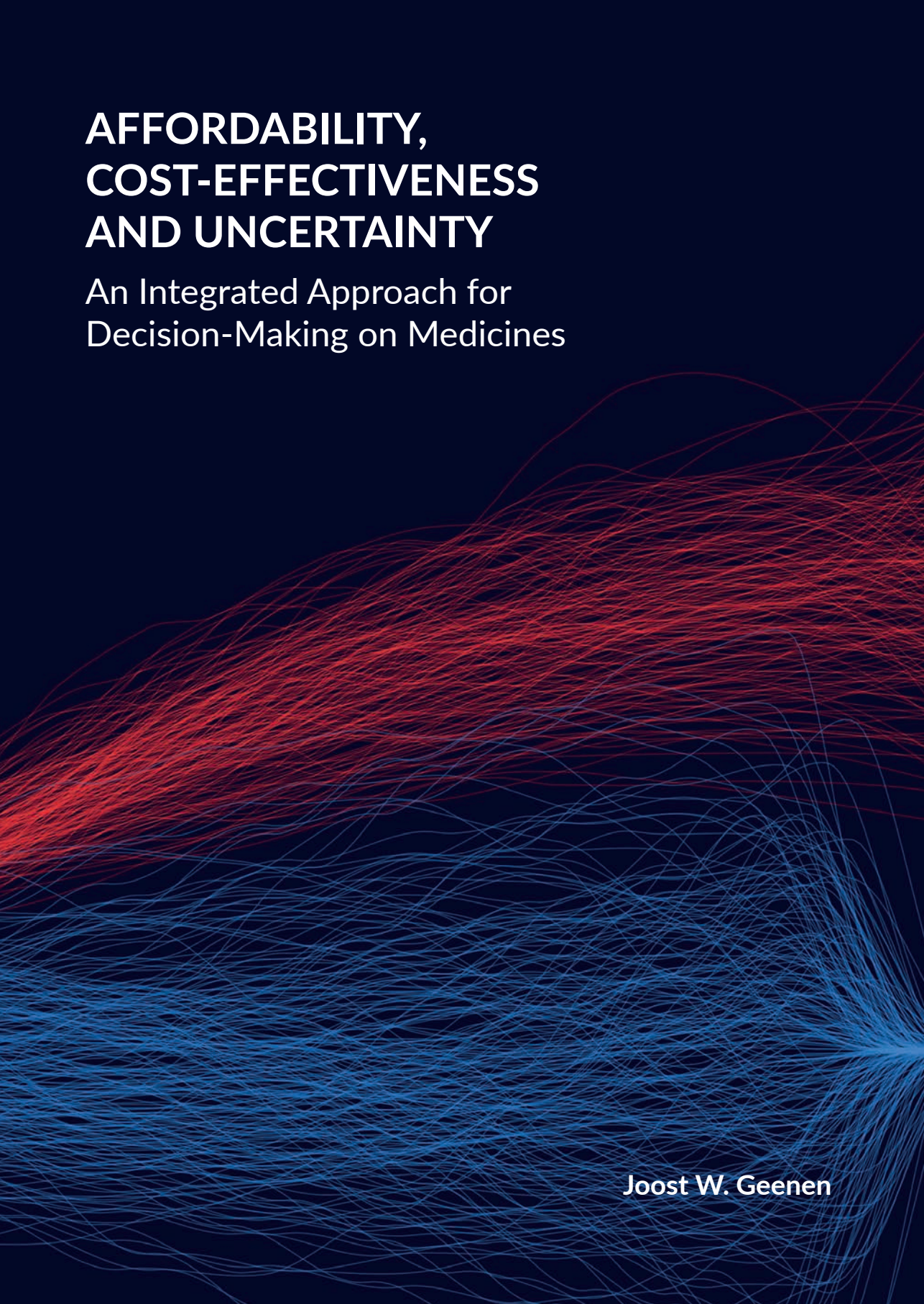
AFFORDABILITY, COST-EFFECTIVENESS AND UNCERTAINTY

An Integrated Approach for Decision-Making on Medicines



Joost W. Geenen

Affordability, Cost-Effectiveness and Uncertainty: An Integrated Approach for Decision-Making on Medicines

Joost Willem Geenen

Colophon

The research presented in this thesis was performed at the division of Pharmacoepidemiology and Clinical Pharmacology of the Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, the Netherlands.

