued some remarks also need to be made, next to the general observations about stated and revealed preference as well as contingent valuation mentioned above. When developing any valuation measure it is important to consider the impact that the design of the study could have on the results. One design feature that has been found to be relevant

in safety valuation, related to the issues discussed above, is the information provided in the survey. Having a clear and comprehensive valuation exercise is important especially when using indirect methods, as respondents can easily be overloaded with respondent fatigue. Including too much or too little information about what is being valued could make questions harder for respondents to understand or lead to own interpretations of the question posed. How to present the information is also an important consideration. It can be presented using various survey techniques. For example, Mattea et al. (2016)[13] explore the use of visual information in a stated preference study and find that respondents' preferences exhibited more stability when visual information was used to explain risk probabilities when studying risk reduction valuation in landslide programmes.

In CV studies, ordering effects, embedding effects and internal consistency have been shown to be important.[14] Ordering effects refer to the fact that the way in which a respondent values a certain good is dependent on the order of the information presented to them during the valuation exercise.[15] Embedding effects are most relevant when referring to the valuation of public goods or services, for example a flu-vaccination campaign. By asking an individual their WTP for this campaign, they are implicitly being asked their WTP for an injection, a reduction in the probability of getting the flu, an increase in the probability of side-effects from a vaccine, etc. There are multiple 'products' *embedded* in this one question.[15] Internal consistency is not frequently tested in CV research, which has worried critics. In the case of CV, internal consistency refers to the fact that the same type of survey on different WTP questions should come up with consistent results. Halvorsen (1996) researched ordering effects and internal consistency when testing WTP for reduced health damage from air pollution and found considerable and significant ordering effects but could not reject their hypothesis of internal consistency. Halvorsen (1996) did not specifically research embedding effects but emphasised the complications of combining all the elements of a certain programme into one valuation question.

In terms of *who* is valuing safety, it needs noting that individual characteristics can affect the valuation. The most frequently researched of these individual characteristics is *risk perception*. This refers to how an individual perceives

the level of risk in a situation.[16] High risk-perception (i.e. assuming larger levels of risk than objectively present) has been shown to lead people to value safety (or risk reduction) more highly.[17] An issue related to risk perception is probability weighting, a part of general prospect theory. Individuals are known to not value probabilities linearly but to overestimate small probabilities and underestimate large probabilities.[18] In fact, Bleichrodt and Eeckhoudt (2006)[19] showed that correcting for probability weighting strongly affects the WTP estimates for reductions in health risks. Another individual issue to consider is respondent uncertainty. It has been shown that respondents are frequently uncertain about their preferences when answering contingent valuation questions and it is a concern that this uncertainty may be affecting CV results.[20] However, Logar and van den Bergh (2012) found that incorporating information on respondent uncertainty into the model does not lead to any gains compared to a standard CV model. It is also worth noting that risk perception is rarely equivalent to worry, as worry is based on emotion rather than intellectual judgment. As Sjoberg (1998) puts it: 'One can feel worried about a risk without believing that it is especially large, and *vice versa'*. However, worry and also pessimism have been shown to be small explanatory factors of risk perception that vary in size depending on the risk being studied.[16]

Another issue that is frequently thought of as causing bias in CV results is public opinion. Critics have contested the assumption underlying CV that respondents have 'well-defined and self-interested preferences' and argue that respondents are in fact influenced by public opinion. Chanel et al. (2006)[21] attempted to test this by giving a group of respondents the option to revise their answers on how much they were willing to pay for a decrease in air pollution after hearing the mean WTP response from the survey group they were in.[21] They found that at least this type of 'public opinion' had no significant impact on respondents' answers and suggest that it may be a poorly-defined private value structure (or preferences) that leads to a reaction to public opinion.[21] The fact that (ideas about) public opinion may have an impact on valuations of safety at least may be something that those developing a CV study may wish to bear in mind.

From the above, it is clear that valuations of safety may depend on the context provided in describing what is being valued, on how safety is valued and by whom. So far, a golden standard for performing valuation studies of safety

emerging from theory is lacking. Hence, it is important to consider how safety is valued in practice.

Methods

In October 2016, a comprehensive literature search for papers related to the valuation of safety was performed. We assumed that alongside papers related to health, there would also be interesting methods on the valuation of safety outside of the biomedical fields. Therefore, one biomedical database, Embase, and one 'broader' database, Scopus, were used. Embase was chosen as the biomedical database as it holds the largest number of indexed records (in comparison to PubMed and Medline) and also includes all records that are present in Medline. Practically, Embase has a somewhat more advanced search filter than other biomedical databases. Scopus was chosen as it covers a broad range of subject fields: life sciences, social sciences, physical sciences and health sciences. It is comparable to Web of Science.

The results are reported according to the PRISMA guidelines.[22] There was no restriction on time period. Book chapters, dissertations, and theses were not considered. The following terms were used for the search: value, valuation, review, shadow price, willingness to pay, willingness to accept, discrete choice experiment, stated preference, revealed preference, and contingent valuation. The above terms were used in combination with these search terms: Safety, security, uncertainty reduction, risk reduction. The exact search strings are provided in Appendix 5.A. Secondary references were found by searching the references of the already included papers in order to find relevant papers that the databases may not have included.

Papers retrieved from the search were selected for review if they fitted both of the following inclusion criteria: Firstly, the research is empirical, and secondly, the research deals with the valuation of safety, security, risk reduction, uncertainty reduction or reduction of some event that is stated to decrease safety. Papers were excluded if safety valuation was not a main objective of the paper or if the paper was not in English (Table 5.1). It is worth noting here that as this research aims to examine the effect of *increasing* safety (or *decreasing* risk*),* papers that research monetary benefits for *decreases* in safety would not be included.

Table 5.1 Results of Search Terms

Embase	*Safety*	*Security*	*Uncertainty Reduction*	*Risk Reduction*	*Total*
Value	29099	2409	15	3312	34835
Valuation	173	61	1	84	319
Shadow Price	1	2	0	0	3
Review	177856	9016	15	24150	211037
WTP	252	24	0	141	417
WTA	41	4	0	8	53
DCE	61	1	0	25	87
Stated Preference	32	1	0	21	54
Revealed Preference	2	0	0	3	5
CV	10	5	0	10	25
Total (incl. Value & Review)					246835
Total (excl. Value & Review)					963
Scopus	*Safety*	*Security*	*Uncertainty Reduction*	*Risk Reduction*	*Total*
Value	82152	30435	4535	25783	142905
Valuation	706	1218	143	531	2598
Shadow Price	11	41	4	8	64
Review	194236	20204	1990	67514	283944
WTP	632	181	97	497	1407
WTA	135	58	5	59	257
DCE	93	16	4	70	183
Stated Preference	274	82	13	138	507
Revealed Preference	310	128	8	101	547
CV	85	37	11	87	220
Total (incl. Value & Review)					432632
Total (excl. Value & Review)					5783

One of the authors (MP) screened the title and abstract of each paper, checking for inclusion and exclusion criteria. After this screening a second check was performed in which entire texts were scanned to ensure the papers were eligible for the review. The following information was extracted and entered into a table (Table 5.2) for all included papers:

1. Author(s)
2. Title of Paper
3. Year
4. Academic Field
5. Definition of safety
6. Method

Table 5.2 General Paper Information

Author(s)	Title of Study	Year	Academic Field	Definition of Safety	Elicitation Format
Alberini et al. (Alberini et al. 2006)	Willingness to Pay to Reduce Mortality Risks: Evidence from a Three-Country Contingent Valuation Study	2006	Health	Risk reduction	Contingent Valuation
Andersson (Andersson. 2007)	Willingness to pay for road safety and estimates of the risk of death: Evidence from a Swedish contingent valuation study	2012	Transport	Risk reduction	Contingent Valuation
Atkinson et al. (Atkinson et al. 2005)	Valuing the costs of violent crime: A stated preference approach	2015	Crime	Incidence reduction	Contingent Valuation
Carlsson et al. (Carlsson et al. 2004)	Is Transport Safety More Valuable in the Air?	2004	Transport	Risk reduction	Contingent Valuation
Carlsson & Johansson-Stenman (Carlsson and Johansson-Stenman. 2000)	Willingness to pay for improved air quality in Sweden	2000	Environment	Incidence reduction	Contingent Valuation
Carson & Mitchell (Carson and Mitchell. 1993)	The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, and Swimmable Quality Water	1993	Environment	Incidence reduction	Contingent Valuation
Chanel et al. (Chanel et al. 2006)	Does public opinion influence willingness-to-pay? Evidence from the field	2006	Environment	Risk reduction	Contingent Valuation
Corso et al. (Corso et al. 2013)	A Comparison of Willingness to Pay to Prevent Child Maltreatment Deaths in Ecuador and the United States	2013	Health	Incidence reduction	Contingent Valuation
Dealy et al. (Dealy et al. 2013)	The Economic Impact of Project MARS (Motivating Adolescents to Reduce Sexual Risk)	2013	Health	Risk reduction	Contingent Valuation
Determann et al. (Determann et al. 2014)	Acceptance of Vaccinations in Pandemic Outbreaks: A Discrete Choice Experiment	2014	Health	Incidence reduction	Discrete Choice Experiment

Author(s)	Title of Study	Year	Academic Field	Definition of Safety	Elicitation Format
Dickinson & Paskewitz (Dickinson and Paskewitz. 2012)	Willingness to Pay for Mosquito Control: How Important Is West Nile Virus Risk Compared to the Nuisance of Mosquitoes?	2012	Environment	Incidence reduction	Conjoint Analysis
Enneking (Enneking. 2004)	Willingness-to-pay for safety improvements in the German meat sector: the case of the Q&S label	2004	Food Safety	Safety	Discrete Choice Experiment
Flügel et al. (Flügel et al. 2015)	Car drivers' valuation of landslide risk reductions	2015	Environment	Risk reduction	Discrete Choice Experiment
Garza-Gil et al. (Garza-Gil et al. 2016)	Marine aquaculture and environment quality as perceived by Spanish consumers. The case of shellfish demand	2016	Environment	Safety	Contingent Valuation
Georgiou et al. (Georgiou et al. 1998)	Determinants of individuals' willingness to pay for perceived reductions in environmental health risks: a case study of bathing water quality	1998	Environment	Risk reduction	Contingent Valuation
Gerking, et al. (Gerking et al. 1988)	The marginal value of job safety: A contingent valuation study	1998	Labour	Risk reduction	Contingent Valuation
Gyrd-Hanssen et al. (Gyrd□ Hansen et al. 2008)	Willingness-to-pay for a statistical life in the times of a pandemic	2007	Health	Risk reduction	Contingent Valuation
Haddak et al. (Haddak et al. 2016)	Willingness-to-pay for road safety improvement	2014	Transport	Risk reduction	Contingent Valuation
Halvorsen (Halvorsen. 1996)	Ordering effects in contingent valuation surveys: willingness to pay for reduces health damage from air pollution	1996	Environment	Risk reduction	Contingent Valuation
Henson (Henson. 1996)	Consumer Willingness to Pay for Reductions in the Risk of Food Poisoning in the UK	1996	Food Safety	Risk reduction	Contingent Valuation

Author(s)	Title of Study	Year	Academic Field	Definition of Safety	Elicitation Format
Hunter et al. (Hunter et al. 2012)	The effect of risk perception on public preferences and willingness to pay for reductions in the health risks posed by toxic cyanobacterial blooms	2012	Environment	Risk reduction	Contingent Valuation
Iraguen & de Dios Orutzar (Iragüen and de Dios Ortúzar. 2004)	Willingness-to-pay for reducing fatal accident risk in urban areas: An Internet-based Web page stated preference survey	2004	Crime	Risk reduction	Discrete Choice Experiment
Khan et al. (Khan et al. 2014)	Household's willingness to pay for arsenic safe drinking water in Bangladesh	2014	Environment/ Health	Risk reduction	Contingent Valuation
Loureiro & Umberger (Loureiro and Umberger. 2007)	A choice experiment model for beef: What US consumer responses tell us about relative preferences for food safety, country-of-origin labeling and traceability	2007	Food Safety	Safety	Discrete Choice Experiment
Mattea et al. (Mattea et al. 2016)	Valuing landslide risk reduction programs in the Italian Alps: The effect of visual information on preference stability	2016	Environment	Risk reduction	Discrete Choice Experiment
Mofadal et al. (Mofadal et al. 2015)	Analysis of pedestrian accident costs in Sudan using the willingness-to-pay method	2015	Transport	Risk reduction	Contingent Valuation
Patil et al. (Patil et al. 2016)	Public preference for data privacy - A pan-European study on metro/train surveillance	2016	Transport	Security	Discrete Choice Experiment
Pham et al. (Pham et al. 2008)	Households' willingness to pay for a motorcycle helmet in Hanoi, Vietnam	2008	Transport	Incidence reduction	Contingent Valuation
Rizzi & Ortuzar (Rizzi and de Dios Ortúzar. 2003)	Stated preference in the valuation of interurban road safety	2003	Transport	Safety	Discrete Choice Experiment

Author(s)	Title of Study	Year	Academic Field	Definition of Safety	Elicitation Format
Smith et al. (Smith et al. 2014)	How Should the Health Benefits of Food Safety Programs Be Measured?	2014	Food Safety	Risk reduction	Discrete Choice Experiment
Viscusi (Viscusi. 2009)	Valuing risks of death from terrorism and natural disasters	2009	Environment	Risk reduction	Discrete Choice Experiment
Yabe (Yabe. 2016)	Students, Faculty, and Staff's Willingness to Pay for Emergency Texting	2016	Crime	Safety	Contingent Valuation
Yun et al. (Yun et al. 2016)	Analysis of the Relationship between Risk Perception and Willingness to Pay for Nuclear Power Plant Risk Reduction	2016	Environment	Risk reduction	Contingent Valuation

Two separate tables (Table 5.3 and Table 5.4) were made for each type of method with columns for:

1. Paper
2. Scenario Description
3. Question asked to respondents
4. Measurement scale (CV) or Attributes (DCE)
5. Econometric Model(s)
6. Covariate results

The comprehensive search yielded a total of 679,467 results. Because the search terms 'value' and 'review' produced many seemingly irrelevant results, any results using these search terms were not included in the abstract screening, leaving 6,746 results for further screening. This first involved evaluating whether paper titles appeared to fit the inclusion criteria, which resulted in the exclusion of 6,659 papers (99%). If the title of the paper was relevant, then the abstract was checked to confirm that the paper did indeed fit the inclusion criteria. This was frequently not the case, leaving 49 papers (5%) after this screening. The reference lists of these papers were searched for additional papers empirically examining the valuation of safety. Nine additional papers were added after this step; hence, 58 papers were included in the next step of the review process. This involved a more thorough check, which showed that 24 of the 58 papers were either a non-empirical paper or did not focus on the value of safety. One additional paper was excluded as it only measured relative values of safety rather than absolute, using a ranking method. Therefore, 33 papers were finally included and summarized in the review.

The main aim of this review, as mentioned previously, was to examine the various methodologies used for valuing safety. Therefore, in both the table and the findings section of this paper, most weight will be placed on study methodology. Due to the variety of topics covered by the papers, the comparison of WTP values seemed nonsensical (since incomparable). However, to give some insight into possible results from similar studies, the covariate results that can be compared across fields are discussed in the findings.

Table 5.3 Contingent Valuation Method

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Crime					
Atkinson et al.	Respondents shown injury descriptions for 3 types of assault: common assault, other wounding, serious wounding. Also informed of pre-policy risk of the incident occurring.	Asked WTP to reduce chance of being a victim to one of the three offences (randomized per respondent) by 50% over the next 12 months. Payment vehicle is a one-off increase in local changes for law enforcement.	Payment card: £0-5000	Interval data model.	Severity of the risk increases WTP. Higher incomes and education both increase WTP.
Corso et al.	Respondents are asked to imagine that there is a program available in their city that reduces the risk of a child being killed by a parent or caretaker by 50%.	Asked WTP for this program through (i) taxes or (ii) donations.	Double-bounded dichotomous choice: Initial WTP value between $10 and $300. Second WTP values are $25 higher (lower) if response is 'yes' ('no').	Maximum likelihood function.	Those reporting histories of child maltreatment have lower WTP.
Yabe	Respondents are told that a Text-to-911 service would be paid for by a few charges to students, staff and faculty at the university.	Respondents are asked if they would be willing to pay X$ for an emergency text messaging service.	Dichotomous choice: bids - $1, $2, $3, $5, $10	Logit model.	Being interested in emergency texting, having experience in campus emergencies, being older, having a higher income, and being American (rather than international) leads to higher WTP.

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Environment					
Carlsson & Johansson-Stenman	No scenario given; researchers want respondents to judge the information about air pollution from various sources.	Asked WTP for a 50% reduction in concentration of harmful substances where they live and work.	Open-ended questions.	Probit. Tobit type I. Tobit type II. Independent models.	WTP increases in income, wealth and education. WTP is larger for: men, members of environmental organizations, people living in big cities, and those who own their house or apartment. WTP is lower for retired people.
Carson & Mitchell	Respondents told that although present minimal water level is 'boatable', most of the nation's freshwater bodies are fishable and 70% are swimmable. Used the 'Resources for the Future' water quality index to show physical water quality parameters.	Asked WTP in taxes 'to keep the nation's freshwater bodies from falling below the boatable/fishable/swimmable level where they are now'. Four WTP amounts solicited for each of the three water quality questions: (i) amount given for each of the WTP questions (ii) WTP given after first amount is repeated and respondents encouraged to make desired corrections (iii) WTP after respondents informed of the range of the amounts households in their income group were already paying for water quality (iv) WTP given after respondents pushed to increase their bid.	Payment card: 0$ to a 'very high amount'. Five points on the card show average amounts households pay in taxes for non-environmental public goods.	Cobb-Douglas.	N/A

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Chanel et al.	Respondents are given the hypothetical choice to move with their family to a less polluted city. Two cities are proposed that are equivalent with the exception of level of pollution and the cost of living.	Four steps: (i) WTP to live in less polluted city. (ii) WTP after shown mean WTP of all respondents. (iii) WTP after receiving scientific and quantitative information on health effects of pollution. (iv) WTP after new mean shown to respondents.	Dichotomous choice questions. Open-ended questions.	Wilcoxon sign-ranked tests.	Public opinion has no effect on WTP. Information provided leads to higher WTP.
Garza-Gil et al.	No scenario provided.	Asked WTP for an enhanced safety guarantee programme for shellfish quality and environmental conditions.	Dichotomous choice questions. WTP 5%, 10%, 20% or more than 20% more than normal price.	N/A	The higher the price the lower the number of people WTP for the intervention.
Georgiou et al.	Respondents informed about sewage contamination of bathing water and health risks from bathing, EC bathing water standards and actual quality of water at a particular beach.	Asked WTP for (i) for a gain (ii) for a loss in bathing water standards - dependent on the beach at which applicants are surveyed.	Open questions.	Semilog model.	Higher income and education lead to a higher WTP. The more unacceptable the respondent finds the risk, the higher their WTP. Having a family member who has suffered due to poor bathing water leads to a higher WTP.

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Halvorsen	Respondents given description of benefits from a 50% reduction in air pollution. These are (i) reduction in in the risk of becoming ill and (ii) a reduction in damage due to acid rain.	Asked maximum WTP for 50% reduction in air-pollution. Four sub-samples with two splits: (1) Sub-samples B and D are told that the government will subsidize electric cars. A and C are told that the government uses a package of unspecified tools. (2) Those in A and B are given all information, then asked WTP. Those in C and D are first given health effect information then asked WTP, then are given all other effects and asked if they wish to change their WTP.	Open questions.	Tobit. Cragg (i) Probit model (ii) Truncated model.	Income, living in a major city, having a university degree and being concerned with the environment all have a positive effect on WTP. Age has a negative effect on WTP.
Hunter et al.	Respondents informed about (i) what cyanobacteria are (ii) the ecological and human health problems they cause and (iii) the practical options available for health-risk mitigation at Loch Leven.	Asked max WTP towards measures to reduce number of Risk Days from 90 to (i) 45 or (ii) 0. Payment vehicle is the cost of domestic water supply set by the council.	Payment card (values not stated).	Binary logit model. Non-parametric models: normal, logistic, lognormal, Weibull, and spike model.	Those with higher concern for environmental health risks have higher WTP. Income has a positive effect on WTP.

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Khan et al.	No scenario provided.	Asked WTP for: (i) a communal deep tube well, (ii) one-time-off capital investment costs of the well (iii) one-time-off investment costs and (iv) operation and maintenance costs.	Double bounded dichotomous choice. Capital costs: Min. bid - 50 BDT. Max. Bid - 250 BDT. O&M costs: min. bid - 10 BDT. Max. Bid - 100 BDT.	Bivariate probit model. Random effects probit.	When respondents are male or earn higher incomes WTP is higher. If households are exposed to higher risk levels, if respondents are aware that their water is contaminated, and if household members are affected by arsenic exposure then WTP increases.
Yun et al.	Respondents are first asked to rank an image about nuclear power plants on a Likert scale of 'very good image/safe (5)' to 'very bad image/ unsafe (1)'.	Respondents are asked if they would pay *A$* to reduce NPP hazard.	Dichotomous choice. Bids are not described.	Log-linear. Linear. Linear-log. Power regression models.	Higher scientific background/ low risk perception led to a lower mean WTP. Mean WTP decreased with increasing quality of informational image.

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Food Safety					
Henson	Respondents are informed that chicken/ egg consumption can cause food poisoning. They are told about two brands of chicken/eggs in the shop. Brand A and Brand B are identical except Brand A has been thoroughly tested and thus has a lower risk of giving one food poisoning.	Maximum additional amount WTP for a risk reduction in (i) fatal food poisoning (ii) mild food poisoning (iii) moderate food poisoning (iv) severe food poisoning in chicken or eggs.	Open question.	Ratios calculated.	More severe outcome leads to higher WTP. Personal experience of food poisoning has a negative effect on WTP. Mean WTP is higher for female respondents. Age and education both have a negative effect on WTP. WTP is positively affected by income.
Health					
Alberini et al.	Respondents are shown their baseline risk of death (that varies with gender and age) over the next 10 years.	Asked WTP for a risk-reduction in death of (i) 5-in-1000 incurred over the next 10 years, (ii) 1-in1000 incurred over the next 10 years, (iii) 5-in-1000 that begins at age 70 and is spread over next 10 years. Payment would be made every year.	Dichotomous choice questions.	Accelerated-life Weibull model.	Income increases WTP. WTP increase with age until age 60 and then plateaus. Hospitalization for cardiovascular or respiratory illness leads to higher WTP.