Cover image: Joost W. Geenen, R 3.6.1 for Windows, ggplot2

Cover, layout & printing: Off Page, Amsterdam

ISBN: 978-94-6182-984-9

Financial support for printing of this thesis was kindly provided by the Koninklijke Nederlandse Maatschappij ter bevordering der Pharmacie (KNMP) and the Utrecht Institute for Pharmaceutical Sciences (UIPS). This PhD project was funded by an unrestricted grant from GlaxoSmithKline. The work presented in Chapter 5 was supported by a grant from the Dutch Health Care Institute (ZIN) as part of the Health Technology Assessment and Appraisal collaboration.

Affordability, Cost-Effectiveness and Uncertainty: An Integrated Approach for Decision-Making on Medicines

Betaalbaarheid, Kosteneffectiviteit en Onzekerheid: Een Geïntegreerde Aanpak voor Besluitvorming over Geneesmiddelen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 11 december 2019 des middags te 4.15 uur

door

Joost Willem Geenen
geboren op 5 februari 1990
te Tilburg

PROMOTOREN

Prof. dr. O.H. Klungel
Prof. dr. J.A.M. Raaijmakers†

COPROMOTOREN

Dr. A.M. Hövels
Dr. C. Boersma

TABLE OF CONTENTS

1

General introduction

HEALTH TECHNOLOGY ASSESSMENT

In the majority of nations in the world, promotion of health and wellbeing as well as consequent prevention of ailment or disease of citizens is a legal obligation to governments laid out in national constitutions [1–3]. Given the vast breadth of the concept of health and wellbeing, the task of safeguarding and advancing these now widely adopted fundamental human rights is by no means an easy one. Unsurprisingly, the resulting amount of resources that are globally devoted to this cause are immense [4,5].

Indeed, the world spent $7.5 trillion on healthcare in 2016 which translates into nearly 10% of global Gross Domestic Product (GDP) [5]. In the European Union (EU), 9.8% of the total GDP was devoted to healthcare in 2018 [4,6]. For the Netherlands, the situation is similar at a 9.9% share of the total GDP, equalling €76,9 billion in 2018 [7,8].

In light of these vast but not unlimited resources, choices regarding the investment in specific types, quantity and quality of care are inevitable. To assist and inform policy-makers on the consequences of (not) investing in specific types of care and healthcare related technologies, Health Technology Assessment (HTA) is conducted [9].

HTA is typically defined as the systematic evaluation of the properties, (wanted and unwanted) effects and, for example, budgetary, societal or organisational impacts of a health technology with the aim of informing policy decision-making [9,10]. Health technology is a very broad concept in this regard as it includes medicines, medical devices, procedures, diagnostics but also other clinical, public health and organisational interventions, as well as the efficient use of resources in healthcare [9,10].

Within the scope of new health technologies, medicines play a crucial role. Not only in their contribution to the advancement of health and society, but also by means of its profound impact on national healthcare budgets and ensuing public debate [11]. It is therefore that HTA has, in many jurisdictions, become an important advisory or even formal decision-making criterion in reimbursement or access decisions of new medicines [12,13].

UNCERTAINTY

The concept of uncertainty is inherent of life. This lack of certainty, aside from its more philosophical implications, is reflected in a plethora of real-world materialisations such as the existence of any conceivable type of insurance plan, the need for daily updated predictions of the weather and, for example, the absence of knowing a-priori which patient might benefit from a certain type of treatment.

In cost-effectiveness models, various types of uncertainty can be identified [14,15]. A major type is parameter uncertainty and relates to the precision and validity which an input parameter is estimated [16]. These input parameter estimates should reflect the true (unbiased) value of the population but are typically generated using only a sample of this population. Parameter uncertainty can therefore partly be regarded as a lack of information as a larger sample size would, in general, yield a more precise estimate [16]. Crucially, parameter uncertainty is different from variability, the inevitable difference between individuals and therefore a source of uncertainty, which is irreducible [14-17]. A second major type of uncertainty is structural or modelling uncertainty. It refers to a lack of validity of outcomes due to limitations that are inherent to a type of model, technique or model structure [14,15,18,19].

Given the potential influence of uncertainty on modelling outcomes, various techniques have been developed to quantify, reduce and manage various types of uncertainty [14,20]. Many of those techniques are now mandatory when cost-effectiveness models are used as part of an application for reimbursement of medicines [21–23].

DETERMINANTS OF DECISION MAKING AND THE ROLE OF UNCERTAINTY

Current reimbursement decision-making is driven by the clinical or therapeutic value of the medicine in question but also by economic outcomes as the Incremental Cost-Effectiveness Ratio (ICER) and Budget Impact (BI) [24–26]. The ICER, designed to quantify this clinical or therapeutic value, is generally compared to- or benchmarked against a Willingness to Pay (WTP) threshold. The role, value and legal status of WTP thresholds however differs greatly between jurisdictions [27–29].

As BI, ICER and WTP are core aspects of this thesis, the current status, role and methods for managing the uncertainty of these three major determinants will be elaborated further.

ICER

The models underlying ICER outcomes are prone to parameter and structural uncertainty. For quantifying parameter uncertainty, probabilistic sensitivity analysis (PSA) and the resulting Value of Information (VOI) analysis have been developed and are now widely used [30]. In a PSA, all input parameters are simultaneously varied along predefined ranges according to their probability distribution, with the outputs generally presented as a scatterplot in the cost-effectiveness plane [31]. It has been shown that decision makers are less likely to reimburse drugs with a highly uncertain ICER [24].

The traditional scatterplot is however limited in its capacity to display differences in the relative density of ICER samples [32]. This obfuscates the true distribution and could hide intricate

details of the underlying distributions or small areas with differing densities. Furthermore, due to overdrawing of samples in high density areas, the actual probability of individual PSA samples (especially outliers) within the total sample is often over-estimated [32]. As PSA is often used for supporting decision-making, the validity of this method is crucial but can currently be questioned.

WTP

The WTP, or ICER threshold, is not an outcome of a cost-effectiveness model and could therefore currently be regarded as a fixed or constant value. The WTP does however have different fixed values or fixed ranges in various jurisdictions [27,28]. In England, an upper limit of £20,000 - £30,000 per QALY is deemed cost-effective whilst an informal threshold varies between €20,000 - €80,000 in the Netherlands.

In a healthcare system with fixed budgets, new innovations can only be funded by savings or disinvesting in other care and thus cause displacement [25,33,34]. In healthcare systems with less restricted budgets, resources are still not unlimited so at least some opportunity costs will exist and policy-makers are likely to prefer lower BI over higher BI [24,25]. Research shows that a high BI and / or highly uncertain BI is a potential risk to decision makers and that they are then more likely to limit reimbursement or to issue a type of managed entry agreement (MEA) [35–37].

These displacement effects and opportunity costs should, at least to some degree, be reflected in the WTP threshold [25,33–35,38]. The WTP is typically used to reflect the maximum amount society is willing pay for one additional Quality Adjusted Life Year (QALY). WTP could however also be regarded as the marginal cost per additional QALY [33]. When assuming that decision-makers have a preference for displacing high ICER care before low ICER care and are able to implement this preference, the following could be deducted: A high BI would displace more care than a low BI would. So, when care is displaced from high to low ICER, the higher the BI, the lower the average ICER of the total care that is displaced. When WTP would include displacement, a higher BI would thus result in a lower WTP [33]. Currently used WTP ranges are established somewhat arbitrarily or, for example, based on a jurisdictions' gross domestic product or on disease severity but they do not include BI as a factor [21,28,39,40].