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Dealy et al.	Participants are randomly assigned to receive one of three treatments: (i) sexual risk reduction intervention (ii) sexual risk reduction plus alcohol risk reduction (iii) sexual risk reduction intervention including both an alcohol and a marijuana risk reduction component.	Asked WTP 'not to get' (i) a curable STD (ii) an incurable non-fatal STD (iii) a fatal STD. Asked before and after intervention.	Open question with a bound of $0-100,000	Anova.	WTP increases after receiving the intervention. WTP increases with the severity of the STD.
Gyrd-Hanssen et al.	No scenario provided.	Asked maximum WTP in order to have a course of Tamiflu drug available in case they would need it.	Open questions.	Linear regression analysis.	Age and being female increase WTP. Household income has a positive impact on WTP. Being uncertain of baseline risk has a positive impact on WTP. Being uncertain of the perceived benefit has a negative impact on WTP.
Labour					
Gerking, et al.	Respondents are asked what their current job is.	Asked (i) how large an increase in annual wages would lead to respondent voluntarily working 'one step up' the risk ladder (WTA) (ii) how large a decrease in annual wages would a respondent forego in order to move one step lower on the risk ladder.	Payment card: $0 to $6000.	Two-limit tobit procedure.	Higher income and perceived likelihood of death at work leads to higher WTP/WTA. Older workers have a higher WTP/WTA. WTP decreases with formal educational levels.

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Transport					
Andersson	Respondents shown overall death risk for an individual. Also shown risk of dying in a traffic accident.	Asked one of two questions: (i) WTP for reducing personal annual risk of death by a third. (ii) WTP for reducing personal annual risk of dying in a traffic accident by one third.	Open-ended questions	Non-linear models. Log-linear models.	WTP increases as baseline risk increases. WTP declines with age. WTP declines with background risk. WTP increases with income.
Carlsson et al.	Two scenarios: (i) The respondent is going to take a taxi alone. They have two taxi options which are identical except for the risk of a fatal accident - 1 in 1 million (AAA) or 0.5 in 1 million (BBB). (ii) The respondent will take a plane alone. They have two airline options which are identical except for the risk of a fatal accident - 1 in 1 million (AAA) or 0.5 in 1 million (BBB).	Cases: Asked WTP for safer air trip compared to AAA at (i) 500 SEK (ii) 3000 SEK. Asked WTP for safer taxi ride compared to AAA at (iii) 50 SEK (iv) 500 SEK. (v) Asked both (i) and (iv). (vi) Asked both (ii) and (iv).	Open-ended questions anchored with baseline-risk prices (AAA)	Tobit type II.	Cost of trip leads to a higher WTP. Higher income leads to higher WTP. Male respondents have lower WTP. Fear of flying leads to a higher WTP for both air and taxi questions.

Paper	Scenario Description	CV Question(s) asked to respondents	Measurement Scale	Econometric Model(s)	Covariate results
Haddak et al.	Three projects presented to respondents: reduces risk of being a victim of (i) a road accident that causes minor injuries (ii) a road accident resulting in serious injuries (iii) a road accident that results in moderate injuries.	Asked how much they would be willing to pay for a (i) 25% reduction (ii) 50% reduction in risk of experiencing various non-fatal types of injuries following a road accident.	Open questions.	Logit. Tobit.	WTP is higher for more severe injuries. WTP increases with income. Accidental experience of individuals (direct and indirect) leads to increased WTP.
Mofadal et al.	Respondent is told to imagine going to work or performing daily activities and during these they need to cross busy streets to reach their destination. The respondent can choose one of five options to reduce this risk.	The respondent first chooses the optimum scenario regarding crossing behavior and side walking. They are then asked their maximum WTP to reduce the risk of a fatality in that scenario. They are also asked their maximum WTP for a pedestrian safety program that reduces fatality risk by 50%.	Payment card: 0 to more than 3000 SDP	Log-linear.	Age positively affects WTP. Income positively affects WTP. Married respondents have a lower WTP. Males have a higher WTP. Higher education increases WTP.
Pham et al.	Respondents are given the hypothetical situation that the government subsidizes the price of motorcycle helmets.	Respondents are asked the maximum amount they are willing to pay for a motorcycle helmet.	Open question. Dichotomous choice questions - min. 50,000 SDP, max. 150,000 SDP.	Interval regression. Multi-linear regression model.	Age and income have a positive effect on WTP. Those with higher education, those with jobs outside of the office and those with a better knowledge of/attitude towards helmets have a higher WTP.

Table 5.4 Discrete Choice Experiment / Conjoint Analysis

Paper	Scenario Description	Question asked to respondents	Attributes	Econometric Model(s)	Covariate Results
Crime					
Iraguen & de Dios Orutzar	Respondents are asked to imagine they are traveling to work from home. The trip takes place on a regular working day, they arrive at their destination at around 7:45 am, and they drive their own car and are responsible for all costs involved.	Respondents are asked to choose between two different routes with differing attributes.	Travel time. Travel cost. Number of fatal car accidents per year.	Multinomial logit model.	Income negatively affects the perception of the importance of travel cost. Safety valuation is positively affected if the individual travels with somebody else. This is also true if the respondent is female or they have been in a serious accident before.
Environment					
Dickinson & Paskewitz	Respondents are informed of the multiple types of mosquitoes in Madison (some a nuisance, some transmit West Nile virus). Control program which would control mosquito larvae could control one type of mosquito larvae or both.	Asked to choose between pairs of hypothetical control programmes.	West Nile Risk. Type of mosquito targeted. Cost (through taxes): $10-200.	Conditional logit model.	Increased risk level leads to higher WTP. WTP decreases as cost increases.
Flügel et al.	Respondents who had a recent trip by car were presented with different choices for a car trip route.	Respondents are asked to choose between two routes with four differing attributes, 6 times.	Cost: fuel and toll. Travel Time. Casualties: fatalities and serious injuries. Landslides: share of the route with landslide risk.	Mixed logit models.	Men are less likely to choose lower landslide risk. People with a higher education tend to choose the option with the lowest risk more often.

Paper	Scenario Description	Question asked to respondents	Attributes	Econometric Model(s)	Covariate Results
Mattea et al.	No scenario given.	Respondents are given six choice sets of seven alternatives, each of which consists of five attributes. These questions are asked twice, and respondents are given visual information on the possible options before being asked a second time.	Four alternatives represent devices to protect against landslides: Diverging channel, Retaining basin, Video cameras, and Acoustic sensors. The fifth alternative is a hypothetical road toll.	Mixed logit in WTP spaces. Multinomial logit model. Mixed logit in preference space.	The 'status quo' is negatively perceived.
Viscusi	Traffic - On an average day 100 people die due to traffic accidents. These risks are isolated deaths. Natural disasters - These are national catastrophes and large numbers of people can die at the same time. Hurricane Katrina killed over 1,000 people. Terrorism - Attacks by terrorists can also be catastrophes. The 9/11 attacks killed 2,976 people.	Respondents are asked 'risk-risk' tradeoff questions (traffic accident-terrorist attack, traffic accident-natural disaster) in two sets of 6 question blocks.	Type of deaths prevented. Average number of deaths prevented.	Conditional logit models. Mixed logit models	More education raises the utility coefficient in every instance, and more so with terrorism. Income has a negative effect on utility. Seatbelt usage increases the utility of reducing all deaths.

Paper	Scenario Description	Question asked to respondents	Attributes	Econometric Model(s)	Covariate Results
Food Safety					
Enneking	Participants are given a short introduction to the Quality and Safety labelling system (regarding liver sausages).	Asked to name three choices from a set of 6 sausages	Brand A: national premium brand (with/ without Q&S label). Brand B: National brand (with/without label). Brand C: National premium brand - reduced fat (no label). Brand D: Private label - organic (no label). Brand E: National organic umbrella brand name (no label). Brand F: Private label.	Maximum likelihood.	Those who find low prices important avoid the more expensive labelled brands.
Loureiro & Umberger	No scenario given.	Asked to choose between two steaks (Option A and Option B) with five varying attributes.	Price ($/lb). Country of origin labeled. Traceability to the farm. Food safety inspected. Guaranteed tender.	Multinomial conditional logit model.	Increasing price of option leads to lower utility. Steaks inspected by US food inspectors carry the highest premium.
Smith et al.	No scenario given.	Regarding improvement of food safety respondents are asked to choose between: the 'status quo', 'Hire more inspectors' and 'purchase medicine'. Each subject is asked 12 choice questions, where each option consists of five attributes.	Annual risk of food borne illness. Average amount of time you will be sick. Extra time needed to prepare food. Cost. Annual increase in income tax.	Multinomial logit models.	Consumers prefer reduction ex ante risk than ex post. Those who are more willing to accept risk, are not as likely to accept risk reduction policies. Respondents prefer private control over the risk reduction.

Paper	Scenario Description	Question asked to respondents	Attributes	Econometric Model(s)	Covariate Results
Health					
Determann et al.	Respondents are presented with some combination of two scenario variables (i) susceptibility to the disease (ii) severity of the disease.	Asked to choose between Vaccine A, Vaccine B and 'No Vaccine' in 16 choice sets. Vaccines are comprised of different levels of 5 attributes.	Effectiveness of vaccine. Safety of the vaccine. Advice regarding the vaccine. Media coverage. Out-of-pocket costs.	Latent class model.	Females and individuals who stated they would never get vaccinated were more influenced by media and more sensitive to costs. WTP is higher for more effective vaccines, especially if the outbreak was more serious.
Transport					
Patil et al.	No scenario given.	Each respondent answered five choice exercises regarding their security preferences when traveling by train or metro.	Type of CCTV cameras. How long CCTV information is stored. Who can access CCTV information. Security personnel at the station. Type of security checks at the station. Time to go through security checks. Security surcharge.	Multinomial logit model.	All preferred CCTV over no CCTV. Preference is weaker for younger people. Females have a stronger preference for CCTV.
Rizzi & Ortuzar	Survey is disguised as a survey to improve interurban route policy and road safety. Respondents are given an identical trip in which: they drive their own car, they pay for the total cost of the trip, and they have to return after 20:00.	Respondents are asked to answer nine choice situations. They are asked to choose between two routes with differences in the three attributes.	Travel time. Toll charge. Annual accident rate (represents "general level of safety").	Binary logit models.	Women have a higher preference for safety than men, as do older people. There is a higher preference for safety if the trip takes place at night. A person driving with others in the car is more aware of risk.

Findings

Table 5.2 shows general information about the papers extracted from the review process. Regarding the fields of the papers, the most popular field is Environment (39%), followed by Transportation (21%) and Health (15%). Twenty-two of the papers (67%) used the contingent valuation (CV) method for their valuation of safety and 11 (33%) used a form of discrete choice experiment (DCE) or conjoint analysis. Of the 33 papers, 20 (60%) used 'risk reduction' as the definition of safety, seven (21%) simply referred to a 'reduction in [unwanted outcome]', five papers (15%) used the term 'safety', and one paper (3%) valued 'security'.

Table 5.3 synthesizes the more specific results of the papers that use CV methods. All papers used one of three types of measurement scale: open-ended questions, payment cards or dichotomous choice questions. Dichotomous choice questions can be broken down into single- or double-bounded questions, where a double-bounded question means that, after being given an initial 'yes or no' WTP price, as in a single-bounded question, the respondent is then given a second WTP option dependent on his first answer.[23] The most popular question format of the 22 papers is an open-ended question (48%),[3,14,17,21,24-28] followed by dichotomous choice (35%),[21,28-34] and payment card.[35-39] Two of the papers use both open-ended questions and dichotomous choice.[21,28] Of the six papers using dichotomous choice, two use double-bounded questions.[30,32]

Table 5.3 also includes findings concerning covariates and their effect on WTP for safety. These covariates can be categorised into three groups: individual characteristics, individual relationship with risk, and aspects of the study design. Regarding individual characteristics, the findings show that higher income was associated with a higher WTP in every case in which it was investigated.[3,14,17,26-28,32,34,35,37-41] Many papers investigating this relationship (70%) report that having a higher level of education is associated with a higher WTP,[14,24,26,28,29,39] while others (30%) report the opposite result.[3,34,37] Age and gender are variables for which ambiguous effects were reported. Several papers (54%) find that increasing age is associated with increased WTP,[27-29,33,37,39] however others (46%) report the opposite result.[3,14,34,40] In papers where gender was considered, sometimes men reported a higher WTP[24,39] and sometimes women did.[3,27,41]

Secondly, we can consider the group of variables that concern the individual and their relationship with the risk. For example, if an individual is more susceptible to the outcome,[29] has been previously exposed to the outcome,[32] or has a family member who has experienced the situation,[26] they are associated with reporting a higher WTP according to some of the papers reviewed. There are several other factors that could lead to an increased WTP. For example, if an individual is more concerned about the issue at risk,[14,28] finds the risk unacceptable,[26] has a higher perceived risk,[27,37] is uncertain of the benefit or risk of the outcome,[37] or is aware of,[32] interested in,[33] or knowledgeable about[28] the issue. Those with experience of the outcome sometimes report higher WTP (60%)[17,25,33] and sometimes report lower WTP (40%)[3,30] than those who have not experienced the outcome. The studies, in which WTP is lower with experience of the outcome, cover the topics of child maltreatment risk reduction[30] and the risk reduction of food poisoning.[3] Corso et al. indicate that the finding is not what was expected but they do not come up with a concrete explanation for the mechanism underlying the result. Henson explained their result through two mechanisms: the first is that those who have recently suffered from food poisoning believe that they have a smaller chance of getting food poisoning in the future, and the second is that many suffered only mild symptoms and so may underweight the probability of having moderate to severe food poisoning symptoms.[3]

Thirdly, we can consider the group of variables related to aspects of the study design. Using a higher baseline risk[40] or severity of risk[3,24,25] is associated with individuals reporting a higher WTP. From the two CV studies that place a price on the intervention, one study finds that increased cost price is associated with higher WTP[41] while the other study finds the opposite result.[31] Carlsson et al. (2004) give no explanation as to why a higher cost price suggests a higher WTP in their paper. Since, however they research choices between taxi rides and flights, it may be due to people assuming that the more expensive the journey, the safer it is. Two studies also investigated the effects of more information on individuals' WTP. Chanel et al. found that giving more information regarding pollution levels is associated with higher WTP, whereas Yun et al. found that providing people with better quality informational images is associated with lower WTP for reduced nuclear power plant hazard. Because they approach the study from the point of view that

nuclear power plants are safer than assumed by some of the public, they do not explicitly discuss why better-quality information is associated with lower WTP. [34] In general, however, better information should have no a priori effect; it simply depends on whether prior expectations were too high or too low. As previously mentioned, the second most popular method for valuing safety is DCE or conjoint analysis. Table 5.4 summarizes the main traits of the papers in which DCE or conjoint analysis is used. The most obvious difference between DCE (or conjoint analysis) and CV methods is that DCE and conjoint analysis use attributes so as to *indirectly* measure the value of what is being researched. Since the papers in this review came from many different fields, it is not possible to directly compare attributes. However, there were three types of attribute which almost all DCE studies used and can be described in broad terms as: one which considers the cost price (81%),[13,42-49] one which considers the level of risk or risk reduction (72%),[42,43,45,46,48-51] and one which considers the type of intervention (81%).[13,42-49,51]

Looking at the results from the DCE papers, the effects of covariates on WTP can, once again, be split into three groups — personal characteristics, individual relationship with risk and aspects of the study design. From Table 5.4 we can see that higher age,[47,48] education[45] and income[50] all increase WTP. The only personal variable that differed from the CV results is that in the DCE studies that investigated gender differences (36%), women[42,45,50] always reported a higher WTP. Regarding the interaction of individuals and risk, experience of the event[50] is associated with higher WTP. Finally, looking at the variables which relate to the effectiveness of the method: a higher cost price was associated with lower WTP,[43,46] while a more severe outcome,[35] a higher risk level[43] and a more effective treatment[42] were all associated with higher WTP.

Many of the papers in the study consider some theoretical issues that come with the methodology used. Out of the CV papers, most of those that do consider theory look at the use of visual aids to represent risk.[29,35-37,40,41] Other issues considered are sample size limitation,[28,33] embedding effects,[14,26,41] the interpretation of risk,[27,39] and interviewing effects.[38] The most commonly considered theoretical issues in the DCE papers were sample bias,[43,50] the use of visual aids[13] and behaviour comparability.[46,48]

Discussion

This review aimed to synthesize the methodology and study design used in empirical research valuing safety. This issue is becoming more and more relevant as economic evaluations are increasingly used in the context of informing governmental policy, and as potential threats to our safety in different areas increasingly become a subject of policy. As can be seen from the results section above, there are several main findings regarding the valuation of safety. Firstly, the two main methods used are CV and DCE (or conjoint analysis), with CV being the most frequently used. Secondly, most studies used 'risk reduction' as a definition of safety when valuing it. Thirdly, there are covariate results other than the main variable of interest that are measured across papers, all of which fell under three categories: individual characteristics, the relationship between the individual and risk, and aspects of the study design. Overall, it was the covariate results related to individual characteristics that led to the most ambiguous conclusions, while the results concerning the individual's relationship with risk mostly ran in the same direction across papers. Finally, while most papers did *mention* at least one of the theoretical issues related to valuing safety, few attempted to *tackle* the issues they mention.

Something that is not directly discussed in the findings but is noteworthy, is that all papers use an individual perspective when valuing safety, and none consider or mention using a societal perspective. Doing this would allow the measurement of how individuals value the safety of others and not just themselves, which is clearly relevant when policies are designed to improve the safety of citizens in general and use taxes as the payment vehicle. However, one may then encounter the issue of double-counting, where an individual not only values their utility, but also the utility of someone else.[52] Using a societal perspective in the methodological design would involve additional scenario description and questions. For example, one can include information in the scenario description about who is at risk and who benefits from the intervention, and also ask questions about the individual's WTP if others are also paying (e.g. through raising taxes), or if the individual themselves does or does not benefit (i.e., distinguishing between social values that do or do not take self-interest into account).[53,54]

Several further observations can be made on the basis of this literature review. Firstly, there is the limited number of papers retrieved from the literature search. Therefore, it is difficult to make strong conclusions or recommendations from any of the results, especially those stemming from DCE experiments, of which relatively few were included. To comment on similarities in methodologies used within fields would require a higher number of papers per field as well. Secondly, there is the complexity of defining safety. Even though most papers define safety as 'risk reduction' when valuing it, not all do, and so this muddles any comparison between papers that use different definitions. In addition, acknowledging that feelings of safety may be important for people's wellbeing next to objectively improved safety, it should be noted that valuations of feelings of safety were not present in the current review. Of course, improved objective risk reduction may result in feeling safer as well, but the two need not coincide. Moreover, we may have excluded risk reduction papers that do not allude to safety, even if methodologically very similar to papers included in this review. Lastly, there is the wide range of fields used in this research. Although the diversity of topics does show that the valuation of safety is relevant in many different areas, it limits the comparison of results.

The above observations show us how useful the (evidence based) standardisation of some elements of safety valuation methodology would be. Governments are presented with many policy options while they have a restricted budget. Consequently, they must make choices about which policies to implement and which not, potentially concerning different departments, such as health and education. When making such choices, information about the value for money different policies generate is relevant information and, in this context, a somewhat standardised methodology for valuing safety would be beneficial for the comparability of information between policies. For example, it could be beneficial to have a standardised number and order of questions or attributes and levels, to require the assessment of individual risk perception and to control for probability weighting, just to name a few options.

As with any study, there are of course limitations: first, by only including research on increasing safety, literature in the area of 'wage risk' trade-offs and the Value of a Statistical life is excluded. While this does lead to a

smaller number of results in our review, it is important to note that values for decreasing something beneficial like risk reduction or health can be very different to values for an increase in the same good. A similar issue arises for literature on drug safety — while literature on this topic was not purposefully excluded, it may be that by searching solely for papers that explicitly mentioned *safety* or other permutations, much of this literature may have been excluded. By enforcing strict criteria there is the benefit of clarity, the disadvantage being the exclusion of interesting research. Fortunately, multiple literature reviews have been carried out for both value of a statistical life and drug safety literature, the latter of which are usually drug-type or situation specific, and these can be used for insights, from different angles, into the safety valuation process.[55-58]

Finally, the process would have been strengthened by a second author reviewing abstracts, or the inclusion of more types of research like theses, papers in a language other than English and grey literature. In a similar vein, the chosen databases have their own limitations; neither database contains all records from their relevant fields, and the methods could have been strengthened by searching at least additional journals, for example Web of Science which contains some records Scopus does not (and vice versa). Searching PubMed may also have been beneficial, although Embase is the largest of the three top biomedical databases. We would argue, nevertheless, that this is a thorough review in keeping with PRISMA guidelines which warrants a comprehensive reporting of the findings. Furthermore, as this is the first literature review on safety valuation, the results definitely do provide insight into an area of research that has not been often studied.

Overall, it has become clear that there is little to no standardisation in safety valuation. Regarding which is 'the best' methodology to use, this literature review brings to light more questions than it does answers: Which definition of safety is the best for its valuation? Which stated preference method should be used, CV or DCE, and which methodological issues should be considered in study design? Should the individual or the societal view be applied in the context of valuing public goods? Which covariates should be added to gain the most insight into an individual's WTP? In other words, there still appears to be a long way ahead before consensus can be attained about a standardised

methodology for valuing safety. In the meantime, forthcoming safety valuation research can build upon the findings of this review of the literature and contribute to the development of more standardised methods by addressing questions about definition of safety, choice and design of method, perspective for valuation, and selection of covariates, thoroughly and clearly.