Based on the aforementioned displacement effects, the exact value of the WTP threshold has been a topic of scientific debate as for example Shiroiwa et al. and Claxton et al. and Lomas et al. have described [28,41,42]. WTP values of, for example, £12,936 (England) and €74,000 (the Netherlands) have been proposed to better reflect opportunity costs and could therefore be more suitable for decision-making [33,43]. A recent literature review reported an even wider range of WTP estimates and found a mean of €24,226 per QALY [39].

In the Netherlands, the proportional shortfall (PS) underlies the value of the WTP and is based on the normative standpoint that (investment in) treatment should be prioritised to those who lose the largest proportion of their remaining QALYs due an illness [40]. This is an equity-based paradigm where social values or social preferences drive budgetary prioritisation on (new) treatments. Other examples of equity-based policy tools are increased WTP thresholds for end of life- and oncology care in England and regulatory benefits for orphan drugs in Europe [40,44,45]. Inherent to these choices is a loss of efficiency, as additional investment on specific 'preferred' diseases or patients must be compensated by reduced spending in other potentially more cost-effective areas.

When the goal is to maximise health output given the available resources and irrespective of its impact on equity of care, efficient spending should be prioritised. Efficiency is achieved when the WTP threshold reflects the cost per QALY at the margin and when this WTP threshold is then strictly enforced. Claxton, Lomas, Adang, Sculpher and others have all argued that opportunity cost should be reflected in decision-making and therefore in the WTP threshold [33,42,46–48].

As mentioned, the Dutch system is primarily based on the concept of equity. The current restrictive policies based on affordability do however not fit within this concept as affordability does not influence PS. Instead, affordability reflects the existence of a balance between equity and efficiency in health-care systems: affordability concerns, which in the Netherlands influence reimbursement decisions, must indeed be caused by maintaining or striving for some level of efficiency and therefore imply that opportunity cost must play a role and must therefore influence the WTP. Given the crucial role of WTP in relation to the ICER and potentially to BI, this topic is of major interest to this thesis.

It is henceforth assumed that for any healthcare system, at least some efficiency is strived for so that for any healthcare system, opportunity costs and marginal benefits are relevant. From a decision-making standpoint, we think that the eventual WTP threshold (be it based on equity, efficiency or both) should be adhered to in practice. Also, for interpretability and simplicity, we henceforth assume a strict enforcement of WTP in relation to the ICER.

BI

Budget Impact Analysis (BIA) is required for reimbursement applications in many jurisdictions [25,35,49]. Whilst submitting a BIA is often mandatory, the role of Budget Impact (BI) outcomes in decision-making is less clear than, for example, the role of cost-effectiveness [24,26,50,51]. Although the role of BI is often informal, many recent cases, (e.g. new drug introductions for hepatitis C such as Sovaldi[1] (sofosbuvir)), have shown that BI can be a crucial and even a decisive factor in reimbursement decisions [24,36,38,49,52–54].

As BIA are generally constructed using point estimates of various uncertain parameters and time-horizons, uncertainty in BI estimations is inevitable [55,56]. Mainly due to limited data, quantification of BI uncertainty remains limited to scenario analyses [55]. Therefore, BI is typically presented (for one or more scenarios) as a point estimate accompanied by a minimum and maximum value. A highly uncertain BI could lead to less or deferred access by means of, for example, managed entry agreements (MEA) [35,57,58].

According to a review by Van de Vooren et al., many published BIAs still fail to reach an acceptable quality [59]. Many BIAs are short term (one year), quite subjective or based on expert opinion and determined by estimations of population size and eventual treatment regimen [60,61]. If the general methodological quality of BI analyses is low, one would expect the predictive accuracy of these analyses to also be low.

Broder et al., who evaluated BI forecasts of US drug launches between 1-Sep-2010 and 1-sep-2015, concluded that the average predicted BI was 5.5 times the observed BI [60,61]. Cha et al. concluded that 60% of the drug forecasts were off by more than 40% [60]. Keeping et al. recently reported that BI estimates used by Welsh payers that were specifically produced to inform access decisions were off by more than 40% in 80% of the cases [62]. We believe that these findings illustrate that the methodological quality as well as the predictive accuracy of current BIAs can be considered as low.

Not only are these estimations insufficient in providing adequate clarity on the costs of a new drug, they also fail to quantify the uncertainty that is associated with these predictions. In other words, the current point estimates or ranges given are not based on an underlying probability distribution and thus provide insufficient insight in the possible range of financial outcomes. Especially given the concerns regarding accuracy and methodological quality mentioned previously, insights into uncertainty surrounding BI estimates could prove to be a crucial step in increasing the use and validity of BIA.

[1.] As budget impact is estimated and used in a product specific context, medicines are designated by their product name when they are discussed in the context of budget impact.

AFFORDABILITY, COST-EFFECTIVENESS AND UNCERTAINTY IN DECISION-MAKING

Affordability and cost-effectiveness are typically appraised separately whilst they both inform on the same decision: does the new technology deliver a health gain? Various studies have attempted to integrate the appraisal of these two aspects for decision-making. Still, none of these efforts have specifically included the uncertainty in affordability (as BI) and cost-effectiveness (as ICER). The criticality of the integration of uncertainty will be shown by means of the following examples:

An imaginary €10 bet which has a 50% chance to yield €25 and 50% change to yield €0. The expected return of this bet is positive at €2.50. To most people, €10 would be a loss without detrimental impact to their lives, making this bet very acceptable.

This second imaginary bet is different as it has two stages: The first stage determines the amount one should bet, where there is a 50% chance to a required bet of €10 and a 50% chance to a required bet of €100.000. The second stage determines the outcome, with a 50% chance of a 2.5-fold multiplication of the initial bet and a 50% chance of losing the entire bet. Still, this bet retains the 25% profit margin of the first bet. Clearly though, the potential impact on one's life are vastly different in this second bet thereby presumably altering the willingness to invest to many people.

In the second example, the amount one must bet reflects BI and the associated chance is the uncertainty in BI. Similarly, the potential return (25% profit) is the ICER and the probability of both outcomes reflect the ICER uncertainty. The potential €100.000 loss is the opportunity cost and can also be seen as displacement (having to sell a car for example). The hypothetical 'willingness to bet' is a reflection of the WTP threshold. Many more examples of bets reflecting reimbursement decision and the synergy between affordability, cost-effectiveness and uncertainty can be devised.

As a final example, consider a 80% chance on 500% return and 20% chance on 0% return, combined with a 80% chance on a €100.000 investment and a 20% chance on a €10.000.000 investment. Clearly, this is a very profitable investment opportunity. If however the potential return (ICER) is appraised separately from the amount required to invest (BI), the risk (a 4% chance of losing €10.000.000) is inadequately considered.

In reality, potential outcomes are much less defined as the dichotomous examples presented, making management of risk much less straightforward. Still, the outlined principles remain valid: uncertainty in affordability has a synergistic relationship with uncertainty in cost-effectiveness. The only way therefore to properly integrate risk as well as potential value of an innovation is

an integrated appraisal of affordability, cost-effectiveness and their individual uncertainties. As WTP should be driven by opportunity costs and therefore BI, an innovation's value should be composed of an integration of the three core components including their uncertainty, being ICER, WTP and ICER.

TEMPORAL ASPECTS OF DECISION-MAKING

When access decisions have to be taken, data on outcomes is often immature and/or significant uncertainty in outcomes remains [63–66]. In light of new, potentially effective treatments, decision-makers have to balance rapid access to patients and thereby accepting higher uncertainty with postponing access and waiting for more mature data [23,51,63–68]. Formal schemes have been developed for coping with these scenarios regarding marketing authorisation and reimbursement decisions, herein collectively denoted as Managed Entry Agreements (MEA) and, within a reimbursement decision-making scope, refers to concepts such as coverage with evidence development, various risk-sharing schemes and conditional coverage [23,51,63–68].