In conclusion, there is no 'golden standard' for safety valuation — there are many different approaches to research methods, survey design, biases and context in the literature. Moreover, given the number of unresolved issues, many aspects of valuing safety are not yet fully understood. What this shows is that there is more work to be done on methodologies for the valuation of safety, theoretically and empirically. That way, it may be able to work towards something more closely resembling a 'golden standard' for safety valuation, which is especially relevant in the field of health economics and economic evaluations addressing health related issues. Investing in this important area, therefore, appears to be a safe bet.

References

1. "safety". *Merriam-Webster.com*. Merriam-Webster, 2011. Web. 24 October 2016.
2. Boithias L, Terrado M, Corominas L, et al. Analysis of the uncertainty in the monetary valuation of ecosystem services—A case study at the river basin scale. *Sci Total Environ*. 2016;543:683-690.
3. Henson S. Consumer willingness to pay for reductions in the risk of food poisoning in the UK. *J Agri Econ*. 1996;47(1□4):403-420.
4. Herzog Jr HW, Schlottmann AM. Valuing risk in the workplace: Market price, willingness to pay, and the optimal provision of safety. *Rev Econ Stat*. 1990:463-470.
5. Ball DJ. Consumer affairs and the valuation of safety. *Accid Anal Prev*. 2000;32(3):337-343.
6. Cellini, S.R., Kee, J.E. Cost-effectiveness and cost-benefit analysis. In: *Handbook of Practical Program Evaluation*. Jossey-Bass & Pfeiffer Imprints, Wiley; 2015.
7. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford University Press; 2015.
8. Brown TC. Introduction to stated preference methods. In: *A primer on nonmarket valuation*. Dordrecht, Netherlands: Springer; 2003:99-110.
9. Mitchell RC, Carson RT. *Using surveys to value public goods: The contingent valuation method*. Resources for the Future; 1989.
10. Ryan M, Gerard K, Amaya-Amaya M. *Using discrete choice experiments to value health and health care*. Vol 11. Springer Science & Business Media; 2007.
11. Chilton S, Jones-Lee M, Kiraly F, Metcalf H, Pang W. Dread risks. *J Risk Uncertainty*. 2006;33(3):165-182.
12. Zeckhauser R. The economics of catastrophes. *J Risk Uncertainty*. 1996;12(2-3):113-140.
13. Mattea S, Franceschinis C, Scarpa R, Thiene M. Valuing landslide risk reduction programs in the italian alps: The effect of visual information on preference stability. *Land Use Policy*. 2016;59:176-184.
14. Halvorsen B. Ordering effects in contingent valuation surveys. *Environ Resour Econ*. 1996;8(4):485-499.
15. Kahneman D, Knetsch JL. Valuing public goods: The purchase of moral satisfaction. *J Environ Econ Manage*. 1992;22(1):57-70.
16. Sjöberg L. Worry and risk perception. *Risk analysis*. 1998;18(1):85-93.
17. Haddak M, Lefèvre M, Havet N. Willingness-to-pay for road safety improvement. *Transportation Research Part A: Policy and Practice*. 2016;87:1-10.
18. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica*. 1979:263-291.
19. Bleichrodt H, Eeckhoudt L. Willingness to pay for reductions in health risks when probabilities are distorted. *Health Econ*. 2006;15(2):211-214.

20. Logar I, van den Bergh, Jeroen CJM. Respondent uncertainty in contingent valuation of preventing beach erosion: An analysis with a polychotomous choice question. *J Environ Manage*. 2012;113:184-193.
21. Chanel O, Cleary S, Luchini S. Does public opinion influence willingness-to-pay? Evidence from the field. *Applied Economics Letters*. 2006;13(13):821-824.
22. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS medicine*. 2009;6(7):e1000097.
23. Hanemann M, Loomis J, Kanninen B. Statistical efficiency of double-bounded dichotomous choice contingent valuation. *Am J Agric Econ*. 1991;73(4):1255-1263.
24. Carlsson F, Johansson-Stenman O. Willingness to pay for improved air quality in sweden. *Appl Econ*. 2000;32(6):661-669.
25. Dealy BC, Horn BP, Callahan TJ, Bryan AD. The economic impact of project MARS (motivating adolescents to reduce sexual risk). *Health Psychology*. 2013;32(9):1003.
26. Georgiou S, Langford IH, Bateman IJ, Turner RK. Determinants of individuals' willingness to pay for perceived reductions in environmental health risks: A case study of bathing water quality. *Environ Plann A*. 1998;30(4):577-594.
27. Gyrd-Hansen D, Halvorsen PA, Kristiansen IS. Willingness□to□pay for a statistical life in the times of a pandemic. *Health Econ*. 2008;17(1):55-66.
28. Pham KH, Le Thi QX, Petrie DJ, Adams J, Doran CM. Households' willingness to pay for a motorcycle helmet in hanoi, vietnam. *Appl Health Econ Health Policy*. 2008;6(2-3):137-144.
29. Alberini A, Hunt A, Markandya A. Willingness to pay to reduce mortality risks: Evidence from a three-country contingent valuation study. *Environ Resour Econ*. 2006;33(2):251-264.
30. Corso PS, Ingels JB, Roldos MI. A comparison of willingness to pay to prevent child maltreatment deaths in Ecuador and the united states. *Int J Environ Res Public Health*. 2013;10(4):1342-1355.
31. Garza-Gil MD, Vázquez-Rodríguez MX, Varela-Lafuente MM. Marine aquaculture and environment quality as perceived by Spanish consumers. the case of shellfish demand. *Mar Policy*. 2016;74:1-5.
32. Khan NI, Brouwer R, Yang H. Household's willingness to pay for arsenic safe drinking water in Bangladesh. *J Environ Manage*. 2014;143:151-161.
33. Yabe M. Students, faculty, and staff's willingness to pay for emergency texting. *Journal of Applied Security Research*. 2016;11(4):437-449.
34. Yun M, Lee SH, Kang HG. Analysis of the relationship between risk perception and willingness to pay for nuclear power plant risk reduction. *Science and Technology of Nuclear Installations*. 2016;2016.
35. Atkinson G, Healey A, Mourato S. Valuing the costs of violent crime: A stated preference approach. *Oxford Economic Papers*. 2005;57(4):559-585.
36. Carson RT, Mitchell RC. The value of clean water: The public's willingness to pay for boatable, fishable, and swimmable quality water. *Water Resour Res*. 1993;29:2445-2454.

37. Gerking S, De Haan M, Schulze W. The marginal value of job safety: A contingent valuation study. *J Risk Uncertainty*. 1988;1(2):185-199.
38. Hunter PD, Hanley N, Czajkowski M, et al. The effect of risk perception on public preferences and willingness to pay for reductions in the health risks posed by toxic cyanobacterial blooms. *Sci Total Environ*. 2012;426:32-44.
39. Mofadal AI, Kanitpong K, Jiwattanakulpaisarn P. Analysis of pedestrian accident costs in sudan using the willingness-to-pay method. *Accid Anal Prev*. 2015;78:201-211.
40. Andersson H. Willingness to pay for road safety and estimates of the risk of death: Evidence from a swedish contingent valuation study. *Accid Anal Prev*. 2007;39(4):853-865.
41. Carlsson F, Johansson-Stenman O, Martinsson P. Is transport safety more valuable in the air? *J Risk Uncertainty*. 2004;28(2):147-163.
42. Determann D, Korfage IJ, Lambooij MS, et al. Acceptance of vaccinations in pandemic outbreaks: A discrete choice experiment. *PloS one*. 2014;9(7):e102505.
43. Dickinson K, Paskewitz S. Willingness to pay for mosquito control: How important is west Nile virus risk compared to the nuisance of mosquitoes? *Vector-borne and Zoonotic Diseases*. 2012;12(10):886-892.
44. Enneking U. Willingness-to-pay for safety improvements in the german meat sector: The case of the Q&S label. *Eur Rev Agric Econ*. 2004;31(2):205-223.
45. Flügel S, Rizzi LI, Veisten K, Elvik R, de Dios Ortúzar J. Car drivers' valuation of landslide risk reductions. *Saf Sci*. 2015;77:1-9.
46. Loureiro ML, Umberger WJ. A choice experiment model for beef: What US consumer responses tell us about relative preferences for food safety, country-of-origin labeling and traceability. *Food Policy*. 2007;32(4):496-514.
47. Patil S, Patruni B, Potoglou D, Robinson N. Public preference for data privacy–A pan-european study on metro/train surveillance. *Transportation Research Part A: Policy and Practice*. 2016;92:145-161.
48. Rizzi LI, de Dios Ortúzar J. Stated preference in the valuation of interurban road safety. *Accid Anal Prev*. 2003;35(1):9-22.
49. Smith VK, Mansfield C, Strong A. How should the health benefits of food safety programs be measured. *Adv Health Econ Health Serv Res*. 2014. 2014;24:161-202.
50. Iragüen P, de Dios Ortúzar J. Willingness-to-pay for reducing fatal accident risk in urban areas: An internet-based web page stated preference survey. *Accid Anal Prev*. 2004;36(4):513-524.
51. Viscusi WK. Valuing risks of death from terrorism and natural disasters. *J Risk Uncertainty*. 2009;38(3):191-213.
52. Bergstrom TC. Benefit-cost in a benevolent society. *Am Econ Rev*. 2006;96(1):339-351.
53. Bobinac A, van Exel J, Rutten F, Brouwer W. Caring for and caring about: Disentangling the caregiver effect and the family effect. *J Health Econ*. 2010;29(4):549-556.

54. Dolan P, Olsen JA, Menzel P, Richardson J. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Econ*. 2003;12(7):545-551.
55. Viscusi WK, Aldy JE. The value of a statistical life: A critical review of market estimates throughout the world. *J Risk Uncertainty*. 2003;27(1):5-76.
56. Hultkrantz L, Svensson M. The value of a statistical life in sweden: A review of the empirical literature. *Health Policy*. 2012;108(2-3):302-310.
57. Cuzzolin L, Atzei A, Fanos V. Off-label and unlicensed prescribing for newborns and children in different settings: A review of the literature and a consideration about drug safety. *Expert opinion on drug safety*. 2006;5(5):703-718.
58. Becker ML, Kallewaard M, Caspers PW, Visser LE, Leufkens HG, Stricker BH. Hospitalisations and emergency department visits due to drug–drug interactions: A literature review. *Pharmacoepidemiol Drug Saf*. 2007;16(6):641-651.

Appendix

Appendix 5.A – Exact search strings

Search Strings:

1. valu* AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
2. valu* AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
3. “shadow price” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
4. “shadow price” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
5. “willingness to pay” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
6. “willingness to pay” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
7. “willingness-to-pay” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
8. “willingness-to-pay” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
9. “willingness to accept” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
10. “willingness to accept” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
11. “willingness-to-accept” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
12. “willingness-to-accept” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
13. “discrete choice experiment” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
14. “discrete choice experiment” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
15. “stated preference” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
16. “stated preference” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review
17. “revealed preference” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
18. “revealed preference” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review

19. “contingent valuation” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”)
20. “contingent valuation” AND (safety OR security OR “uncertainty reduction” OR “risk reduction”) AND review

CHAPTER 6

Willingness to pay for health gains from mitigating infectious disease outbreaks

With Sebastian Himmler, Job van Exel & Werner Brouwer.

Submitted

Abstract

Recently, owing to the corona virus outbreak, pandemics and their effects have been at the forefront of the research agenda. While their benefits may be large, estimates of the perceived value of early warning systems for identifying, containing and mitigating infectious outbreaks remain scarce. This paper aims to show how potential health gains due to an international integrated early warning system might be valued, while also covering methodological issues such as probability weighting and sensitivity to scale.

This paper reports on a study into willingness-to-pay (WTP) in six European countries (Denmark, Germany, Hungary, Italy, the Netherlands and the UK) for health gains due to such an early warning system. The context in which health is gained (i.e. one year decrease in health or immediate death), those affected and the reduction in risk of contracting the disease (4% to 2%, 4% to 0% and 60% to 20%) generated by the early warning system are varied across seven scenarios. Using linear regression that includes various characteristics (such as income, age, awareness of outbreaks) we analyse this 'augmented' willingness-to-pay for a QALY (WTP-Q) for each of the seven scenarios. 'Augmented' refers to the possible inclusion of context specific elements of value, such as feelings of safety.

An initial WTP-Q value, for the basic scenario is estimated at €37,000. This can be interpreted as a threshold for investment per QALY into an early warning system for infectious disease outbreaks. Overall, WTP estimates move in the expected directions (e.g. higher risk reduction leads to higher WTP). However, changes in respondents' WTP for reductions in risk were not proportional to the magnitude of the change in risk reduction. It appears that differences between the obtained WTP-Q estimates were predominantly driven by changes in (expected) health gain, rather than WTP itself.

This study provided estimates of the monetary value of health gains in the context of a pandemic under seven scenarios which differ in terms of outcome, risk reduction and those affected. It also highlights the importance of future research into optimal ways of eliciting thresholds for investments in public health interventions, such as those necessary for infection identification, containment and mitigation of disease outbreak.

Introduction

Recently, pandemics and their effects, both on health and the economy, have been at the forefront of the research agenda. The current coronavirus pandemic, along with other recent infectious outbreaks such as Ebola, SARS and H1N1, has highlighted the importance of infectious disease control.[1] In 2015, an interdisciplinary research network that investigates the potential of an international integrated early warning system (EWS) for identifying, containing and mitigating large infectious outbreaks was initiated by the EU.[2] Currently only 81 countries have a national strategy for disaster risk reduction (such as an EWS), few explicitly mention pandemic threats,[3] and few, predominantly island nations (e.g. New Zealand, Taiwan, Singapore), have successfully mitigated the impact of COVID-19. An international EWS would upgrade existing regional systems and would strengthen the international cooperation required to address current and future threats of infectious diseases.[4] This would be especially beneficial to nations that do share (open) borders with other countries and may therefore find the migration of a virus harder to control.

Given the excessive costs of a pandemic – the coronavirus pandemic has been estimated to cost 1-2% of global GDP so far[5]– it is clearly both logical and desirable to prevent such situations, even though EWSs themselves are costly. The potential benefits of an effective warning system are clear and include a reduction in disease burden,[6] increased feelings of safety[7] and a smaller negative impact on the economy.[8] These potential benefits, however, are difficult to calculate as they are abstract, uncertain and, at least partially, may occur in the distant future. Additionally, if outbreaks are stopped early on, benefits of the system will not be visible to the general public. This, in turn means that acceptance of high spending contributions to EWSs may be limited, even if the benefits far outweigh the potential costs of outbreaks. Quantifying the perceived value of these systems informs us of what level of contributions society would find acceptable.

How to elicit a value for improved (feelings of) health safety in the context of recurrent threats of pandemics was addressed in a recent review by Perry-Duxbury et al.[9] They examined the methodologies commonly used for safety valuation across various topics within applied economics: specifically

transportation, environment and health. Of the 33 papers reviewed, 22 used willingness-to-pay (WTP), a form of the contingent valuation method, while the other 11 used discrete choice experiments. Himmler et al. (2020), building on this review, subsequently investigated individuals' WTP for an EWS and reported a mean estimate of €21.80 (median €10.00) per household per month from a sample from six European countries – Denmark, Germany, Hungary, Italy, the Netherlands and the UK (as measured *before* the current COVID-19 outbreak). They found that 80-90% of people would be willing to pay at least some additional tax towards an increase in health safety via an EWS.[7] Notably, these results were obtained from a survey question in which the magnitude of the health gain generated by an EWS was not explicitly mentioned or defined.

We can provide additional insights into the 'acceptable' level of contributions towards an EWS by exploring the value of an EWS while explicating its potential health gains will provide. This will further allow us to calculate implied WTP per unit of health gain (such as a quality adjusted life year (QALY)). Mean WTP (for a QALY) can then be viewed as an upper bound for spending on this preventative system, with the aggregate upper-bound being mean WTP multiplied by the number of contributing citizens. This estimate may differ from similar estimates focusing on curative interventions for several reasons. First, in the case of an unidentified pandemic, it is harder to predict who will be affected by the virus and to what extent. Second, preventative programmes such as an EWS lead to health benefits in the future (i.e. QALY gains) but potentially also to positive externalities now, such as the feeling of safety. It is therefore likely that WTP for an EWS does not solely consider (future) QALY gains, but also fewer tangible gains in (current) well-being.

This paper reports on a study into the WTP in six European countries for improved (feelings of) health safety in the context of preventing outbreaks. The study includes tests for sensitivity to scale and context using seven separate scenarios. They vary by the context in which health is gained (i.e. a one year decrease in health or immediate death), social inclusivity, and by how much an EWS reduces the risk of contracting the disease (4% to 2%, 4% to 0% and 60% to 20%). This is relevant as research has shown that, in WTP questionnaires, elicited values may vary substantially across the context in which QALYs are gained.[10]

Background

When estimating the WTP for a particular good like an EWS, and also when estimating the implied WTP per QALY (WTP-Q), there are various methodological issues to bear in mind. While some relate to the interpretation of the elicited value by the researcher, such as the definition of outliers and protest answers, there are also more conceptual concerns surrounding this approach. Anchoring - a pervasive judgement bias in human decision making – is one such issue that may systematically influence elicited WTP values.[11] While this can relate to the impact of a random value being shown before a valuation question is asked,[12] in the case of contingent valuation studies, it often refers to the initial bid for the first WTP question asked in the survey. This bid then works as an anchor in subsequent WTP answers.

Other relevant methodological concerns regarding WTP elicitation include sensitivity to scale and probability weighting. The probability of infection varies dramatically across different infectious diseases and is both relevant in itself and in its impact on the health gain generated by preventing the disease. Prospect theory suggests that people tend to overweight small probabilities and underweight larger ones.[13] Probability weighting, where probabilities are corrected for being treated non-linearly, is one way to address this issue. Bobinac et al.[14] demonstrate the impact of probability weighting on WTP for a Quality Adjusted Life Year (QALY) and report improved validity of the corrected estimates.

Sensitivity to scale relates to whether WTP estimates adequately reflect the size of the good on offer. Even if WTP responses are 'theoretically valid', (i.e. they increase with the size of the (health) gain offered), Bobinac et al. (2012) argue this is not a sufficient condition for the practical use or 'theoretical plausibility' of the estimate. For example, someone may be willing to pay 50€ for a health gain of 1 QALY but only 51€ for a health gain of 2 QALYs. While the determination of a 'practically meaningful' result may be arbitrary, we can test whether WTP estimates are (at least close to) proportional to the size of the gain.[15] Bobinac et al. found that WTP-Q estimates were highly insensitive to duration and type of health gain, which precludes establishing a 'unique' value per QALY.[15]

If there is insufficient sensitivity to scale in WTP responses, the valuation of WTP-Q will be strongly influenced by the size of the health gain offered in the WTP question. This could lead to a large range of possible estimations of WTP-Q if the EWS in question were to cover a variety of infectious diseases with varying rates of infection. Acknowledging sensitivity to scale and probability weighting (along with anchoring effects) in analysing WTP data, can enhance our understanding of the extent of uncertainty surrounding the estimates and may bring about improvements in eliciting and analysing WTP data. Given that the values elicited for health gains in the context of an EWS can inform the decision-making process, said insights will further inform investment choices for the warning system mentioned earlier.

Methods

Data

A contingent valuation experiment on the monetary valuation of a change in quality-of-life or life-years — asked within the context of an EWS for infectious disease — was conducted via a survey in general population samples from six European countries: Denmark, Germany, Hungary, Italy, The Netherlands and the UK, from February to March of 2018. The survey covered citizens between the ages of 18 and 65 so as to limit the population to income taxpayers, given that tax was the payment vehicle in the questionnaire. The samples aimed to be representative of national populations regarding age, gender and level of education. The sample size from each country was around 500 respondents. More specific information on the data collection itself and reasoning behind the survey design can be found in Himmler et al. (2020).

The survey consisted of WTP questions for seven different scenarios, (preceded by a warm-up question and a question in which risk was excluded; reported elsewhere (Himmler et al., 2020)), which differed in the risk of infection and the outcome of infection in relation to the effectiveness of an EWS. Some scenarios were asked from the socially-inclusive-personal (SIP) perspective, while others were asked from the socially exclusive perspective. Social exclusive scenarios covered situations where (1) only members of the opposite sex or (2) only children benefit from the warning system and receive the

health gain. It is important to note that the social perspective, while excluding the respondent, captures elements of altruism and possible closeness to the specific situation. For example, if the respondent is in a partnership with someone of the opposite sex, or has children, this may affect the observed WTP. These different groupings help to provide additional insight into both valuing health gains under a societal perspective and the question of equity weighting – whether or not health gains need to be weighted based on demographic characteristics such as age. For example, Bobinac et al. (2013)[16] found that WTP-Q values varied substantially (between €52,000 to €83,000) dependent on whether the societal perspective used includes the individual responding to the question or not. Regarding the scenario in which children alone are affected, it has been shown that individuals attach a higher weight to treatments for relatively younger patients (and the more severely ill).[17]

The seven scenarios are referred to as the *Basic, Certainty, High risk, Death, Equity, Social exclusive* and *Catastrophic* scenarios, and are summarised in Figure 6.1. Not all scenarios were posed to all respondents; all respondents answered the basic scenario before they were randomized to one of the three scenarios in stage 2, followed by one of the three scenarios in stage 3.

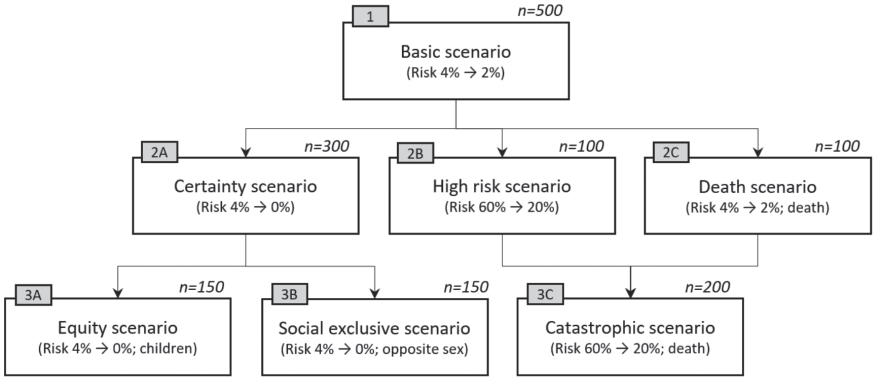


Figure 6.1 Scenarios & questionnaire flow (n = respondents per country)

The risk presented in each scenario is based on a change from one health state to another, i.e. becoming infected by the disease in question would cause a drop in health from 'Health state 1' (HS1) to 'Health State 2' (HS2).