Such MEAs are now frequently used to grant patients early access to promising treatments whilst, towards payers, they assist in managing uncertainty in BI and clinical and/or cost-effectiveness [23,51,63–68]. By design, these schemes specifically recognise time and timing as a factor and eventual access decisions are relatively dynamic. Also, active management of new innovations or label changes (e.g., changes in indication) do happen over time [69]. Naturally, these temporal events (e.g., changes in price, indication, population, coverage status) influence cost-effectiveness. The phenomenon of time is thus a major factor and should therefore be included in economic analyses [69,70].

Current CEA and BIA however employ the Net Present Value (NPV) paradigm; future benefits and costs are discounted towards a present-day value and the investment or reimbursement decision is to be taken now or never [71]. Flexibility can be added by means of scenario analyses or reperforming a CEA or BIA after some time, but in the essence of these analyses, delaying the decision is not an available option. Therefore, current CEA or BIA methodology does not fully incorporate the role of active management of healthcare related projects or the development of uncertainty over time. Clearly, this is a hiatus in current decision-analytic modelling and consequent reimbursement decision-making.

TEMPORAL ASPECTS OF DECISION-MAKING: EARLY HTA

Economic evaluations by HTA bodies and concomitant price negotiations can take up to a year, delaying patient access [72]. To reduce delays in patient access, it is possible to start early with the assessment of added value of a new therapy. This is a different approach than the aforementioned MEA, where the decision itself is taken earlier (or more dynamically)

whereas in early HTA, the economic evaluation itself is performed earlier. The process of early HTA can start during preclinical development but is more regularly executed during phase I or II clinical trials. One method is the use of early cost-utility analyses, which can clarify the relative impact of the parameters that drive cost-effectiveness. Early cost-utility analyses allow decision makers and manufacturers to streamline clinical development and reimbursement processes [73–75].

Also, during the development of novel diagnostics, the price of a diagnostic as well as the eventual influence on treatment pathways and health outcomes is often unknown. From a health system perspective it is therefore important to assess, at an early stage, the potential impact of a test in daily practice. When price and the sensitivity (true positive rate) and specificity (true negative rate) of the test are still unknown, the estimation of cost-effectiveness is done in a turn-around analysis: investigate the required parameter-values that would make the test a cost-effective diagnostic. By evaluating an intervention at an early stage, its value becomes clear early and this can inform either further research, an implementation trajectory or an exit strategy.

In both these examples, parameter uncertainty plays a vital role. As early HTA has become an essential tool in advancing patient access, it is crucial that the current decision-analytical frameworks are suitable for these types of early analyses and that parameter uncertainty is handled in an optimal manner.

AN INTEGRATED APPROACH FOR DECISION-MAKING ON MEDICINES

The previous paragraphs have outlined the three core components of reimbursement decision-making. The ICER is typically characterised well, including the associated uncertainty. BI on the other hand, is methodologically less mature and appears to be estimated inaccurately. The WTP thresholds used to appraise the ICER currently do not include opportunity costs whilst evidence states that this should be done when a healthcare system (at least partly) strives for efficient spending.

Cost-effectiveness, affordability and their uncertainty are now however appraised separately. As described above, this leads to incorrect valuation of potential risks and benefits. Furthermore, the aspect of timeliness of data availability and uncertainty is hardly integrated in current decision-making practice. These issues form the motivation for the studies presented in this thesis.

OBJECTIVES

The main objective of this thesis is to develop an integrated approach of cost-effectiveness, affordability and the associated uncertainty. Uncertainty in this regard also pertains to the temporal aspects of evidence and uncertainty.

To achieve this, the current magnitude of uncertainty and its management in reimbursement decision-making needs to be assessed. Furthermore, it is imperative to identify and to develop methods for improved quantification as well as improved management of various sources of uncertainty, thereby contributing to improved decision-making on access to innovative medicines. The incorporation of timing of decisions and the temporal development of uncertainty will be crucial aspects, as well as the role of a dynamic WTP within the integration of affordability and cost-effectiveness.