Respondents were asked to value a description of two health states (HS1 and HS2), presented using the EQ-5D framework, at the start of the questionnaire. Respondents valued the health states using a visual analogue scale (VAS) with a range of 0 (worst imaginable health) to 100 (best imaginable health). The two health states differed on all dimensions, with HS1 dominating HS2 on all dimensions. The level scores for HS1 and HS2 were 11221 and 21433, respectively and HS1 was thus expected to be valued higher than HS2. After valuation, respondents were also explicitly asked to choose which of the two health states they thought was 'better'. In addition to being asked to value HS1 and HS2, respondents were asked to value perfect health, death, and their current health on a VAS scale.

In the WTP questions themselves, respondents were asked, using payment scales, what they would be willing to pay per month for a reduction in risk, with taxes as the payment vehicle. The risk reduction referred to an outcome of either one year in the worse health state (from the better health state) or immediate death (from the better health state). A complete example of a scenario description and the respective questions asked can be found in Appendix B. After completing the three WTP questions, questions were asked about demographic characteristics and awareness of infectious disease outbreaks.

Before analysing the data, all monetary values (i.e. WTP and income) from UK, Danish and Hungarian samples, were converted to euro values using average exchange rates during the month of sampling. Additionally, all monetary answers were adjusted for purchasing power parity (PPP) using the European 28 countries index as a base.[18] Further pre-processing of the data related to whether respondents understood the questionnaire, specifically when valuing health states. If a respondent valued HS1 higher than HS2 on the VAS scale but marked HS2 as the 'better' health state (and vice versa), there was reason to believe that they did not understand the questionnaire, and their responses were therefore labelled as 'illogical'. The other answer type that was labelled as illogical was if an individual valued perfect health the same as death. Around 24 percent of the sample responded 'illogically' according to at least one of the criteria, and their answers were removed from further analysis. There were additional forms of inconsistent responses that were included in the analysis

as they were not considered to be 'illogical', for example valuing HS2 higher than HS1 but being consistent in these answers. Further data on inconsistent responses is available in Appendix 6.A (Table 6.A.1). Finally, protest zero answers and WTP outliers were set to missing per question. Individuals who chose 0 as their maximum WTP had to select one of the following options to specify the reason for this answer: (i) not worth more than 0, (ii) unable to pay more than 0, (iii) government task, or (iv) the option to formulate another reason in an open text field. The first two options were considered to indicate a true WTP of zero, while "government task" indicated a protest zero. Entries in the open text field were evaluated and labelled as either true zero or protest zero. WTP values were labelled as outliers when they were greater than five percent of an individual's monthly household income (Table 6.1).

Table 6.1 True zeros, protest answers and outliers

Scenarios	# Initial observations	Excluded		Included	
		# Protest zeros (%)	# Responses WTP >5% of Income (%)	# Observations after cleaning	# True zeros (%)
Base	2148	186 (8.66)	79 (3.68)	1883	94 (4.99)
Certainty	1165	102 (8.75)	45 (3.86)	1018	66 (6.48)
High Risk	539	58 (10.76)	21 (3.90)	460	17 (3.69)
Death	444	26 (5.86)	20 (4.50)	398	14 (3.52)
Equity	589	40 (6.97)	20 (3.40)	529	23 (4.34)
Social Exclusive	576	99 (17.19)	15 (2.60)	462	36 (7.79)
Catastrophic	983	94 (9.56)	61 (6.21)	828	29 (3.50)

If an exact monthly household income was not provided by a respondent after selecting their monthly income range (in 1,000 decrements), the missing income was imputed based on the mean response from all participants within the income group chosen by the respondent. Income was also trimmed to exclude outliers that lay outside the 1st and 99th percentile. After this, descriptive statistics were prepared to show an overview of the sample (Table 6.2). The mean age of the sample was 42 years old and half of the sample was male. Mean income was €2,800 and the majority of the sample were married and employed (or self-employed).

Self-Assessed Health was measured as the EQ-5D-5L score provided by each

respondent in the questionnaire. Country specific value sets were used where available. When unavailable, value sets from countries in the same region in Europe for which a value set is available were used (Spain and Germany for Italy and Denmark, respectively).[19-23] Awareness of outbreaks was assessed using 12 questions on a Likert scale from 0 to 7, with higher values indicating a higher level of awareness of the existence, impact and likelihood of infectious diseases outbreaks. For more detail on this measure, we refer to Himmler et al. (2020).

Table 6.2 Summary statistics

	Mean (sd)	Min.	Max.	Observations
Age	41.61 (13.82)	18	65	2148
Monthly Household Income in EUR	2,531	234	10,796	1855
EQ-5D-5L score†	0.86 (0.18)	-0.2	1	2148
Awareness of outbreaks	54.75 (9.42)	12	84	2148
Male	0.49 (0.50)			2148
Married	0.57 (0.49)			2148
Education				
No tertiary education	0.59 (0.49)			2148
Tertiary education	0.41 (0.49)			2148
Employment				
Employed	0.55 (0.50)			2148
Self-employed	0.10 (0.30)			2148
Unemployed	0.06 (0.24)			2148
Homemaker	0.06 (0.24)			2148
Student	0.10 (0.31)			2148
Retired	0.08 (0.28)			2148

†Country specific value sets were used where available. Value sets from Spain and Germany were used for Italy and Denmark, respectively).

Willingness-to-pay differences for scenarios and determinants of WTP

A model was estimated with WTP values as the outcome in order to see the effects of the different elements that make up the seven scenarios on WTP, along with the effects of other determinants such as age and income (Equation 6.1). The elements of the scenarios considered in this question can all be traced back to Figure 6.1, the elements are: whether the risk difference is large (40%)

or small (4%,2%), whether it is certain that the EWS prevents all infections (4%-0%), whether the perspective is socially exclusive or SIP and whether the outcome of the disease is death or 1 year in worse health. The seven scenarios were designed by varying these elements – for example, the certainty scenario has a small risk difference, a certain drop in risk to zero percent and an outcome of 1 year in worse health. All of the aforementioned variables were treated as dummy variables. We also included a variable for the difference between better and worse health (or better health and death) to investigate whether the size of the health loss had an impact on individual WTP responses. Education, employment and country are all categorical variables and were treated as such. Given the skewed nature of income, log of income was used for ease of interpretation. We also transformed WTP to ln(WTP+1) — as WTP data were skewed and the plus one allowed for the inclusion of (true) zero WTP answers in the analysis.

$$\begin{aligned}\ln(WTP+1) = {} & \alpha \\ & + \beta_1 LargeRisk \\ & + \beta_2 CertainRisk + \beta_3 SocialExclusive + \beta_3 OutcomeDeath \\ & + \beta_4 HealthChange + \beta_5 married + \beta_6 \ln(income) + \beta_7 age + \beta_8 age^2 \\ & + \beta_9 male + \beta_{10} SAH + \beta_{11} education + \beta_{12} employment \\ & + \beta_{12} awareness + \beta_{13}\, country + \varepsilon\end{aligned} \tag{6.1}$$

Weighting, rescaling, and transforming of WTP data for WTP-Q estimation

As mentioned in previous sections, the aim of this paper was not only to assess the effect of different scenarios on WTP, but also to present WTP-Q estimates. To do so, we needed estimates for an individual's expected health gain for each scenario. In this survey, the outcome of introducing an EWS for infectious diseases led to some reduction in risk of either (1) a loss in quality of life for one year — from the previously designated better health state to the worse health state — or (2) a large loss in life years — from the better health state to immediate (and permanent) death. Therefore, the expected QALY gain (E(Q)) per scenario was be calculated as:

$$E(Q)_{scenario} = (Q(HS_{better}) - Q(HS_{worse\ or\ death})) * p_{scenario} * LY \tag{6.2}$$

Where $Q(HS_{better}) - Q(HS_{worse\ or\ death})$ is the quality of life loss from becoming infected based on respondent's VAS scores for the health states, and

$P_{scenario}$ represents the reduction in risk from the introduction of the warning system for the specific scenario. *LY* stands for life years and refers to the years gained due to the reduction in risk. These varied depending on the type of risk; in scenarios where there was only a change in QoL (for 1 year) before returning to better health, the increase in life-years was zero. However, for scenarios where death was the risk posed by the infectious disease, (Death and Catastrophic scenarios), the change in QoL was permanent and covered all the remaining life years. In the equation, *LY* represents remaining period life expectancy conditional on age, country, and gender, taken from the Human Mortality Database lifetables.[24]

Given that respondents valued death and perfect health at scores other than 0 and 100 on a VAS score,[25] the raw responses to HS1 and HS2 were rescaled, as recommended by and done in previous studies,[15,15,25] so that the endpoints of the VAS scale were effectively labelled as 'Perfect health' and 'Death' (the QALY scale), rather than their original labels of 'best imaginable health' and 'worst imaginable health' (Equation 6.3).

$$Rescaled\ HS_{1,2} = \frac{Raw\ HS_{1,2} - HS_{death}}{HS_{perfect} - HS_{death}} \tag{6.3}$$

Scores for health states worse than death were capped at -1, and at 1 for health states 'better than perfect health' which occur when a respondent values the health state in question higher than their score for perfect health.

In this paper, the effects of probability weighting on WTP-Q are also shown, as it has been found that adjusting for probability weighting improves the validity of WTP-Q estimates.[14] To do this, both non-weighted probabilities (i.e. those shown in the questionnaire) and weighted probabilities were used to estimate expected QALY gains. In order to be able to compare results with Bobinac et al.[14], we used the same three specifications of the probability weighting function: the one-parameter Tversky and Kahneman (TK) function,[13] the two-parameter Gonzalez and Wu (GW) function,[26] and the one-parameter Prelec (P) function,[27] using the function parameters estimated by Bleichrodt and Pinto Prades for TK and P functions,[28] and the parameters estimated by Abdellaoui for the GW function.[29] The weighted probabilities for the questionnaire (Appendix 6.A, Table 6.A.2) were used for $P_{scenario}$ in Equation 6.2.

Estimating WTP-Q using three different approaches

Once the (probability weighted and rescaled) expected QALY estimates were calculated, we calculated the WTP-Q per scenario. This is simply the scenario specific yearly WTP divided by the scenario specific expected QALY gain (Equation 6.4). There were several options for the calculation of WTP-Q: the disaggregated approach (i.e. mean of ratios) (Eq. 6.4), the aggregated approach (i.e. ratio of means) (Eq. 6.5), or linear regression of scenarios and control variables on WTP-Q estimates (Eq.6.6). Given that probability weighting was used to calculate expected QALY gain, the regression was run four separate times, once with no probability weighting and then with the three probability weighting specifications.

$$WTP\_Q_{scenario} = mean\left(\frac{WTP_{scenario}*12}{E(Q)_{scenario}}\right) \tag{6.4}$$

$$WTP\_Q_{scenario} = \frac{mean(WTP_{scenario} * 12)}{mean(E(Q)_{scenario})} \tag{6.5}$$

$$\begin{aligned}\ln(WTP\_Q) = {} & \alpha + \beta_1 scenarios + \beta_2 married + \beta_3 \ln(income) + \beta_4 age + \beta_5 age^2 \\ & + \beta_6 male + \beta_7 SAH + \beta_8 education + \beta_9 employment + \beta_{10} awareness \\ & + \beta_{11}\, country + \varepsilon\end{aligned} \tag{6.6}$$

Using the mean of ratios leads to a larger range of estimates due to more extreme values taken through to the final estimation. The ratio of means is the more common approach, as extreme responses have less of an impact on final estimates. These approaches can lead to markedly different outcomes.[30] A key benefit of using linear regression to predict WTP-Q is correcting for individual characteristics. Given the somewhat smaller sample sizes for some of the scenarios (e.g. death and high-risk) the respondent groups may not have identical characteristics. Therefore, if we wish to compare WTP-estimates for each scenario, the prediction approach is arguably the most apt for our current purposes.

Results

Willingness-to-pay per scenario

First, summary statistics of raw WTP by scenario can be examined. This not only signals general willingness to pay for an EWS under different scenarios, but also provides first insight into the theoretical validity and plausibility of the questionnaire and respondents' answers. Table 6.3 first of all shows that, for every scenario, 92% or more of respondents were willing to pay something (i.e. > 0) for an EWS. Moreover, from Table 6.3 it appears that the mean WTP for all other scenarios was higher than in the basic scenario, with all means lying between 19 and 26 Euros. For instance, an EWS would be valued about 34% higher when designed to prevent a catastrophic scenario, than in case of a 'basic' scenario. Given that there were also fewer protest answers for the catastrophic scenario, the aggregated value for an EWS in this case would be valued even higher than in a basic scenario. Standard deviations ranged between 26 and 35 Euros dependent on the scenario, showing the considerable heterogeneity between responses, and suggesting that differences in means are surrounded with uncertainty. The scenario with the largest monthly mean WTP (€25.67) was the catastrophic scenario, where respondents were asked about their WTP for a 40% reduction in risk (form 60%-20%) of immediate death. Finally, we present an approximation of an aggregate WTP per scenario across all six countries, using the same assumptions as Himmler and colleagues used – i.e. taking the median WTP and assuming that 50% of households (excluding the percentage of protest answers) would be eligible to pay the additional tax (calculations shown in Table 6.A.5 of Appendix 6.A).[7] These aggregate estimates range between 4.4 and 7.7 billion euros per year.

The regression on ln(WTP+1) - shown in Table 6.4 - highlights how each element of the scenarios affects WTP along with other key determinants and shows face validity for several of the elements in question. The results show that a large risk reduction (e.g. 40% rather than 2%) was significantly correlated with a 15% higher WTP. Given that the large risk reduction is 20-fold the small risk reduction, it is fair to conclude that this increase in WTP is not proportional to changes in risk. We also see that certainty (risk dropping to 0%) was significantly correlated with a 10% higher WTP. Death being the

outcome of the infectious disease (as opposed to a 1-year reduction in health) was correlated with a 20% increase in WTP. The social exclusive perspective (in which respondents themselves are not the beneficiaries of the EWS) had a limited negative, although insignificant, relationship with WTP (-4%). The magnitude of the difference between 'better' and 'worse' health states (or 'better' and death health states) had no visible effect on WTP responses. With regards to demographic characteristics, we found that income had a positive effect on WTP while self-assessed health (EQ-5D score) had a negative effect on WTP.

Table 6.3 Summary statistics monthly WTP by Scenario

Scenario	Mean (sd)	Median	Min.	Max.	% WTP > €0	Obs.	Total in millions
Basic	19.22 (27.70)	8.23	0	273.44	95.04	1894	4,908m
Certainty	21.88 (29.61)	10.80	0	212.50	93.55	1023	6,434m
High Risk	20.27 (26.84)	10.31	0	161.83	96.34	464	6,007m
Death	20.00 (27.40)	9.77	0	205.15	96.49	399	6,005m
Equity	22.78 (30.85)	11.72	0	257.73	95.67	531	7,118m
Social exclusive	20.21 (29.68)	8.25	0	214.84	92.26	465	4,460m
Catastrophic	25.71 (35.65)	13.00	0	323.77	96.52	834	7,676m

We also found an effect of home country on WTP, using Hungary as the base case. As hypothesized by Himmler et al. in the context of the initial estimate of WTP for an EWS, differences between countries can, at least partly, be explained by variations in Hofstede's cultural dimensions and trust in public institutions. Countries included in this study were in fact selected using the three most relevant dimensions of Hofstede's cultural dimensions: individualism vs collectivism, masculinity, and uncertainty avoidance.[7,31,31]

Table 6.4 OLS regression on WTP

	ln(WTP + 1)	
	Estimates	**Standard Error**
Large Risk	0.15*	0.04
Certain Risk (Risk reduced to 0%)	0.10*	0.04
Socially Exclusive	-0.03	0.05
Outcome Death	0.20*	0.07
Difference in health states	-0.01	<0.01
Ln(income)	0.31*	0.02
Age	-0.07*	0.01
Age^2	<0.01*	<0.01
Male	0.08*	0.03
EQ-5D Score	-0.18	0.12
Tertiary education[a]	0.06*	0.03
Married	0.17*	0.03
Employed[b]	0.16*	0.03
Self-employed[b]	0.08	0.06
Awareness of Outbreaks[†]	0.02*	0.00
Denmark[c]	0.13*	0.06
Germany[c]	0.12*	0.05
Italy[c]	0.41*	0.05
Netherlands[c]	0.19*	0.05
UK[c]	-0.01	0.05
Constant	0.73*	0.27
Observations		4824
R2 / R2 adjusted		0.205/0.201
AIC		13138.012

**P-values < 0.05* [a] Base case: No tertiary education, [b] Base case: Not working, [c] Base case: Hungary [†]scored from 12 to 84 (12 questions with 7 levels)

WTP-Q estimates

The OLS regression shown in Equation 6.6 allowed the isolation of the effects of different scenarios on log-transformed WTP-Q. The model was fitted for both probability weighted, and non-probability weighted WTP-Q estimates, so that comparisons between weighting approaches could be investigated. The results are displayed in Table 6.5. This regression-based approach used

the same control variables (not displayed) as in Table 6.4 and resulted in coefficients similar to those shown in Table 6.4 across models.

All scenarios had a significant (negative) effect on WTP-Q in comparison to the base scenario. In this case it is the difference in magnitude that is the most interesting. We see that (without weighting) the high risk, death and catastrophic scenarios led to a WTP-Q estimate that was at least 96% lower than the WTP-Q estimate in the base case, ceteris paribus. Certainty, equity, and social exclusive scenarios lead to WTP-Q that was around 30% lower.

Table 6.5 OLS regression on ln(WTP-Q) estimates

	No PW†		TK†		Prelec†		GW†	
Predictors	**Estimates**	**SE**	**Estimates**	**SE**	**Estimates**	**SE**	**Estimates**	**SE**
Certainty scenario[a]	-0.28*	0.04	-0.36*	0.03	-0.59*	0.03	-0.37*	0.03
High risk scenario[a]	-0.96*	0.05	-0.59*	0.04	-0.67*	0.04	-0.63*	0.04
Death scenario[a]	-1.12*	0.05	-0.85*	0.04	-0.94*	0.04	-0.87*	0.04
Equity scenario[a]	-0.28*	0.04	-0.36*	0.04	-0.59*	0.04	-0.37*	0.04
Social exclusive scenario[a]	-0.36*	0.05	-0.41*	0.04	-0.63*	0.04	-0.42*	0.04
Catastrophic scenario[a]	-1.15*	0.03	-0.87*	0.03	-0.96*	0.03	-0.89*	0.03
Observations	3966		3966		3966		3966	
R2 / R2 adjusted	0.342 / 0.338		0.289 / 0.285		0.326 / 0.323		0.293 / 0.290	
AIC	8910.016		7347.718		7305.335		7452.630	

N.B. Same control variables used as in Table 6.4
[a] Base case: Basic scenario, [b] Base case: No secondary school diploma, [c] Base case: Not working, [d] Base case: Hungary
[‡] UK Value set used, [†]scored from 12 to 84 (12 questions with 7 levels)
Note: WTP-Q estimates only calculated where EQ was not equal to 0, so as to avoid infinite values.
[†]No PW = No probability weighting. TK = functional form from Tversky & Kahneman[13] $w(p) = \frac{p^{\gamma}}{[p^{\gamma}+(1-p)^{\gamma}]^{\frac{1}{\gamma}}}$;

P = functional form Prelec[27]: ;
GW = functional form Gonzalez and W[26]: $w(p) = \frac{\delta p^{\gamma}}{[\delta p^{\gamma}+(1-p)]^{\gamma}}$

While the linear regression estimates displayed in Table 6.5 provided us with the relevant information on the impact of scenarios and probability weighting on WTP-Q, Table 6.6 gives absolute values for the WTP for a health gain in the context of a pandemic, for each scenario, methodological approach and probability weighting functional form.

Results varied significantly depending on methodological approach and the scenario in question. For example, in the Basic scenario, the ratio of means approach (weighted using TK form) led to a WTP-Q of €19,000, which was about half that of the estimate reached using the mean of ratios, €43,000. There was a similar trend for all scenarios, with the linear regression estimates lying between estimates from the other two approaches. Looking at the scenarios, WTP-Q ranged from €76,000 (basic scenario) to €26 (catastrophic scenario) across methods and weighting-approaches. Even when considering a single methodological approach and probability-weighting functional form differences in WTP per QALY across scenarios were still large.

Scenarios in which health gains were largest, such as where death was the result of infection, and where reduction in risk was greatest, had the lowest WTP-Q estimates. Certainty, equity and social exclusive scenarios, which all had the same reduction in risk and health gains, resulted in similar estimates, in the area of €14,000-€15,000.

Table 6.6 WTP-Q for each scenario calculated using Mean of Ratios, Ratio of Means and Linear Regression (LR) Prediction

	Basic			Certainty			High risk			Death			Equity			Social exclusive			Catastrophic		
	Mean of ratios	Ratio of means	LR	Mean of ratios	Ratio of means	LR	Mean of ratios	Ratio of means	LR	Mean of ratios	Ratio of means	LR	Mean of ratios	Ratio of means	LR	Mean of ratios	Ratio of means	LR	Mean of ratios	Ratio of means	LR
No PW[a]	75,979	33,002	65,717	40,411	18,811	34,531	4,337	1,745	3,566	435	385	435	42,902	19,502	36,877	38,817	17,398	34,215	27	25	26
TK[a]	43,323	18,818	37,471	16,017	7,456	13,686	6,890	2,772	5,665	248	220	248	17,004	7,730	14,616	15,385	6,896	13,561	43	39	42
P[a]	53,084	23,057	45,914	10,432	4,856	8,914	7,834	3,152	6,442	304	269	304	11,076	5,035	9,520	10,021	4,491	8,833	49	45	47
GW[a]	45,398	19,719	39,266	16,802	7,821	14,357	6,475	2,606	5,324	260	230	260	17,837	8,108	15,332	16,139	7,233	14,226	40	37	39
N	1,608	2,083	1,419	873	1,141	766	396	501	355	438	438	388	449	594	397	399	515	347	905	905	800

[a] No PW = No probability weighting. TK = functional form from Tversky & Kahneman[13]: $w(p) = \frac{p^{\gamma}}{[p^{\gamma}+(1-p)^{\gamma}]^{\frac{1}{\gamma}}}$;
P = functional form Prelec[27]: $w(p) = \exp(-(-\ln)^{\alpha})$;

GW = functional form Gonzalez and W[26]: $w(p) = \frac{\delta p^{\gamma}}{[\delta p^{\gamma}+(1-p)]^{\gamma}}$

Discussion

The need for and potential benefits of EWSs for infectious disease and foodborne outbreaks have rarely been more apparent than in recent times, and thus research on investment into EWSs is of relevance. The aim of this paper was to add to the available literature by providing estimates of WTP for EWSs as well as WTP-Q in the context of infectious disease outbreaks, both for seven different scenarios. These scenarios were chosen to investigate the impact of changes in risk reduction, the context in which health is gained (decrease in health or death) and whether a SIP or social exclusive perspective is taken. In doing so, some additional methodological issues came to light, specifically the (in)sensitivity to scale of WTP and the subsequent effects on WTP-Q.

Our design allowed us to consider both directly elicited WTP for an EWS in different contexts as well as calculating the implied WTP-Q estimates. In that respect, it is important to note that WTP values are the estimates that were directly elicited via the questionnaire. WTP-Q estimates were calculated on the basis of the indicated WTP and our calculations of the implied health gains (in terms of QALYs). We found that WTP was driven, at least in part, by the size and certainty of the risk presented to respondents, the outcome of the disease in question, and various demographic variables such as income and age. Given that all coefficients moved in the expected direction, our WTP results can be considered to be 'theoretically valid'. Using these results to estimate aggregate values of an EWS, we found total values for an EWS between 4.9 and 7.7 billion euros per year for the six European countries covered in the questionnaire – the largest value in relation to the catastrophic scenario and the smallest when presenting the social exclusive scenario (a difference of 58%). When translating these results into WTP-Q estimates, the range of estimates produced was large, especially due to the fact that changes in WTP were not (near) proportional to changes in implied health gains, and, therefore, the practical plausibility of some estimates may be questioned.

WTP-Q was estimated using three separate approaches (mean of ratios, ratios of means and linear regression), and three approaches for probability weighting. This showcased the effects of different methodological approaches

on the WTP-Q estimates. Given that the linear regression method controls for demographic variables that could affect WTP-Q, we focus on those results in this discussion section, specifically those that were probability weighted with the Tversky & Kahneman function. Regarding the results, an initial WTP-Q value, for the basic scenario was estimated at €37,000. This can be interpreted as a threshold for investment per QALY into an EWS for infectious disease outbreaks. Note that this WTP-Q estimates may include broader benefits related to the EWS, and therefore could be seen as a 'WTP for a QALY+', as for instance feelings of safety may be expected to be generated by an EWS. If so, respondents could consider and value these elements when providing their WTP. It is important to note that these included elements could contain intangible benefits, like feelings of safety, but also cost-savings — respondents may for instance be willing to pay more money now in order to avoid a reduction in income later. Given that this questionnaire was answered before the COVID-19 outbreak, this may have been less of a consideration than post-March 2020, but still it cannot be excluded. This may also affect some scenarios more than others. WTP responses may also include reflect expectations about income or productivity in particular situations that would normally be captured separately in an economic evaluation, highlighting the need to avoid double-counting.

The base-case was only one of seven scenarios examined in this paper. When comparing the results from these seven scenarios, the implied WTP-Q estimates varied very considerably. This was not only the case when comparing different methodological approaches to estimating WTP-Q from WTP data and implied health gains in the different scenarios, but especially related to the fact that respondents' WTP was clearly not proportional to the magnitude of the implied health gains. This holds both in terms of the size of the risk reduction and in terms of the size of the difference between health states. We therefore found that differences between the obtained WTP-Q estimates, in which WTP is divided by expected health gain per scenario, were predominantly driven by this denominator, rather than the differences in WTP responses themselves. This lack of sensitivity (at least in relation to implied health gains) may be related to the fact that we were valuing a system for disease surveillance, which is preventative (rather than curative) in nature, and is concerned with

hypothetical, future emergencies. Perhaps that somewhat drew attention away from the size of the health gains on offer, although sensitivity to scale is a more commonly observed phenomenon. It would be interesting to see the same scenarios attached to a curative intervention when eliciting WTP per QALY.

The main results of this paper are in line with previous work on the topic.[14,14,15,32,32] 'Raw' WTP results are theoretically valid, and move in the directions expected; individuals are willing to pay more for the certainty and death scenarios than the basic scenario, although only results for the certainty scenario were significant. These estimates are, however, far from theoretically plausible, given their lack of near-proportionality in relation to the implied size of the health gains. This is worrisome, given the consequences of this combination on the subsequent WTP-Q measures. When WTP estimates lack sensitivity to scale, it results in WTP-Q estimates that are highly variable and hard to consider theoretically valid, when comparing across scenarios. For example, while the non-probability weighted basic scenario estimate of €66,000 per QALY is in line with previous estimates, [14,14,15,32,32] in the same sample the estimates from the certainty and catastrophic scenarios result in much lower WTP-Q estimates. Even though low WTP-Q estimates have been reported before — where QoL improving interventions were valued lower than life-extending interventions[33] — the fact that a such a wide range of values was found in our results raises the issue of which value is 'correct'.

If WTP were sufficiently sensitive to scale, then the WTP-Q estimates should remain relatively similar across scenarios. Given that this is not the case, and it appears unlikely that the resulting WTP-Q reflect 'true valuations' of QALY gains in different contexts, this emphasises the difficulty of finding a unique and universally valid WTP-Q estimate. It is clear that methodological assumptions made, such as whether to use probability weighting (and even which weighting function to use), the framing of WTP questions, possible anchoring effects, and the approach used to estimate WTP-Q, strongly affect the outcomes of the research.

There are limitations to our study. First, while anchoring may be responsible for later WTP estimates being close in magnitude to the WTP for the basic scenario,

it is not possible to fully empirically investigate this using the data available – there was no change in ordering of questions for any respondents. That being said, in auxiliary analysis (Appendix 6.A, Tables 6.A.3 & 6.A.4) we found that the higher the responses to the initial question posed to respondents, the greater the WTP for the next questions. However, separating anchoring effects from other issues, such as ordering effects or simply individual preferences, is not possible. It is advisable, when collecting data for the estimation of WTP-Q, to provide multiple orderings of scenarios so that anchoring can be investigated. Second, while our sample is representative of the population for the basic scenario with regards to sex, age and level of education, it is not representative of other factors which may impact WTP-Q responses. Furthermore, it is likely, when respondents are randomly assigned to the other six scenario questions, that the sample will not be fully representative of each country's population. If, however, linear progression models are used to predict WTP-Q, many of these demographic characteristics are accounted for. Third, life-tables were used to estimate expected health gains when death was an outcome in a specific scenario. It may however be that life-expectancy from the Human Mortality Database is different from the subjective life-expectancy an individual had in mind when filling in the questionnaire (for example respondents may over-estimate their life-expectancy).[34] Fourth, the data used were collected via online questionnaires in March of 2018, when there was no global pandemic or even knowledge of COVID-19. If the questionnaire were rerun today, it is possible that WTP (and therefore WTP-Q) estimates would be higher, as people are more aware of the real-life implications of a pandemic. This would be a beneficial starting point for future research, especially given the now more obvious need for EWSs. Future research would also benefit from further investigation into questionnaire design when eliciting WTP-Q, and into the effects that using such estimates of cost-effectiveness thresholds may have on policy. For example, what precisely is included in the WTP estimates is currently somewhat of a black box – it would be useful to try and disaggregate some of the elements that may be included in the valuation (e.g. QALYs, safety, cost-savings, productivity) in future questionnaires. Furthermore, it is possible that by carrying out the questionnaire online, there was less engagement with the questionnaire and (hence) a lower quality of the answers than if the questionnaire had been completed in person. Carrying out (at least

a percentage of) in person questionnaires alongside the online questionnaire would have highlighted any systematic differences in responses.[35]

There were also strengths to this study, a key one being that data were collected from six European countries, which were chosen specifically to cover different ends of the spectrum with regards to three of Hofstede's cultural dimensions: individualism vs collectivism, masculinity, and uncertainty avoidance.[31] This may enhance the generalisability (to the European level) of our results. Furthermore, the use of a two-step payment scale approach and an open question to elicit WTP leads to more precise estimates of WTP, as it combines two (linked) elicitation questions.[36-38] Additionally, the use of three separate estimation approaches, provides a thorough overview of the impact of methodological choices on WTP-Q estimation.

This study has provided estimates of the monetary value of a health gain in the context of a pandemic under seven scenarios which differ in terms of outcome, risk reduction and those affected. The effects of probability weighting were also investigated. While the (probability-weighted) WTP-Q for the basic scenario lies somewhere within an expected range for previous estimates of WTP-Q in the literature (€37,000), the large variation in estimates for the six other scenarios is concerning if this threshold were to be used in the decision-making context. That being said, it is clear from the aggregate estimates of WTP that an EWS is an intervention that is valued in European populations. This is encouraging when we consider that EWS interventions are one of the ways to avoid repeat pandemics and their negative consequences on health and welfare.

WTP-Q estimates also represent first estimates of a possible cost-effectiveness threshold for the reimbursement of preventive interventions, specifically those that generate feelings of safety along with QALY gains. However, they also raise several questions around the choice of questionnaire design when trying to elicit such values, the reliability of using contingent valuation to elicit WTP for a QALY estimates, and whether using an 'augmented' WTP-Q estimate brings value or overcomplicates an already complex area of decision-making. Given that these are initial estimates with much uncertainty, we recommend, at the very least, that these results be considered with caution, and that there

is a need for research into how intervention specific thresholds (that may cover outcomes outside of the QALY) can or should fit into decision making around public health.

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The Lancet*. 2020;395(10223):470-473.
2. Skiby JE. About COMPARE. Updated 2015. Accessed October, 2016. http://www.compare-europe.eu/About.
3. Fearnley CJ, Dixon D. Early warning systems for pandemics: Lessons learned from natural hazards. *International Journal of Disaster Risk Reduction*. 2020.
4. Nuzzo J, Shearer M. International engagement is critical to fighting epidemics. *Health Security*. 2017;15(1):33-35.
5. Ali SA, Baloch M, Ahmed N, Ali AA, Iqbal A. The outbreak of coronavirus disease 2019 (COVID-19)—An emerging global health threat. *Journal of infection and public health*. 2020.
6. Hernandez M, Milechin LE, Davis SK, DeLaura R, Claypool KT, Swiston A. The impact of host-based early warning on disease outbreaks. *medRxiv*. 2020.
7. Himmler S, van Exel J, Perry-Duxbury M, Brouwer W. Willingness to pay for an early warning system for infectious diseases. *The European Journal of Health Economics*. 2020:1-11.
8. Kellerborg K, Brouwer W, van Baal P. Costs and benefits of early response in the ebola virus disease outbreak in sierra leone. *Cost Effectiveness and Resource Allocation*. 2020;18(1):1-9.
9. Perry-Duxbury M, van Exel J, Brouwer W. How to value safety in economic evaluations in health care? A review of applications in different sectors. *The European Journal of Health Economics*. 2019;20(7):1041-1061.
10. Ryen L, Svensson M. The willingness to pay for a quality adjusted life year: A review of the empirical literature. *Health Econ*. 2015;24(10):1289-1301.
11. Van Exel J, Brouwer W, van den Berg B, Koopmanschap MA. With a little help from an anchor: Discussion and evidence of anchoring effects in contingent valuation. *The journal of socio-economics*. 2006;35(5):836-853.
12. Ariely D, Loewenstein G, Prelec D. "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly journal of economics*. 2003;118(1):73-106.
13. Tversky A, Kahneman D. Advances in prospect theory: Cumulative representation of uncertainty. *J Risk Uncertainty*. 1992;5(4):297-323.
14. Bobinac A, van Exel J, Rutten F, Brouwer W. The value of a QALY: Individual willingness to pay for health gains under risk. *Pharmacoeconomics*. 2014;32(1):75-86.
15. Bobinac A, van Exel J, Rutten F, Brouwer W. Get more, pay more? an elaborate test of construct validity of willingness to pay per QALY estimates obtained through contingent valuation. *J Health Econ*. 2012;31(1):158-168.
16. Bobinac A, van Exel J, Rutten F, Brouwer W. Valuing QALY gains by applying a societal perspective. *Health Econ*. 2013;22(10):1272-1281.

17. Reckers-Droog V, van Exel J, Brouwer W. Equity weights for priority setting in healthcare: Severity, age, or both? *Value in Health*. 2019;22(12):1441-1449.
18. Eurostat. GDP per capita in PPS. Updated February 2021. Accessed February 2021. https://ec.europa.eu/eurostat/databrowser/view/tec00114/default/table?lang=en.
19. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5 D-5 L value set for England. *Health Econ*. 2018;27(1):7-22.
20. Versteegh MM, Vermeulen KM, Evers SM, De Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value in health*. 2016;19(4):343-352.
21. Ludwig K, von der Schulenburg, J-Matthias Graf, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663-674.
22. Ramos-Goñi JM, Craig BM, Oppe M, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value in Health*. 2018;21(5):596-604.
23. Rencz F, Brodszky V, Gulácsi L, et al. Parallel valuation of the EQ-5D-3L and EQ-5D-5L by time trade-off in Hungary. *Value in Health*. 2020;23(9):1235-1245.
24. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Human mortality database. Available at www.mortality.org or www.humanmortality.de. Accessed October, 2020.
25. Oppe M, Devlin NJ, Szende A. *EQ-5D value sets: Inventory, comparative review and user guide*. Springer; 2007.
26. Wu G, Gonzalez R. Curvature of the probability weighting function. *Management science*. 1996;42(12):1676-1690.
27. Prelec D. The probability weighting function. *Econometrica*. 1998:497-527.
28. Bleichrodt H, Pinto JL. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management science*. 2000;46(11):1485-1496.
29. Abdellaoui M. Parameter-free elicitation of utility and probability weighting functions. *Management science*. 2000;46(11):1497-1512.
30. Gyrd-Hansen D, Kjær T. Disentangling WTP per QALY data: Different analytical approaches, different answers. *Health Econ*. 2012;21(3):222-237.
31. Hofstede G. National cultures in four dimensions: A research-based theory of cultural differences among nations. *International Studies of Management & Organization*. 1983;13(1-2):46-74.
32. Gyrd-Hansen D. Willingness to pay for a QALY. *Pharmacoeconomics*. 2005;23(5):423-432.
33. Pennington M, Baker R, Brouwer W, et al. Comparing WTP values of different types of QALY gain elicited from the general public. *Health Econ*. 2015;24(3):280-293.
34. Brouwer W, van Exel J. Expectations regarding length and health related quality of life: Some empirical findings. *Soc Sci Med*. 2005;61(5):1083-1094.

35. Wright KB. Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of computer-mediated communication*. 2005;10(3):JCMC1034.
36. Johnson FR, Banzhaf MR, Desvousges WH. Willingness to pay for improved respiratory and cardiovascular health: A multiple-format, stated-preference approach. *Health Econ*. 2000;9(4):295-317.
37. Donaldson C, Thomas R, Torgerson DJ. Validity of open-ended and payment scale approaches to eliciting willingness to pay. *Appl Econ*. 1997;29(1):79-84.
38. Bobinac A, Van Exel J, Rutten F, Brouwer W. Willingness to pay for a quality-adjusted life-year: The individual perspective. *Value in Health*. 2010;13(8):1046-1055.

Appendices

Appendix 6.A

Table 6.A.1 Expected answers for health state valuation

Statement	# Responses	% of sample
Health State 1 > Health State 2 — Consistently	2267	72.20
Inconsistent answers for Health State 1 > Health State 2	749	23.85
Health State 1 > Death	2105	67.04
Health State 2 > Death	1970	62.74
All answers as expected (HS1>HS2, HS1 > Death, HS2> Death)	1422	45.29

Table 6.A.2 Probability weighting and parameter estimates

Non weighted probabilities for each scenario	Weighted probabilities		
	TK[a] γ = 0.674 for losses	P[a] α = 0.533	GW[a] δ = 0.84, γ = 0.65 for losses
Basic scenario			
0.04	0.10	0.15	0.10
0.02	0.07	0.13	0.06
Certainty			
0.04	0.10	0.15	0.10
0.00	0.00	0.00	0.00
High risk			
0.60	0.51	0.50	0.52
0.20	0.26	0.13	0.25
Death			
0.04	0.10	0.15	0.10
0.02	0.70	0.13	0.06
Equity			
0.04	0.10	0.15	0.10
0.00	0.00	0.00	0.00
Social exclusive			
0.04	0.10	0.15	0.10
0.00	0.00	0.00	0.00
Catastrophic			
0.60	0.51	0.50	0.52
0.20	0.26	0.13	0.25

[a] TK = functional form from Tversky & Kahneman[13]: $w(p) = \frac{p^{\gamma}}{[p^{\gamma}+(1-p)^{\gamma}]^{\frac{1}{\gamma}}}$;

P = functional form Prelec[27]: $w(p) = \exp(-(-\ln)^{\alpha})$;

GW = functional form Gonzalez and W[26]: $w(p) = \frac{\delta p^{\gamma}}{[\delta p^{\gamma}+(1-p)]^{\gamma}}$

Table 6.A.3 Effect of initial EWS answer on Basic Scenario answer

	Basic Scenario answer			
Predictors	**Estimates**	**std. Error**	**Estimates**	**std. Error**
Initial answer (reported elsewhere)[†]	0.43*	0.01	0.43*	0.01
Income	5.04*	0.92	5.00*	0.92
Age	-0.29*	0.04	-0.29*	0.04
Male	2.76*	1.13	2.66*	1.13
Tertiary education	0.13	1.20	0.11	1.20
Married	1.05	1.24	1.08	1.24
Employed[a]	1.09	1.33	1.23	1.33
Self-employed[a]	2.44	2.26	2.33	2.28
Denmark[b]	-0.89	2.28	-1.22	2.29
Germany[b]	-0.03	2.06	-0.45	2.06
Italy[b]	2.71	2.06	2.18	2.09
Netherlands[b]	-0.16	2.09	-0.54	2.10
UK[b]	-1.30	1.94	-1.71	1.95
Answer Pattern 2A-3B [a]			2.37*	1.52
Answer Pattern 2B-3C [a]			-0.50	1.52
Answer Pattern 2C-3C [a]			2.86	1.62
Constant	-18.32*	6.34	-18.64	7.32
Observations		1222		1222
R2 / R2 adjusted		0.491 / 0.485		0.494 / 0.487
AIC		10720.443		10719.853

**P-values < 0.05*

[†] Himmler et al.[7] Descriptive statistics using our selection criteria:

Mean (sd): 19.75 (36.34) , Min: 0 , Max: 1031

[a] Base case: Answer patterns 2A-3A (see Figure 6.1) [b] Base case: Not working, [c] Base case: Hungary

Table 6.A.4 Effect of Basic Scenario answer on following answer

	Second scenario answer			
Predictors	**Estimates**	**std. Error**	**Estimates**	**std. Error**
Initial answer (reported elsewhere)[†]	0.88	0.02	0.88*	0.02
Income	2.31	0.68	2.37*	0.68
Age	-0.07	0.03	-0.07*	0.03
Male	1.21	0.80	1.19	0.81
Tertiary education	1.58	0.85	1.55	0.86
Married	0.08	0.89	0.08	0.88
Employed [b]	-1.69	0.95	-1.72	0.95
Self-employed [b]	-3.03	1.61	-2.88	1.62
Denmark [c]	0.02	1.63	0.06	1.64
Germany [c]	-2.77	1.47	-2.59	1.48
Italy [c]	-1.67	1.48	-1.59	1.50
Netherlands [c]	-2.84	1.48	-2.60	1.49
UK [c]	-1.63	1.38	-1.44	1.39
Answer Pattern 2A-3B [a]			0.37	1.08
Answer Pattern 2B-3C [a]			1.64	1.08
Answer Pattern 2C-3C [a]			-0.94	1.15
Constant	-8.97	4.62	-9.73	4.66
Observations		1194		1194
R2 / R2 adjusted		0.763 / 0.760		0.764 / 0.760
AIC		9636.755		9637.663

**P-values < 0.0*
[†] Himmler et al.[7] Descriptive statistics using our selection criteria:
Mean (sd): 19.75 (36.34) , Min: 0 , Max: 1031
[a] Base case: Answer patterns 2A-3A (see Figure 6.1) [b] Base case: Not working, [c] Base case: Hungary

Table 6.A.5 – Aggregate WTP

Scenario	Median WTP in € per month[1]	No. of household in million[2]	% Protest zeros[3]	% HH paying tax	Total in millions
Basic	8.23	108.81m	8.66%	50%	4,908m
Certainty	10.80	108.81m	8.75%	50%	6,434m
High Risk	10.31	108.81m	10.76%	50%	6,007m
Death	9.77	108.81m	5.86%	50%	6,005m
Equity	11.72	108.81m	6.97	50%	7,118
Social exclusive	8.25	108.81m	17.19	50%	4,460m
Catastrophic	13.00	108.81m	9.56	50%	7,676m

[1] Based on Table 6.A.3

[2] Assumption based on the share of households with income taxpayer who are eligible for additional taxation

[3] Based on Table 6.A.1

Appendix 6.B

Example of scenario description and resulting questions

[Introduction presented before module 3]

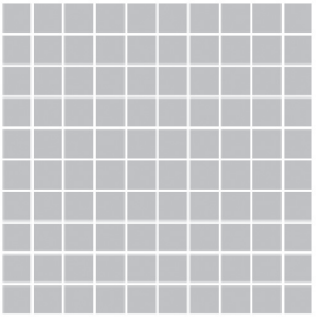
The following three questions are similar to the two previous questions and are based on the two health states (health state 1 and health state 2) that you rated on a rating scale in the beginning of this survey. For your convenience, the health states that you rated on a rating scale will follow you through this section of the survey.