The following specific objectives will be addressed in this thesis:

- To assess the possibility of accurate establishment of parameter or outcome thresholds in early HTA and whether these can be used to inform reimbursement- or market access decision-making (Chapter 2).
- To assess the accuracy of Dutch BI estimations that are used for informing reimbursement decisions, to assess the resulting influence of this potential source of uncertainty on decision-making and to evaluate whether it is justified to use current BI estimations for informing decision-making (Chapter 3).
- To develop methodology to reduce BI uncertainty, incorporating timeliness in BI estimation and to improve quantification and visualisation of ICER uncertainty (Chapter 4).
- To develop conceptual frameworks that unify cost-effectiveness, affordability, willingness to pay and the associated uncertainty and where the aspect of time regarding data availability, uncertainty and decision-making is explicitly included (Chapter 5).

REFERENCES

1. Kinney ED, Clark BA. Provisions for Health and Health Care in the Constitutions of the Countries of the World. Cornell Int Law J. 2004;37:285–356.
2. WHO | Countries [Internet]. WHO. [cited 2019 Jun 5]. Available from: http://www.who.int/countries/en/
3. WHO. Constitution [Internet]. [cited 2019 Jun 5]. Available from: https://www.who.int/about/who-we-are/constitution
4. OECD/EU (2018). Health at a Glance: Europe 2018: State of Health in the EU Cycle [Internet]. Paris: OECD Publishing; Available from: https://doi.org/10.1787/health_glance_eur-2018-en
5. WHO | Public Spending on Health: A Closer Look at Global Trends [Internet]. WHO. [cited 2019 Jun 5]. Available from: http://www.who.int/health_financing/documents/health-expenditure-report-2018/en/
6. Eurostat - Tables, Graphs and Maps Interface (TGM) table [Internet]. [cited 2019 Jul 17]. Available from: https://ec.europa.eu/eurostat/tgm/refreshTableAction.do?tab=table&plugin=1&pcode =tec00001&language=en
7. Centraal Bureau voor de Statistiek. Zorguitgaven stijgen in 2018 met 3,1 procent [Internet]. Cent. Bur. Voor Stat. [cited 2019 Jul 17]. Available from: https://www.cbs.nl/nl-nl/nieuws/2019/25/zorguitgaven-stijgen-in-2018-met-3-1-procent
8. StatLine - Zorguitgaven; kerncijfers [Internet]. [cited 2019 Jul 17]. Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84047NED/table?dl=1D372
9. WHO | Health technology assessment [Internet]. WHO. [cited 2019 Jun 5]. Available from: http://www.who.int/health-technology-assessment/en/
10. About HTAi | Health Technology Assessment International (HTAi) [Internet]. HTAi. [cited 2019 Jun 5]. Available from: https://htai.org/about-htai/
11. OECD. Pharmaceutical Innovation and Access to Medicines [Internet]. 2018. Available from: https://www.oecd-ilibrary.org/content/publication/9789264307391-en
12. Garrido MV, Zentner A, Busse R. Health systems, health policy and health technology assessment. Health Technol Assess Health Policy-Mak Eur. 2008;53.
13. Oortwijn W, Mathijssen J, Banta D. The role of health technology assessment on pharmaceutical reimbursement in selected middle-income countries. Health Policy. 2010;95:174–84.
14. Briggs AH. Handling Uncertainty in Cost-Effectiveness Models: Pharmaco Economics. 2000;17:479–500.
15. Briggs AH, Claxton K, Sculpher MJ. Decision Modelling for Health Economic Evaluation. Oxford University Press; 2006.
16. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Lippincott Williams & Wilkins; 2008.
17. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis: Pharmaco Economics. 2008;26:781–98.
18. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. Value Health. 2009;12:739–49.
19. Jackson CH, Sharples LD, Thompson SG. Structural and parameter uncertainty in Bayesian cost-effectiveness models. J R Stat Soc Ser C Appl Stat. 2010;59:233–53.
20. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value Health. 2012;15:796–803.
21. Zorginstituut Nederland. Kosteneffectiviteit in de praktijk - Rapport [Internet]. 2015 [cited 2019 Apr 24]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk
22. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal [Internet]. 2013 [cited