1A. [Introduction]

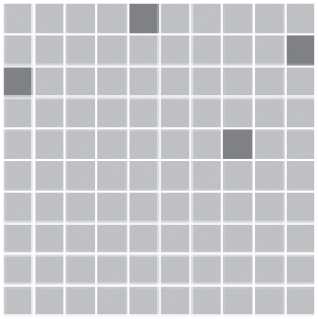
Suppose 100,000 people in your country are in the health state that you previously selected as the better one. Due to a virus, these people face a 4% risk of becoming infected with a virus within the next three months. It is not known who belongs to this group. The health of those infected deteriorates to the health state you just selected as the worse one. They will remain in the worse health state for a year, and then return to the better health state.

[explanation of risk with and without an early warning system]

To illustrate what the percentage risk means, click on one of the squares in the picture below [shown in greyscale here].



[The following text and picture will appear when respondent clicks on one of the squares]

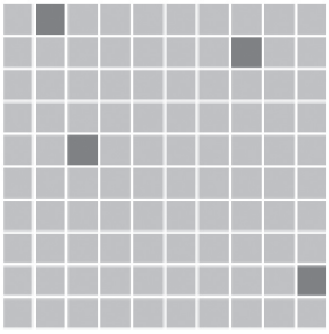


In a group of 100 people, 4 in 100 (4 %) will become sick and 96 will not become sick. To get a sense of what an integrated early warning system would mean for a group of 100 people, please click again on one of the squares in the picture below.

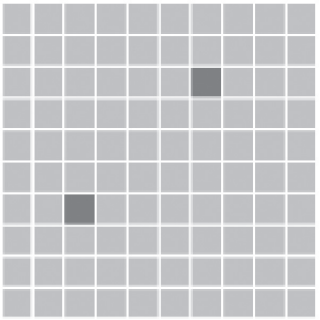
The chance that the square you selected turned red was 4%

Imagine that the risk of being infected by this virus can be reduced from 4% to 2% through an international, integrated early warning system that helps contain and mitigate [mouse cursor] the spread of the virus. However, establishing and maintaining such a system is not without costs. Suppose the funding would take place through taxation, paid by all eligible people in your country (above 18 years of age). This taxation would be collected through monthly installments for the duration of one year.

To illustrate what the percentage risk means, click on one of the squares in the picture below [The following text and picture will appear]



[The following text and picture will appear when respondent clicks on one of the squares]



In a group of 100 people, 2 in 100 means (2 %) will become sick and 98 will not become sick.

The chance that the square you selected turned red was 2%

1B [Payment scale]

Suppose you would have to pay this monthly instalment starting now. Please consider the amounts on the scale below, from left to right, and select the amounts you would definitely be willing to pay per month for establishing and maintaining this integrated early warning system. This amount would then be a taxation paid by all eligible people in your country. Please keep in mind your ability to pay (your net monthly household income).

€0	€1	€2	€3	€4	€5	€10	€15	€20	€25	€30	€40	€50	€60	€70	€80	€100	€120	€150	€200	more
o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o

[Save selected column as a variable. If € 0 is selected, proceed to 1D. If 'more' is selected, proceed to 1E. Otherwise, save selected amount as 'X' (for later use in 1G) and proceed to 1C.]

1C. [Payment scale]

Now consider the same amounts below, now from right to left, and select the amounts you would definitely be willing to not pay per month for establishing and maintaining

this integrated early warning system. This amount would then be a taxation paid by all eligible people in your country. Please keep in mind your ability to pay (your net monthly household income).

€0	€1	€2	€3	€4	€5	€10	€15	€20	€25	€30	€40	€50	€60	€70	€80	€100	€120	€150	€200	more
o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o

[Save selected column as a variable. If 'more' is selected, proceed to 1F. Otherwise, save selected amount as 'Y' (for later use in 1G) and proceed to 1G.]

1D. [Follow-up question if 3B = € 0]

You have indicated that the maximum amount you would be willing to pay to establish and maintain an integrated early warning system is € 0.

Please indicate below, the reason behind this preference.

- ❑ Having an integrated early warning system is not worth more than € 0 to me
- ❑ I am unable to pay more than € 0
- ❑ Government should pay for this
- ❑ Other reason, namely _____ (please specify; open text field, required if option is selected)

1E. [Follow-up question if respondent indicates 'more' in questions 1B; if possible on the same screen]

You have indicated that you would be willing to pay more than the maximum amount presented on the scale for establishing and maintaining an integrated early warning system, **which decreases your risk of becoming infected from 4% to 2%**. Please indicate in the box below the maximum amount that **you would be willing to pay** per month. This amount would then be a taxation paid by all eligible people in your country. Please keep in mind your ability to pay (your net monthly household income).

€

1F. [Follow-up question if respondent indicates 'more' in questions 3C; if possible on the same screen]

Please indicate in the box below the maximum amount that **you would not be willing to pay** for establishing and maintaining an integrated early warning system, **which decreases your risk of becoming infected from 4% to 2%.** Please keep in mind your ability to pay (your net monthly household income).

€

[Save indicated amount as 'Y3' (for later use in 3G) and proceed to question 3G]

1G. [Open-ended question; on new screen]

You have indicated that **you would definitely pay €[insert X]** per month and that **you would definitely not pay €[insert Y]** per month for establishing and maintaining an integrated early warning system, **which decreases your risk of becoming infected from 4% to 2%**. Please indicate in the box below the amount between €[insert X] and €[insert Y] that is closest to the maximum that you would be willing to pay per month. This amount would then be a taxation paid by all eligible people in your country. Please keep in mind your ability to pay (your net monthly household income).

€

CHAPTER 7

Discussion

The aim of this thesis was to investigate how several future health events, in terms of costs and benefits, can be accounted for in the context of health care decision-making. Issues relating to both sides of the decision rule – the ICER and the cost-effectiveness threshold(s) - were addressed in the chapters in this thesis, where some chapters focused on a health care perspective and some focused on a societal perspective. The impact of future health events on decision-making is relevant as scarcity of resources and the increasing demand for health care necessitate choices regarding which programmes and interventions to reimburse. This also requires choices between programs that yield costs and benefits at different moments in the future, directly or indirectly.

This thesis started with an investigation into the standardization of including so-called future unrelated medical costs and future non-medical costs in economic evaluation and in the accompanying cost-effectiveness threshold, specifically the *k*-threshold. The second half of this thesis covered methods that may be used when eliciting the value of safety and estimated the value of QALY gains, potentially including this value of safety, in the context of an early warning system for infectious disease outbreaks. This chapter will first answer the research questions posed in the introduction and subsequently discusses the limitations of the presented research and highlights some relevant implications for policy and future research.

Findings of this thesis

Estimating and standardizing the inclusion of future costs in economic evaluation

The inclusion of future costs, both medical and non-medical, in economic evaluation is necessary to ensure an optimal allocation of resources across interventions that extend life and those that improve quality of life (future costs refer to costs, medical and non-medical, that occur during the life-years gained from an intervention). As highlighted in this thesis, without the inclusion of future costs, life-extending interventions incorrectly appear relatively cost-effective relative to interventions that only increase quality-of-life (and therefore do not affect survival). The latter are, by definition,

unaffected by the inclusion or exclusion of future costs. If the future costs generated in life-years gained are ignored, then comparing the two types of interventions is biased, leading to sub-optimal allocation of resources and, subsequently, lower overall health and/or welfare. Chapters 2, 3 and 4 looked into the inclusion of future costs in economic evaluations of health care interventions. Chapters 2 and 3 focused on how future costs can be estimated for use in economic evaluations and how their inclusion can be standardized. Our results presented in these two chapters showed that it is both possible and impactful to estimate future costs for and include these in economic evaluation of life-extending interventions.

Chapter 2 provided updated estimates for future unrelated medical costs in The Netherlands, stratified by age, gender, disease category and time-to-death (TTD).[1] The initial estimates were provided by van Baal et al. (2011).[2] First estimates were also provided for future non-medical consumption, specified by age and gender. These are relevant to the Dutch context where the societal perspective is used in economic evaluation, and therefore costs outside the health care sector are considered in the evaluation. Alongside presenting methods for the inclusion of future unrelated medical costs in economic evaluation, guidance on how to adjust estimates for 'non-average' patient groups (i.e. patient groups with health care consumption that systematically differs from an average person in the general population) was also provided.

Chapter 3 provided the first standardized framework for the inclusion of future unrelated medical costs for England and Wales. The methods used in this chapter are similar to those in Chapter 2, providing estimates based on age, gender and TTD. As disease-specific cost-of-illness data was not available, this chapter gave guidance on how to adjust future unrelated medical cost estimates for prevalence of the disease treated by an intervention and any additional co-morbidities that might affect future unrelated medical costs. Previous research by Briggs et al. (2018) also provided future unrelated medical cost estimates, although only for costs unrelated to ten specified disease categories. Our estimates, which were estimated using a different methodological approach, follow a similar pattern to this earlier work, yet are consistently larger. This is most likely due to the selection of diseases by Briggs et al., and our estimates being adjusted for 2018 prices (while their estimates are in 2014 prices).

The estimates and methods from Chapters 2 and 3 were also used to create open-access online tools to aid researchers wishing to add future costs to their economic evaluations. Given the respective national guidelines prescribing a societal and a health care perspective, the Dutch tool provides guidance for the inclusion of future unrelated medical costs and future non-medical consumption, and the tool for England and Wales provides guidance for the inclusion of future unrelated medical costs only.[3]

Chapter 4 shifted focus to the right-hand side of the decision rule under a health care perspective and updated previous estimates of the supply-side threshold for England and Wales by including future medical costs. The supply-side or *k*-threshold represents the cost per QALY of current interventions funded through changes in health care expenditure at the margin. This provides us with the average cost-effectiveness of unspecified interventions (potentially) displaced or expanded at the margin, which can also be seen as reflecting the health opportunity costs of spending on a new intervention. Our approach was similar to that used previously in the Netherlands to estimate a supply-side threshold estimate for cardiovascular disease (which included future medical costs),[3] although we updated initial supply-side estimates[4] with future costs rather than calculating an entirely new threshold estimate. If future medical costs are to be included in the ICER, then logic (and the literature) suggests that they should also be included in the supply-side threshold. This would lead to resources being used more optimally and thus generating more health.

What impact does including future costs have on whether interventions are deemed cost-effective?

After estimating future costs and including them on both sides of the decision-rule, the next step was to ask what the impact would be of including these costs on estimated ICERs and thresholds, and (therefore) on the final judgement regarding whether an intervention is considered cost-effective.

To give an indication of the impact of including future costs in economic evaluation, both Chapters 2 and 3 applied their estimates of future costs to existing economic evaluations. The interventions covered were an intervention

[3] The tools can be found here: https://www.imta.nl/paid/

for stage III colon cancer (Chapter 2),[5] two separate chemotherapy treatments for non-small-cell lung cancer (Chapter 2 & 3),[6,7] chemotherapy for acute myeloid leukaemia (Chapter 3),[8] and transcatheter aortic valve implantation (TAVI) (Chapter 3).[9] The interventions covered varied by treatment age and by whether they mostly improved quality-of-life or extended life – this was measured by a ratio of change in life expectancy over change in QALYs. Alongside the cases where future costs were included in specific interventions, both chapters looked at the impact of including future costs on the ICER when death is prevented (at no cost) at a certain age. This showed the average impact of including future costs at different treatment ages. Regarding the impact of including future costs of specific interventions; Chapter 2 found increases in the ICER of between €3,800 and €5,600 per QALY when including future unrelated medical costs, and between €5,000 and €9,100 per QALY when only including non-medical costs, both in the Dutch context. Chapter 3 presented similar findings and found increases in the ICER of between £2,600 and £5,400 per QALY when including only future unrelated medical costs (given the health care perspective).

The main reason for the differences between the estimates of increases in the ICER is, unsurprisingly, the difference in interventions themselves. Both studies found that the impact of including future costs in economic evaluation is dependent on the average age at the start of the intervention, and the ratio of QALYs gained to LYs gained (future unrelated medical costs only). The greater the average age at the start of the intervention, the higher the estimates of future medical costs, as medical costs are higher in the last years of life. If these more expensive years are closer to the current year, the impact of discounting on the associated costs will be smaller and therefore their impact on the ICER larger.

With regard to the second finding, the smaller the ratio of QALYs to LYs gained, implying that the intervention is relatively more life extending than quality of life improving, the larger the impact of future costs on the ICER. Additionally, Chapter 3 found that adjusting for double-counting of future related costs (i.e. removing related costs from estimates) had relatively little impact on the ICER, while adjusting for TTD had a relatively large impact. The estimates presented in both Chapters 2 and 3 were lower than previous estimates of future unrelated medical costs. For example, research into the cost-effectiveness of

cancer screening in the U.S. suggested that ICERs are underestimated by between $10,300 and $13,700.[10] These differences may be explained by differences in applied discount rates and the fact that health care expenditures in the U.S. are usually higher than those in the UK or the Netherlands.

Discount rates themselves also affect the differences between our estimates – a rate of 3.5% is used for both costs and benefits for England and Wales,[11] while the Netherlands use a rate of 4% to discount costs and a rate of 1.5% to discount benefits.[12] Therefore, even if these two regions were to have the same spending pattern by age, the Netherlands would have lower future medical cost estimates, which leads to a smaller ICER. Furthermore, as benefits are discounted at a lower rate in the Netherlands, incremental benefits are higher than they would be for a treatment evaluated in England and Wales, leading, once again, to a smaller ICER. Our results, however, show that changes in the ICER due to the inclusion of future unrelated medical costs are larger in the Netherlands than in England and Wales. This is most likely explained by the notable differences in the age profile of health spending between the Netherlands and England and Wales, which appear to mostly be due to how health care costs are calculated. First, while Dutch cost of illness data used in Chapter 2 are estimated using a top-down approach (i.e. they are disaggregated from total health care costs),[13] data for England and Wales, which are used in Chapters 3 and 4, are estimated using a bottom-up approach.[14] These differences in approach may lead to differences in included cost-categories. Second, the two countries have a somewhat different definition of healthcare with respect to long-term care. In England and Wales, (almost all) long-term care is not covered by the National Health Service (and is often financed privately), and so does not fall under the banner of the healthcare system, whereas in the Netherlands long-term care is publicly financed and considered part of the healthcare system. The magnitude of long-term care spending also differs between these two countries – the Netherlands is known for spending the most on long-term care of all OECD countries at 3.7% of GDP, while the UK spends 1.4% of GDP, which is below the OECD average.[15]

A key message from these two chapters is that including future costs can lead to a shift in the order of which interventions are viewed as most cost-effective; mitigating existing bias towards life-extending interventions and leading to

decisions that better align with the commonly assumed aim of health care systems, i.e. leading to more health and welfare.

Chapter 4 added a valuable piece to the puzzle, in that it showed the impact of including future medical costs on estimates of the *k*-threshold in the UK, based on an approach used in the Netherlands. A previous estimate of the supply-side threshold[4] was updated to include future medical costs using the life-tables method and the future medical cost estimates from Chapter 3. We found that the threshold estimate increases from approximately £13,000 to £13,700 when future medical costs are included. This relatively limited increase appears to be partly due to the fact that in the paper calculating the initial estimate, marginal returns on health spending were predominantly driven by improvements in quality of life, and not by gains in life-years. This estimate (both the initial estimate and our updated one) is significantly lower than the Dutch supply-side threshold estimate for cardiovascular disease mentioned above, which was €41,000, however the impact of including future unrelated medical costs in this threshold estimate was also relatively small.[16] Key reasons for this small impact are that CVD related costs (which are substantial) were already included in the estimate, and that most hospital spending is centred in the last year of life (and costs in the last year of life decrease at higher ages), as the TTD effect is much stronger in the Netherlands than it in in England and Wales.

A key finding from this chapter is that the increase in the threshold from including future medical costs is substantially less than increases in ICERs estimated in previous chapters. While this difference between ICER estimates and the threshold is likely related to the selection of interventions used as case-studies in Chapters 2 and 3, the selected interventions are arguably representative of the technologies frequently appraised by decision-making bodies. For example, of the 55 technology appraisals published by NICE in 2020, 35 covered cancer treatments, five of which treated blood cancers and nine of which treated lung cancers.[19]

This combination of findings therefore implies that if both ICERs and threshold estimates were updated to include future costs, interventions which are quality of life-extending and/or for younger patients, would be more likely

to be reimbursed than they are now relative to interventions that are quality-of-life improving and/or for older patients.

Which methods can be used to value feelings of health safety?

We were not only interested in how future health events impact costs, but also in how they are a part of the benefits that can occur in the present and the future. Chapters 5 and 6 explored this issue further in the context of prevention and infectious diseases.

As mentioned in the introduction, there are many intangible benefits, such as feelings of safety, to preventive programmes such as an early warning system for infectious diseases. These benefits occur both in the present and the future, even if an early warning system will only (potentially) lead to *health* benefits in the (far) future. Chapter 5 presented a literature review of empirical literature where safety (also referred to as risk- or uncertainty-reduction) was valued. Papers were excluded if safety valuation was not the main aim of the paper and all elicitation techniques (such as stated or revealed preference) were included. The chapter showed that all retrieved papers used stated preference techniques, in which individuals are asked, directly or indirectly, what they would pay for a hypothetical increase in safety. From the literature reviewed, the most common method was contingent valuation – an approach which asks people to directly report their (hypothetical) willingness-to-pay to obtain a good or willingness-to-accept to give up a good.[20]

This key finding from Chapter 5 influenced a data collection from six European countries, in which respondents were asked their willingness-to-pay (WTP) for an early warning system, along with more specific questions which focused on WTP for such a system when explicating the beneficiaries and exact health gains generated by it. These data were analysed in Chapter 6, where the value of an early warning system was estimated. More specifically, Chapter 6 looked into seven separate scenarios, which differed by the reduction in risk of infection and the duration of the health gains generated by the early warning system. Scenarios also varied by whether the perspective was socially-inclusive-personal (SIP) or socially exclusive (SE). Eliciting SIP preferences

means that the beneficiaries of the intervention could entail the respondents as well as others, while in eliciting SE preferences, the beneficiaries only concern others and not the respondent him- or herself.

How do people value health gains from programmes related to the prevention of disease outbreaks?

Preventive programmes, like an early warning system to prevent infectious disease outbreaks, typically require current investments while generating outcomes which are uncertain and may occur in the (far) future (such as reduced mortality due to infectious diseases). To inform how much a government might invest in such an early warning system, we estimated how much individuals are willing to pay for it. Our particular approach is only one potential source of information to inform governmental investment decisions and comes with its own set of limitations that will be discussed in the following section.

Chapter 6 found that individuals were willing to pay between €19 and €26 in taxes a month for an early warning system to prevent infectious disease outbreaks, suggesting that most individuals are willing to increase their taxes so as to have more certainty and perhaps better (own or others) future health. Besides the value of the expected health gains, the elicited value potentially included other benefits brought about by an early warning system, such as feelings of safety. Feelings of safety would be present whether or not an early warning system ever detected an outbreak, as it still functions as a form of health protection.

Chapter 6 also estimated an 'augmented' v-threshold in the context of preventing pandemics via an early warning system. This threshold is the estimate of WTP per QALY gained through an early warning system for each of the seven scenarios presented, and may include the value of feelings of safety next to that of QALY gains, as mentioned above. The first estimate for WTP per QALY (when probability-weighted) was approximately €37,000 per QALY, which sits within the cost-effectiveness range used in the Netherlands, (€20,000 to €80,000).[21] However, WTP per QALY estimates varied considerably, especially depending on the size of the expected health gain and

whether the question was posed using the SIP or SE perspective. Variation resulted from changes in WTP per scenario not being proportional to the changes in health gains projected in those scenarios.

These results suggest that while the contingent valuation method, as suggested by the research presented in Chapter 5, can provide estimates that are in line with the current literature, threshold estimates can vary considerably, especially depending on the size of the health gain in question. Even though Chapter 6 provided several possible 'augmented' demand-side thresholds for programmes that prevent infectious disease outbreaks, the variation in estimates does lead to the question of which estimates might be considered to be most accurate in terms of being closest to 'actual preferences'. This remains an issue for future research, also in relation to how WTP values are elicited.

Additionally, there may be cases where a preventive programme is not expected to increase feelings of safety and therefore should not be measured against any of the threshold estimates provided. Using these threshold estimates may then be inappropriate, as they could represent both WTP for a QALY *and* feelings of safety. Furthermore, the questionnaire design itself may have affected (the quality of) our estimates. For example, using online 'self-administered' questionnaires is known to produce different results than when an interviewer asks the questions, owing to, for instance, reduced engagement or additional cognitive strain placed on respondents.[22] Moreover, to increase comparability across programs, one would ideally separate the value of a QALY from the value of safety, which was neither the aim of our study nor was it possible.

While Chapter 6 provided a set of answers to initial questions regarding how people value health gains in the context of preventing infectious diseases, several questions remain concerning the use of contingent valuation, the (divergent) results and questionnaire design.

Limitations

With this thesis I have tried to show how future health events potentially impact all elements of the health care decision-making framework, specifically costs, benefits and the two types of cost-effectiveness thresholds (k and v). In doing so, the proposed research questions have been answered, albeit with

limitations. In this section some noteworthy limitations will be discussed.

First, when making any type of prediction, assumptions need to be made. In the case of Chapters 2, 3 and 4, these assumptions were predominantly about age, gender and disease-specific spending patterns. Some of the data used (for example the household spending survey) were relatively old and may not be fully reflective of current spending patterns. For example, as the retirement age increases and subsequently so do formal years of employment / productivity, so might levels of both medical and non-medical consumption.[23,24] Furthermore, it may be that certain patient groups do not follow average spending patterns, for a variety of reasons. For example, there are patients with kidney disease who are put on dialysis permanently, and thus have far higher background health care costs than the average individual. We would then see a larger impact of including future unrelated medical costs, and therefore a higher ICER. This also emphasises that inclusion of these (like other) costs can lead to important distributional issues.

Second, the first half of this thesis focused solely on data from the Netherlands, England and Wales. We can be fairly confident in our estimates as they reflect average per capita expenditure over a range of diseases and goods and services, which is more straightforward to predict than expenditures for a particular disease. While survival estimates can be a source of uncertainty, this uncertainty is mitigated when presenting a relative measure such as the ICER because survival influences estimates of both future cost estimates (the numerator) and the QALY (the denominator). Nonetheless, spending patterns and decedent-survivor ratios may still be sources of uncertainty[25,26] The spending-patterns by age and decedent cost patterns used as inputs in the chapters are likely to change over time. For example, consider advances in cancer treatments; as treatments become more and more effective at improving quality-of-life and extending life, (relative) spending not only shifts to higher ages, but also into the last year of life. Additionally, the literature has found that the more a disease impacts longevity, the higher the additional health care costs would be after the elimination of the disease.[27]

Furthermore, to estimate any impact of future costs on economic evaluation in a country, estimates of future costs by age and gender (and preferably disease)

are necessary. Given that the corresponding data varies quite substantially between countries, we cannot generalise the specific estimates from Chapters 2 to 4 to other countries. Conversely, the *methods* used in these chapters can be used across countries, and therefore are generalizable. The same is likely to hold for the patterns observed in our results in these chapters. For example, the fact that future costs differed across ages and between survivors and decedents, that future costs need to be included on both sides of the decision-making framework, and that interventions for the elderly and those that prolong life will incur relatively large increases in their ICERs when future costs are included, are expected to be equally relevant for other countries and jurisdictions.

Third, the data collected for Chapters 5 and 6, which focused on how to value an early warning system for infectious diseases, were collected before the COVID-19 pandemic, i.e. when the idea of a pandemic was far more hypothetical than it is today. That being said, a recent replication of the questionnaire used in Chapter 6, repeated during the spring 2020, found only moderate increases in WTP due to COVID-19 burden.[28] Whether this reflects a similar unresponsiveness to the circumstances under which WTP is elicited (as observed in Chapter 6) or signals that the initially elicited WTP was indeed accurate, cannot be assessed. Variation in WTP between respondents was larger in the repeated study, with the most notable change in WTP determinants occurring in self-employed individuals (who were arguably affected strongly by the pandemic). Thus, while our results from these studies – Chapter 6 in particular – must be considered with caution, they do indicate a general willingness to pay for an early warning system for infectious disease outbreaks. The exact height, especially in relation to the benefits of such a system, remains less clear and deserves further attention.

Finally, related to the previous point, while the use of contingent valuation methods is commonplace when estimating the value of an intangible good, the question still stands as to whether these methods sufficiently reflect 'actual' preferences of individuals. First, while our WTP for a QALY estimates may in fact be augmented with individuals' WTP for safety, safety was not explicitly mentioned in the questionnaire nor in any answers to open-ended questions. Future researchers may wish to add questions surrounding feelings

of safety (perhaps after the main body of the questionnaire) to ascertain how much it featured in the valuation process. Ideally, this could lead to a valuation of safety separately from the value of a QALY. Second, our WTP for a QALY estimates varied dependent on the size and duration of the health gain, whether a scenario involves a SIP or SE perspective, and whether the early warning system completely removes any risk of losses to health. While not inherent to contingent valuation, it is likely that these values are affected by issues surrounding questionnaire design, such as sensitivity to scale and anchoring, given the relatively similar estimates provided for WTP across scenarios. This emphasises that caution is needed in the use of any of these estimates, especially in translating them to WTP per QALY, since respondents may not make similar calculations when answering WTP questions.

Implications for policy and future research

The results presented in this thesis are relevant for policy-making, and can be used to improve how future events are considered in economic evaluations and subsequent health care decision-making. Furthermore, we can infer from the limitations mentioned above that more research needs to be done. In this section I will outline some of the most relevant policy implications and areas for future research.

First, this thesis has shown how it is both necessary and of consequence to include future costs on both sides of the decision-making framework. Not only does including future costs change the hierarchy of which intervention is most cost-effective, but the impact on the ICER is far greater than the impact on the supply-side threshold. As mentioned in previous chapters, the v-threshold is not expected to change, as people's WTP for a QALY is elicited independently from the economic evaluation and most likely already includes some (relevant) benefits of future costs. Depending on the perspective taken, I recommend that decision-makers include future unrelated medical costs and, if using a societal perspective, also future non-medical consumption in economic evaluations. This would involve the Dutch health authority (Zorginstituut) explicitly recommending the inclusion of future non-medical costs in economic evaluation; they currently only mention the inclusion of non-medical 'costs incurred in sectors outside the healthcare system' with

no explicit mention of life-years gained.[12] In a similar vein, the National Institute for Health and Care Excellence (NICE), who provide guidance for economic evaluation across England and Wales, would need to change their guidelines to prescribe the inclusion of future unrelated medical costs – which they currently actively discourage. If both countries were to follow these recommendations, they could benefit from using the online tools developed from Chapters 2 and 3, PAID 3.0 and PAID UK, that provide future cost estimates and a standardized approach to their inclusion. In order to keep up with changing guidelines and spending and mortality trends, such tools should be regularly updated and improved where necessary. Including future costs in economic evaluation will inevitably lead to interventions for older treatment groups being deemed relatively less cost-effective. As argued by van Baal et al., distributional issues such as this should be dealt with 'openly and explicitly'.[29] Economic evaluation is only one element in the decision-making process, and is by no means the sole indicator for the reimbursement of health care interventions. By including future costs in economic evaluations decision makers are completely informed about all relevant costs and benefits of an intervention, which they may subsequently weight against other goals, such as ethical and distributional considerations.

Second, after including future costs, the supply-side threshold estimate for the UK increased from approximately £13,000 to £13,700, a relatively limited increase when compared to the changes in the ICERs observed in Chapters 2 to 4. This estimate is still quite a bit lower than the £20,000 to £30,000 range of values provided by NICE. This implies that the current threshold (range) used by NICE is set too high. It is clear that having the most precise estimate of the supply-side threshold leads to optimal health and welfare and that any deviation from this estimate adversely effects health. This can occur directly, by accepting interventions that do not maximise health (if the threshold is too high) or by rejecting health maximizing interventions (if the threshold is too low), or indirectly, by incentivising pharmaceutical companies to set levels of research and development that are wastefully high (if the threshold is too high) or by research and development investment levels being too low relative to their economic value (if the threshold is too low).[30]

Given that it is difficult to ascertain exactly which interventions are displaced

or expanded at the margin, using empirical estimates based on the average cost-effectiveness of interventions (as opposed to being based on a few historically published cost-effectiveness studies) is arguably an appropriate approximation. For all the reasons mentioned in this chapter and previous chapters, future costs should also be included in this threshold estimate, although the ultimate impact on the threshold itself is relatively limited.

Third, initial research into the cost-effectiveness of the response to the COVID-19 pandemic have suggested estimates between €160,000 per QALY and €1 mil1ion per QALY. [32] However, the assumptions made when calculating these different estimates differ substantially; for example, the comparator used in these calculations is often a hypothetical 'no response' from the government. It is, however, difficult to predict how individuals would have responded to the pandemic without government intervention. It seems unlikely that people would go about their daily activities 'as normal', and there would almost certainly still be job losses and absence from work also due to (more) illness, thereby affecting incremental costs between the intervention and comparator. Furthermore, while these initial estimates of cost-effectiveness are presented in cost per QALY format, the COVID response did not only impact health, but also the economy, and most likely other wellbeing outcomes, such as feelings of safety or solidarity and social cohesion. That being said, while there is still much uncertainty surrounding the current cost-effectiveness estimates for the COVID response, which is a *curative* response, they all appear to be much higher than our 'augmented' *v*-threshold estimates, which possibly include feelings of safety in the benefits, for a *preventive* early warning system. If further research were to confirm the magnitude of cost-effectiveness of current interventions as well as the estimates observed in Chapter 6, and if both are deemed to reflect real valuations of health gains, we could infer that individuals and governments value preventive interventions less than they value curative interventions.[33]

The estimates presented in this thesis add to the knowledge base informing national and international policy-makers on how much should be spent to mitigate the economic and medical casualties that are caused by outbreaks such as COVID-19. While both estimates of cost-effectiveness of the COVID-19 response and our estimates of an investment threshold for a system to

prevent infectious disease outbreaks are uncertain and need further study and confirmation, they for now suggest that the (European) response to COVID-19 may either not have been cost-effective or not all elements of value were sufficiently or accurately captured in the existing economic evaluations. It would be prudent for public health departments to work on cost-effective responses (and preventive systems) for future outbreaks, as they are an ever-present threat.

Hopefully, the steps made in this thesis towards intentionally considering future health in the decision-making framework will be built upon by other researchers in health economics and surrounding fields. Here I will give some possible avenues for future research.

First, while chapters 3 and 4 provided initial insights into the effect of intervention treatment age, the QALY gains to LY gains ratio and TTD, implementing the inclusion of future costs in forthcoming economic evaluations will further our understanding of these interactions. It would also be beneficial to estimate future costs by age, gender and disease for every country that uses economic evaluation in the decision-making framework, as health care costs can vary quite dramatically depending on the sectors covered by the health care sector, the price of medicines and so on. Furthermore, comparing the impact of including future costs in multiple countries (once future cost estimates are available) would offer further insight into the differences in data sources and their effect on the size of increases in the ICER.

Second, it would be interesting to find individual-level future medical and non-medical costs for specific patient groups. This is a massive undertaking as it would require a lengthy longitudinal study collecting health spending and consumption data for several cohorts. However, it would provide a cohort perspective on how (future) cost spending patterns change over a lifetime, which, if compared to population averages, would inform which interventions might require 'additional 'adjustment to future cost estimates.

Third, while there is much ongoing research into the estimation of both the supply-side threshold and the demand-side threshold, there is currently little research on how these thresholds might be simultaneously incorporated into the decision-making framework. By comparing these thresholds, we can see

whether the marginal cost-effectiveness of certain sectors or disease areas are aligned with societal preferences and valuations of the related gains. Policy analysis into the simultaneous use of these two thresholds could help to shed light on opinions and hurdles surrounding this issue. Regarding the two thresholds separately, I also have some recommendations for future research. Considering *k,* more research could be done into what exactly *is* displaced or expanded at the margin. Currently, most interventions do not go through the economic evaluation process and thus it is difficult to garner more precise estimates of health forgone. Furthermore, it would be interesting to better understand the differences in *k* estimated between health sectors and between disease areas and what such differences imply. Regarding *v*, it would be interesting to investigate whether individuals' do in fact consider their future health, costs and consumption when answering contingent valuation questions. This could be tested via a think-out loud study, or an additional questionnaire post-valuation. In a similar vein to the above suggestion, we could also confront respondents with the implied value of a QALY that their WTP estimates suggest, and ask them whether they agree. This would provide additional insight into the validity of WTP as a way to elicit *v*-threshold estimates.

Fourth, Chapter 5 reviewed the literature on safety valuation and the methods used to do so. Safety, however, is only one of many elements of value that could arise, if not a main aim, as positive externalities of preventive interventions; there are several outcome measures that may capture more of such elements, which could be used alongside or as replacement of the QALY in economic evaluations (such as the ICECAP,[34,35] ASCOT,[36] and WOOP[37] measures). It would be interesting to see further research into *v*-thresholds for such measures, (capturing broader well-being or direct valuations of elements of value not captured in the QALY), as initial research suggests these different *v*-threshold estimates are significantly different from those for QALY gains.[38] This would provide further insight into how individuals value such outcomes and into the relative importance of these outcomes compared to the QALY. This is especially relevant in a time where threats to global health, impacting societies in profound ways, have been shown to be more than a distant possibility.

Concluding statement

To conclude, this thesis looked at how we can take future health events into account on the ICER side and the threshold side of the decision rule, and from both the health care perspective and the societal perspective. I hope that this thesis has highlighted the importance of comprehensively considering the future in cost-effectiveness research and that the results and insights, highlighted above, will have a positive influence on future health care policy and research.

References

1. Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: A red herring? *Health Econ*. 1999;8(6):485-496.
2. van Baal P, Wong A, Slobbe LCJ, Polder JJ, Brouwer W, De Wit GA. Standardizing the inclusion of indirect medical costs in economic evaluations. *Pharmacoeconomics*. 2011;29(3):175-187.
3. van Baal P, Perry-Duxbury M, Bakx P, Versteegh M, van Doorslaer E, Brouwer W. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Econ*. 2019;28(1):87-100.
4. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the national institute for health and care excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14):1-503.
5. van Gils CW, de Groot S, Redekop WK, Koopman M, Punt CJ, Uyl-de Groot CA. Real-world cost-effectiveness of oxaliplatin in stage III colon cancer: A synthesis of clinical trial and daily practice evidence. *Pharmacoeconomics*. 2013;31(8):703-718.
6. Mok TS, Wu Y, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): A randomised, open-label, controlled, phase 3 trial. *The Lancet*. 2019;393(10183):1819-1830.
7. Bertranou E, Bodnar C, Dansk V, Greystoke A, Large S, Dyer M. Cost-effectiveness of osimertinib in the UK for advanced EGFR-T790M non-small cell lung cancer. *Journal of medical economics*. 2018;21(2):113-121.
8. Tremblay G, Dolph M, Patel S, Brandt P, Forsythe A. Cost-effectiveness analysis for midostaurin versus standard of care in acute myeloid leukemia in the United Kingdom. *Cost Effectiveness and Resource Allocation*. 2018;16(1):33-33.
9. Watt M, Mealing S, Eaton J, et al. Cost-effectiveness of transcatheter aortic valve replacement in patients ineligible for conventional aortic valve replacement. *Heart*. 2012;98(5):370-376.
10. Ratushnyak S, Hoogendoorn M, van Baal P. Cost-effectiveness of cancer screening: Health and costs in life years gained. *Am J Prev Med*. 2019;57(6):792-799.
11. NICE UK. Guide to the methods of technology appraisal. *NICE*; 2013.
12. Zorginstituut Nederland. *Guideline for Economic Evaluations in Healthcare*. Diemen, The Netherlands: Zorginstituut Nederland; 2016.
13. Rijksinstituut voor Volksgezondheid en Milieu (RIVM). Volksgezondheidenzorg.info. https://www.volksgezondheidenzorg.info/. Accessed August, 2020.
14. Asaria M. Health care costs in the English NHS: Reference tables for average annual NHS spend by age, sex and deprivation group. 2017:1-19.
15. OECD 2020. Health Statistics. https://stats.oecd.org/Index.aspx?ThemeTreeId=9. Accessed December 21, 2020.
16. van Baal P, Morton A, Meltzer D, Brouwer W. Future unrelated medical costs need to be considered in cost effectiveness analysis. *European Journal of Health Economics*. 2018:1-5.

17. Meerding W, Polder J, de Hollander A, Mackenbach J. Hoe gezond zijn de zorguitgaven?. zorg voor euro's-6. de kosten en opbrengsten van gezondheidszorg bij infectieziekten, kankers en hart-en vaatziekten. *RIVM rapport 270091002*. 2008.
18. Woods B, Revill P, Sculpher M, Claxton K. Country-level cost-effectiveness thresholds: Initial estimates and the need for further research. *Value in Health*. 2016;19(8):929-935.
19. NICE UK. Guidance and advice list. Updated January, 2021. Accessed January, 2021. https://www.nice.org.uk/guidance/published?type=ta&fromdate=January%202020&todate=January%202021&title=myoc.
20. Mitchell RC, Carson RT. *Using surveys to value public goods: The contingent valuation method*. Resources for the Future; 1989.
21. Raad voor de Volksgezondheid en Zorg. *Rechtvaardige en duurzame zorg*. Raad voor de Volksgezondheid en Zorg*;* 2006.
22. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *Journal of public health*. 2005;27(3):281-291.
23. Van Solinge H. Health change in retirement: A longitudinal study among older workers in the netherlands. *Res Aging*. 2007;29(3):225-256.
24. Aguila E, Attanasio O, Meghir C. Changes in consumption at retirement: Evidence from panel data. *Rev Econ Stat*. 2011;93(3):1094-1099.
25. Wong A, van Baal P., Boshuizen HC, Polder JJ. Exploring the influence of proximity to death on disease-specific hospital expenditures: A carpaccio of red herrings. *Health Econ*. 2011;20(4):379-400.
26. Howdon D, Rice N. Health care expenditures, age, proximity to death and morbidity: Implications for an ageing population. *J Health Econ*. 2018;57:60-74.
27. Grootjans-van Kampen I, Engelfriet PM, van Baal P,H.M. Disease prevention: Saving lives or reducing health care costs? *PLoS One*. 2014;9(8):e104469-e104469.
28. Himmler S, Van Exel J, Brouwer W. Did the COVID-19 pandemic change the willingness to pay for an early warning system for infectious diseases in europe?. *Submitted*.
29. van Baal P, Meltzer D, Brouwer W. Pharmacoeconomic guidelines should prescribe inclusion of indirect medical costs! A response to grima et al. *Pharmacoeconomics*. 2013;31(5):369-373.
30. Vernon JA, Goldberg R, Golec J. Economic evaluation and cost-effectiveness thresholds. *Pharmacoeconomics*. 2009;27(10):797-806.
31. Claxton K, Sculpher M, Palmer S, Culyer AJ. Causes for concern: Is NICE failing to uphold its responsibilities to all NHS patients? *Health Econ*. 2015;24(1):1-7.
32. Frijters P. Vanuit geluk bezien zijn de kosten van een lockdown veel hoger dan de baten. *ESB*. 12 November, 2020.
33. Dranove D. *What's your life worth?: Health care rationing--who lives? who dies? who decides?* FT Press; 2003.

34. Coast J, Peters TJ, Natarajan L, Sproston K, Flynn T. An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Quality of Life Research*. 2008;17(7):967-976.
35. Flynn T, Chan P, Coast J, Peters TJ. Assessing quality of life among British older people using the ICEPOP CAPability (ICECAP-O) measure. *Applied health economics and health policy*. 2011;9(5):317-329.
36. Milte CM, Walker R, Luszcz MA, Lancsar E, Kaambwa B, Ratcliffe J. How important is health status in defining quality of life for older people? an exploratory study of the views of older south Australians. *Applied health economics and health policy*. 2014;12(1):73-84.
37. Hackert MQN. *Care for well-being: Development of the well-being of older people measure (WOOP)*. [PhD thesis]. Rotterdam, The Netherlands: Erasmus University Rotterdam; 2020.
38. Himmler S, van Exel J, Brouwer W. Estimating the monetary value of health and capability well-being applying the well-being valuation approach. *The European Journal of Health Economics*. 2020;21(8):1235-1244.

Population health and health care spending are increasing rapidly across the globe. Consequently, the optimal allocation of resources towards and within the health care sector continues to be important. Here, it is important to recognise that our current actions regarding the allocation of resources may have consequences that stretch far into the future. The aim of this thesis was to investigate how several future health events, in terms of costs and benefits, can be accounted for in the context of health care decision-making, from both a health care perspective and a societal perspective. Issues relating to both sides of the decision rule – the incremental cost-effectiveness ratio (ICER) and the cost-effectiveness threshold(s) - were addressed in the chapters in this thesis, where some chapters focused on a health care perspective and some focused on a societal perspective. The thesis started with an investigation into the standardization of including so-called future unrelated medical costs and future non-medical costs in economic evaluation and in the accompanying cost-effectiveness threshold. The second part of this thesis covered methods that may be used when eliciting the value of safety and estimated the value of quality-adjusted life year (QALY) gains, potentially including this value of safety, in the context of an early warning system for infectious disease outbreaks. **Chapters 2, 3 and 4** looked into the inclusion of future costs in economic evaluations of health care interventions. **Chapters 2 and 3** focused on how future costs can be estimated for use in economic evaluations and how their inclusion can be standardized.

Chapter 2 provided updated estimates for future unrelated medical costs in The Netherlands, stratified by age, gender, disease category and time-to-death (TTD). First estimates were also provided for future non-medical consumption, specified by age and gender, which are relevant for the Dutch context as the societal perspective is taken. **Chapter 3** provided the first standardized framework for the inclusion of future unrelated medical costs for England and Wales. The methods used in this chapter are similar to those in Chapter 2, providing estimates based on age, gender and TTD. The results presented in these two chapters showed that it is both possible and impactful to estimate future costs for and include these in economic evaluation of life-extending interventions. Both chapters also provided guidance on how to adjust estimates for 'non-average' patient groups. A key message from these two chapters is that including future costs can lead to a shift in the order of

which interventions are viewed as most cost-effective; mitigating existing bias towards life-extending interventions and leading to decisions that better align with the commonly assumed aim of health care systems, i.e. leading to more health and welfare. **Chapter 4** shifted focus to the right-hand side of the decision rule under a health care perspective and updated previous estimates of the supply-side threshold for England and Wales by including future medical costs. This provided the average cost-effectiveness of unspecified interventions (potentially) displaced or expanded at the margin, which can also be seen as reflecting the health opportunity costs of spending on a new intervention. If future medical costs are to be included in the ICER, then logic (and the literature) suggests that they should also be included in the supply-side threshold. This would lead to resources being used more optimally and thus generating more health. A key finding from this chapter is that the increase in the threshold from including future medical costs is substantially less than increases in ICERs estimated in previous chapters.

We were not only interested in how future health events impact costs, but also in how they are a part of the benefits that can occur in the present and the future. **Chapters 5 and 6** explored this issue further in the context of prevention and infectious diseases. **Chapter 5** presented a literature review of empirical literature where safety was valued. When searching, papers using any elicitation technique were included and papers were excluded if safety valuation was not the main aim of the paper. The chapter showed that all retrieved papers used stated preference techniques, in which individuals are asked, directly or indirectly, what they would pay for a hypothetical increase in safety. From the literature reviewed, the most common method was contingent valuation – an approach which asks people to directly report their (hypothetical) willingness-to-pay to obtain a good or willingness-to-accept to give up a good. This key finding from **Chapter 5** influenced a data collection from six European countries, in which respondents were asked their willingness-to-pay (WTP) for an early warning system, along with more specific questions which focused on WTP for such a system when explicating the beneficiaries and exact health gains generated by it. These data were analysed in **Chapter 6**, where the value of an early warning system was estimated. More specifically, **Chapter 6** looked into seven separate scenarios, which differed

by the reduction in risk of infection, the duration of the health gains generated by the early warning system and whether or not the perspective was socially-inclusive-personal (SIP) or socially exclusive (SE). Eliciting SIP preferences means that the beneficiaries of the intervention could entail the respondents as well as others, while in eliciting SE preferences, the beneficiaries only concern others and not the respondent him- or herself. **Chapter 6** found that most individuals were willing to increase their taxes so as to have more certainty and perhaps better (own or others) future health, specifically via an early warning system to prevent infectious disease outbreaks. **Chapter 6** also estimated a threshold in the context of preventing pandemics via an early warning system. This threshold was the estimate of WTP per QALY gained through an early warning system for each of the seven scenarios presented, and may include the value of feelings of safety next to that of QALY gains. However, WTP per QALY estimates varied considerably, especially depending on the size of the expected health gain and whether the question posed included the respondent's health or not. Variation resulted from changes in WTP per scenario not being proportional to the changes in health gains projected in those scenarios. These results suggest that while the contingent valuation method can provide estimates that are in line with the current literature, threshold estimates can vary considerably, especially depending on the size of the health gain in question. Even though **Chapter 6** provided several possible 'augmented' demand-side thresholds for programmes that prevent infectious disease outbreaks, the variation in estimates does lead to the question of which estimates might be considered to be most accurate in terms of being closest to 'actual preferences'.

This thesis looked at how we can take future health events into account in the framework of cost effectiveness, from both the health care perspective and the societal perspective. In doing so it aimed to add to the health economic literature and provide insight into how future health events can be accounted for in the health care decision-making process. Overall, I believe that this thesis has provided additional stimulus and context for the study of future health costs and outcomes within health economics, and I hope that this area of research will continue to develop given its relevance in the current (and future) economic climate.

Samenvatting

Uitgaven voor volksgezondheid en gezondheidszorg nemen wereldwijd toe Zodoende blijft de optimale verdeling van middelen binnen de zorgsector van belang, alsmede de afweging hoeveel budget te reserveren voor zorg ten opzichte van andere sectoren. Het is belangrijk om te erkennen dat onze huidige keuzes met betrekking tot de toewijzing van middelen gevolgen kunnen hebben die tot ver in de toekomst reiken. Het doel van dit proefschrift was om te onderzoeken hoe verschillende toekomstige gezondheidsgebeurtenissen, in termen van kosten en baten, in overweging kunnen worden genomen bij vergoedingsbeslissingen over behandelingen en interventies vanuit gezondheidszorgperspectief en vanuit maatschappelijk perspectief. De invloed van toekomstige gezondheidsgebeurtenissen op beide kanten van de besluitvormingskader - de incrementele kosteneffectiviteitsratio (ICER) en de kosteneffectiviteitsdrempel(s) - kwamen aan de orde in de hoofdstukken van dit proefschrift. Hierbij concentreerde een aantal hoofdstukken zich op een gezondheidszorgperspectief en een aantal op een maatschappelijk perspectief. Het proefschrift begon met een onderzoek naar de standaardisatie van het opnemen van zogenaamde toekomstige niet-gerelateerde medische kosten en toekomstige niet-medische kosten in economische evaluatie en in de bijbehorende kosteneffectiviteitsdrempel. De tweede deel van dit proefschrift behandelde methoden die kunnen worden gebruikt bij het uitvragen van de waarde van veiligheid en schatte de waarden van kwaliteit gecorrigeerde levensjaren (‘*quality-adjusted life years*’, QALYs)-winsten, mogelijk inclusief deze waarde van veiligheid, in de context van een vroegtijdig waarschuwingssysteem voor de uitbraken van infectieziekten. In **Hoofdstukken 2, 3 en 4** is gekeken naar het opnemen van toekomstige kosten in economische evaluaties van zorginterventies. **Hoofstukken 2 en 3** richtte zich op hoe toekomstige kunnen worden geschat voor gebruikt in economische evaluaties en hoe hun inclusie kan worden gestandaardiseerd.

Hoofstuk 2 rapporteert geactualiseerde schattingen voor toekomstige niet-gerelateerde medische kosten in Nederland, gestratificeerd naar leeftijd, geslacht, ziektecategorie en tijd tot overlijden. Ook zijn er eerste schattingen gemaakt voor toekomstige niet-medische consumptie, gespecifieerd naar leeftijd en geschat. Deze schattingen kunnen relevant zijn voor Nederlandse

context wanneer bij economische evaluaties een maatschappelijke perspectief wordt ingenomen. **Hoofdstuk 3** bood een erste gestandaardiseerde kader voor het opnemen van toekomstige niet-gerelateerde medische kosten voor Engeland en Wales. De methoden die in dit hoofdstuk worden gebruikt, zijn vergelijkbaar met die in **Hoofdstuk 2** en geven schattingen op basis van leeftijd, geslacht en tijd tot overlijden. Deze twee hoofdstukken lieten zien dat het zowel mogelijk als impactvol is om toekomstige kosten in te schatten en deze mee te nemen in economische evaluaties van levensverlengende interventies. Een kernboodschap uit deze twee hoofdstukken is dat het meenemen van toekomstige kosten kan leiden tot een verschuiving in de volgorde waarin interventies als meest kosteneffectief worden beschouwd; bestaande vooringenomenheid voor levensverlengende interventies verminderen en leiden tot beslissingen die beter aansluiten bij het algemeen aanvaarde doel van gezondheidszorgstelsels, d.w.z. leiden tot meer gezondheid en welvaart. **Hoofdstuk 4** verschoof de aandacht naar de rechterkant van de beslissingsregel door toekomstige medische kosten op te nemen vanuit een gezondheidszorgperspectief en actualiseerde daarmee eerdere schattingen van de grenswaarde, vanuit de aanbodszijde, voor Engeland en Wales. Met dit proces werd gemiddelde marginale kosteneffectiviteit van niet-gespecificeerde interventies die (potentieel) werden verplaatst of uitgebreid berekend. Dit kan ook worden gezien als een weerspiegeling van de gezondheidsopportuniteitskosten van uitgaven voor een nieuwe interventie. Als toekomstige medische kosten in de ICER moeten worden opgenomen, lijkt het logisch (en ook naar de stand van deliteratuur) dat deze ook in de aanboddrempel moeten worden opgenomen. Dit zou ertoe leiden dat middelen optimaal worden benut en dus meer gezondheid genereren. Een belangrijke bevinding uit dit hoofdstuk is dat de verhoging van de grenswaarde als gevolg van het opnemen van toekomstige medische kosten aanzienlijk lager is dan de verhogingen van ICER's die in eerdere hoofdstukken werden geschat.

We waren niet alleen geïnteresseerd in hoe toekomstige gezondheidsgebeurtenissen de kosten beïnvloeden, maar ook in hoe ze deel uitmaken van de voordelen die in het heden en in de toekomst kunnen optreden. In **hoofdstukken 5 en 6** werd deze kwestie verder onderzocht in de context van preventie en infectieziekten. **Hoofdstuk 5** presenteerde

een literatuurreview van de empirische literatuur waarin veiligheid was gewaardeerd. Tijdens vergaring waren publicaties die alle meettechnieken gebruikten opgenomen en publicaties waar veiligheidswaardering niet het hoofddoel was waren uitgesloten. Het hoofdstuk toonde aan dat alle geïdentificeerde publicaties uitgesproken voorkeurstechnieken gebruikten, waarbij individuen, direct of indirect, gevraagd werden wat ze zouden betalen voor een hypothetische verhoging van de veiligheid. Uit de onderzochte literatuur was de meest gebruikelijke methode contingente-waardering; een benadering die mensen vraagt om hun (hypothetische) betalingsbereidheid voor een goed of hun bereidheid een goed op te geven direct te rapporteren. Deze belangrijke bevinding uit **Hoofdstuk 5** leidde tot een dataverzameling in zes Europese landen, waarin respondenten werd gevraagd naar hun bereidheid om te betalen (‘*willingness to pay’*,WTP) voor een vroegtijdig waarschuwingssysteem, samen met meer specifieke vragen die gericht waren op WTP voor een dergelijk systeem waarbij meer detail werd gegeven over de begunstigden en de exacte gezondheidswinst die erdoor wordt gegenereerd. Deze data werd geanalyseerd in **Hoofdstuk 6**, waarin de waarde van een systeem voor een vroegtijdig waarschuwingssysteem werd geschat. **Hoofdstuk 6** onderzocht zeven afzonderlijke scenario’s, die verschilden in de mate van het infectierisico, de duur van de gezondheidswinst gegenereerd door het vroegtijdig waarschuwingssysteem en alsmede of het perspectief sociaal-inclusief-persoonlijk (SIP) of sociaal exclusief (SE) was. Het opwekken van SIP-voorkeuren betekent dat de begunstigden van de interventie zowel de geënquêteerde als anderen kunnen zijn, terwijl bij het uitlokken van SE-voorkeuren de begunstigden alleen anderen betreffen en niet de geënquêteerde zelf. **Hoofdstuk 6** ontdekte dat de meeste mensen bereid waren om hun belastingen te verhogen om via een vroegtijdig waarschuwingssysteem om het uitbreken van infectieziekten te voorkomen meer zekerheid te hebben en mogelijk een betere (eigen of anderen) gezondheid in de toekomst. In **Hoofdstuk 6** werd ook een drempel in de context van het voorkomen van pandemieën via een vroegtijdig waarschuwingssysteem. Deze drempel was de schatting van de WTP per QALY die werd verkregen door middel van een vroegtijdig waarschuwingssysteem voor elk van de zeven gepresenteerde scenario’s, en omvat mogelijk de waarde van veiligheidsgevoelens naast die van QALY winsten. Echter varieerden de WTP per QALY-schattingen aanzienlijk,

wat vooral afhankelijk was van de omvang van de verwachte gezondheidswinst en of de gestelde vraag over de gezondheid de geënquêteerde omvatte of niet. Variatie was het gevolg van het feit dat veranderingen in WTP per scenario niet evenredig waren met de verwachte veranderingen in gezondheidswinst in die scenario's. Deze resultaten suggereren dat hoewel de contingente-waarderingsmethode schattingen kan opleveren die in overeenstemming zijn met de huidige literatuur, drempelschattingen aanzienlijk kunnen variëren, vooral afhankelijk van de omvang van de gezondheidswinst in kwestie. Hoewel **Hoofdstuk 6** verschillende mogelijke 'verhoogde' drempels aan de vraagzijde heeft gegeven voor programma's die uitbraken van infectieziekten voorkomen, leidt de variatie in schattingen tot de vraag welke schattingen 'werkelijke voorkeuren' het beste weergeven.

Dit proefschrift keek naar hoe we met toekomstige gezondheidsgebeurtenissen rekening kunnen houden in het kader van kosteneffectiviteit, zowel vanuit het perspectief van de gezondheidszorg als vanuit het maatschappelijk perspectief. Hiermee beoogde het een aanvulling te zijn op de gezondheidseconomische literatuur en inzicht te bieden in hoe toekomstige gezondheidsgebeurtenissen in overweging kunnen worden genomen in het besluitvormingsproces in de gezondheidszorg. Over het algemeen ben ik van mening dat dit proefschrift een extra stimulans en context heeft geboden voor de studie van toekomstige gezondheidskosten en -resultaten binnen de gezondheidseconomie, en ik hoop dat dit onderzoeksgebied zich zal blijven ontwikkelen, gezien de relevantie ervan in het huidige (en toekomstige) economische klimaat.

Portfolio

Training

2020	Dutch B2.2-C1.1. *Erasmus Language & Training Centre, Rotterdam.*
	Science Communication in Theory & Practice (summer school). *University of Copenhagen*
2019	Analytical Storytelling. *Analytical Storytelling Group*
2018	Active Blended Learning. *Risbo teaching institute, Rotterdam.*
	Drupal & Content Management. *Lectric, Rotterdam.*
	Group Dynamics. *Risbo teaching institute, Rotterdam.*
	Participating in Health Technology Assessment Research. *Erasmus School of Health Policy & Management, ESHPM.*
	Self-presentation. *Erasmus Graduate School of Social Sciences and the Humanities, Rotterdam.*
	Summer Academy on Hospital Economics & Mental Health Economics. *Competent in Competition in Health (CINCH) Essen*
	Your personal PhD work-life balance: how to do less but achieve more. *Erasmus Graduate School of Social Sciences and the Humanities, Rotterdam.*
2017	Applied Microeconometrics. *Tinbergen Institute, Rotterdam.*
	Basic Didactics. *Risbo teaching institute, Rotterdam.*

History of Economic Thought. *Tinbergen Institute, Rotterdam.*

How to Survive Your PhD, *Erasmus Graduate School of Social Sciences and the Humanities, Rotterdam.*

Making an Academic Poster that Stands Out. *Erasmus Graduate School of Social Sciences and the Humanities, Rotterdam.*

Searching, Finding and Managing Your Literature. *Erasmus Graduate School of Social Sciences and the Humanities, Rotterdam.*

Teaching

2017-2020	Working groups in Economics of Health and Health Care, required course MSc Health Policy, Management & Law (Erasmus School of Health Policy & Management)
2017-2020	Master thesis supervisor for the MSc programme Health Economics, Policy & Law (Erasmus School of Health Policy & Management)
2017-2019	Lectures in Public Health Economics, required course MSc Health Policy, Management & Law – specialization Health Economics (Erasmus School of Health Policy & Management)
2017-2019	Working groups in Advanced Research Methods, required course MSc Health Policy, Management & Law (Erasmus School of Health Policy & Management)
2016-2018	Working groups in Global Health Economics, elective course MSc Health Policy, Management & Law (Erasmus School of Health Policy & Management)

Awards

2017 Dean's Award for Multidisciplinary Excellence, Erasmus University Rotterdam

Presentations

Invited Seminars

2020 Perry-Duxbury M., Asaria M, Lomas J & van Baal P. Cured Today, Ill Tomorrow: A Method for Including Future Unrelated Medical Costs in Economic Evaluation in England and Wales. Centre for Health Economics, York. January 2020.

2018 Cash on the Nail: did an increase in hospital spending reduce Dutch mortality? CINCH Summer School, Essen.

Conference and summer school presentations

2020 Lowlands Health Economics Study Group. Groningen (online edition), Netherlands. Discussant & Paper Presentation.

2019 European Health Economics Association PhD-Supervisor & ECR Conference. Porto, Portugal. Paper Presentation.

International Health Economics Association. Basel, Switzerland. Paper Presentation.

Lowlands Health Economics Study Group. Almen, Netherlands. Paper Presentation.

2018 Lowlands Health Economics Study Group. Hoenderloo, Netherlands. Paper Presentation.

European Health Economics Association.
Maastricht, Netherlands.
Paper Presentation.

2017 International Health Economics Association.
Boston, USA.
Paper Presentation.

COMPARE Young Researchers Annual Meeting.
Rotterdam, Netherlands.
Poster Presentation.

List of publications

In this dissertation

Kellerborg, K.*, Perry-Duxbury, M.*, de Vries, L.*, & van Baal, P. (2020). Practical Guidance for Including Future Costs in Economic Evaluations in The Netherlands: Introducing and Applying PAID 3.0. Value in Health, 23(11), 1453-1461. **Equal contribution*

Perry-Duxbury, M., Asaria, M., Lomas, J., & van Baal, P. (2020). Cured Today, Ill Tomorrow: A Method for Including Future Unrelated Medical Costs in Economic Evaluation in England and Wales. Value in Health, 23(8), 1027-1033.

Perry-Duxbury, M., van Exel, J., & Brouwer, W. (2019). How to value safety in economic evaluations in health care? A review of applications in different sectors. *The European Journal of Health Economics*, 20, 1041-1061.

Other academic publications

Himmler, S., van Exel, J., Perry-Duxbury, M., & Brouwer, W. (2020). Willingness to pay for an early warning system for infectious diseases. *The European Journal of Health Economics*, 21, 763–773.

Perry-Duxbury, M., van Exel, J., & Brouwer, W. , Sköldunger, A., Gonçalves-Pereira, M., Irving, K., Mayer, G., Selbaek, G., Woods, B., Zanetti, O., Verhey, F., Wimo, A., Handels, R., & The Actifcare Consortium. (2019). A Validation Study of the ICECAP-O in Informal Carers of People with Dementia from Eight European Countries. *Quality of Life Research,* 29(1), 237-251.

van Baal, P., Perry-Duxbury, M., Bakx, P., Versteegh, M., van Doorslaer, E., & Brouwer, W. (2019). A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health economics*, 28(1), 87-100.

Miscellaneous professional activities

Referee for *Quality of Life Research, Health Economics & Australian Health Review*. 2018-2019 Member of THRIVE Purpose Accelerator, 2017-2018 Communications officer for Rotterdam Global Health Initiative.

Acknowledgements

I feel incredibly fortunate to have had so many brilliant people around me during the four years in which I wrote my PhD. It has been hard to find the words to fully articulate my gratitude but given that that is the point of this section of the book, I will try my best.

I would first like to thank my supervisors, Pieter van Baal and Werner Brouwer:

Pieter, you hired me to write a dissertation on a topic that is quite far removed from the final product we ended up with. While many variables changed during my time at ESHPM, our working relationship remained constant, for which I am grateful. Thank you for your seemingly infinite number of research ideas, your sense of humour, and for not engaging me in too many conversations about football.

Werner, thank you for taking me on as one of your students halfway through my trajectory. I learned a lot from you in terms of both scientific content and professionalism – and of course my general knowledge has increased at least three-fold due to the various red herrings scattered throughout each of our meetings! I am grateful for your invaluable feedback and positive attitude.

Thanks also go to my first supervisor Eddy van Doorslaer. Eddy, while we did not work together all that long, I want to thank you for giving me the opportunity to work at ESHPM. You feedback was always insightful, and your ambition regarding the outcomes of our project was inspiring.

Next up are my co-authors: Job, thank you for not only contributing to this thesis but for hiring me as a research assistant for the months before I started my PhD. James and Miqdad, thank you for providing insight into the UK health care system, and providing invaluable feedback on our papers.

I have had the pleasure of teaching in several courses. Lucas, Pieter Bakx, and Bram, thank you for giving me the opportunity to teach the next wave of heath economists and to grow and learn in settings outside of my dissertation topic.

I would also like to thank my fellow colleagues from the Health Economics group, which I am sure will be one of the friendliest environments I ever work in. Starting with my office mates from over the years, i.e. the people who also

endured the yearly 'ceiling-gate' saga. Mariska and Kaya, thank you both for making me feel so welcome from the beginning of my time at ESHPM. Sebastian, your love of all things sugary (see: cake) and aversion to caffeine made you a trustworthy office mate for the short time you were in J8-45. Friederike, thank you for the nice office chats and tips about new places and events to check out in the city. Stefan, thank you for being a sounding board on numerous occasions and for exuding self-confidence that was at least a little bit contagious!

Next, colleagues with whom I did not share an office, but did share much of my working-day. Linda, you made 'co-first-authoring' a paper a lot of fun. Thanks for sharing the load and covering for me when I needed to leave a meeting to buy Eurovision tickets. Jannis, you added significantly to the lunch time debates about pretty much every series on Netflix. I will miss those heated arguments (see: Star Wars, Avengers). Vivian, thank you for the great advice you gave me on multiple occasions and for suggesting wild social activities like the roller disco. Judith, thank you for always chatting to me when I came and spent far too much time eating sweets and standing in your office avoiding my own work on a Friday afternoon. Klas, thank you for talking/debating with me about any topic under the sun and for starting our short-lived tradition of going to (quite bad) comedy shows in the city.

Now to move away from the university campus:

Gozien, Marieke, en Jannie – hartstikke bedankt voor jullie steun en liefde over de laatste zeven jaar. Ik heb het geluk dat ik zo'n leuke schoonfamilie heb.

Mum and Dad, thank you for all your love, care, and support. The past four years were a rollercoaster for all of us, and I am so grateful that you were there for me throughout it all. You have always accepted where I want to go or what I want to do, providing advice when needed but also standing back and letting me discover the world on my own.

Finally, Wilbert, thank you for always cheering me on, and for being the person I call 'home'. I can always count on you for some light pressure underneath the elbow when crossing the street.

Funding

The work in Chapter 3 was supported by grant 754936 from the Horizon 2020 research and innovation program, and was part of the COMPAR project under work package 6 (https://self-management.eu/compar-eu-project/work-packages/work-package-6/).

The work in Chapters 5 and 6 was part of the COMPARE project, a multidisciplinary research network that aims to become a platform for the rapid identification, containment, and mitigation of emerging infectious diseases and foodborne outbreaks (http://www.compare-europe.eu/). The COMPARE project was funded by the European Commission under the Horizon 2020 research and innovation programme (grant agreement No. 643476).

About the Author

Meg Perry-Duxbury was born in Oxford, United Kingdom on August 27th, 1994. She was accepted to Chetham's School of Music in 2005 as a piano student, and stayed there for 7 years, receiving her A levels and LRSM for Piano Performance in 2012.

Meg moved to the Netherlands and completed a BSc. in Economics and Business Economics at the University of Groningen in 2015 and an MSc. in Economics (specialisation Health Economics) at Erasmus University Rotterdam in 2016. In 2017 Meg started a PhD in the Health Economics department at Erasmus School of Health Policy & Management, under the supervision of Dr. Pieter van Baal and Prof. dr. Werner Brouwer.

Meg has taught in several courses for the MSc programme Health Economics, Policy & Law, and has also supervised six MSc students. She has presented her work at several international conferences and has published in journals such as Value in Health, the European Journal of Health Economics, and Quality of Life Research.