2019 May 13]. Available from: https://www.nice.org.uk/process/pmg9/chapter/foreword

23. Pouwels XGLV, Grutters JPC, Bindels J, Ramaekers BLT, Joore MA. Uncertainty and Coverage With Evidence Development: Does Practice Meet Theory? Value Health [Internet]. 2019 [cited 2019 May 1]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S1098301518363113
24. Ghijben P, Gu Y, Lancsar E, Zavarsek S. Revealed and Stated Preferences of Decision Makers for Priority Setting in Health Technology Assessment: A Systematic Review. PharmacoEconomics. 2018;36:323–40.
25. Cohen JP, Stolk E, Niezen M. Role of Budget Impact in Drug Reimbursement Decisions. J Health Polit Policy Law. 2008;33:225–47.
26. Simoens S. Use of Economic Evaluation in Decision Making: Evidence and Recommendations for Improvement. Drugs. 2010;70:1917–26.
27. Carrera P, IJzerman MJ. Are current ICER thresholds outdated? Valuing medicines in the era of personalized healthcare. Expert Rev Pharmacoecon Outcomes Res. 2016;16:435–7.
28. Shiroiwa T, Sung Y-K, Fukuda T, Lang H-C, Bae S-C, Tsutani K. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ. 2010;19:422–37.
29. Woods B, Revill P, Sculpher M, Claxton K. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. Value Health. 2016;19:929–35.
30. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. Health Technol Assess [Internet]. 2009 [cited 2019 Apr 24];13. Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta13290/
31. Cleveland WS, McGill R. The Many Faces of a Scatterplot. J Am Stat Assoc. 1984;79:807–22.
32. Mayorga A, Gleicher M. Splatterplots: Overcoming Overdraw in Scatter Plots. IEEE Trans Vis Comput Graph. 2013;19:1526–38.
33. Adang E, Stadhouders N, Parsons C, Wammes J, Govaert P, Determann D, et al. Verdringingseffecten binnen het Nederlandse zorgstelsel [Internet]. 2018. Available from: https://www.iqhealthcare.nl/media/124601/eindrapport-verdringing-2018.pdf
34. Pearson SD. The ICER Value Framework: Integrating Cost Effectiveness and Affordability in the Assessment of Health Care Value. Value Health. 2018;21:258–65.
35. Niezen MGH, Bont A de, Busschbach JJV, Cohen JP, Stolk EA. Finding legitimacy for the role of budget impact in drug reimbursement decisions. Int J Technol Assess Health Care. 2009;25:49–55.
36. Barua S, Greenwald R, Grebely J, Dore GJ, Swan T, Taylor LE. Restrictions for Medicaid Reimbursement of Sofosbuvir for the Treatment of Hepatitis C Virus Infection in the United States. Ann Intern Med. 2015;163:215.
37. Niezen MGH, Stolk EA, Steenhoek A, Uyl-De Groot CA. Inequalities in oncology care: Economic consequences of high cost drugs. Eur J Cancer. 2006;42:2887–92.
38. Neumann PJ, Cohen JT, Weinstein MC. Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold. N Engl J Med. 2014;371:796–7.
39. Ryen L, Svensson M. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. Health Econ. 2015;24:1289–301.
40. Reckers-Droog VT, van Exel NJA, Brouwer WBF. Looking back and moving forward: On the application of proportional shortfall in healthcare priority setting in the Netherlands. Health Policy. 2018;122:621–9.
41. 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;24:1–7.
42. Lomas J, Claxton K, Martin S, Soares M. Resolving the "Cost-Effective but Unaffordable" Paradox: Estimating the Health Opportunity Costs of Nonmarginal Budget Impacts. Value Health. 2018;21:266–75.
43. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the NICE cost effectiveness threshold [Internet]. Centre for Health Economics, University of York;

2013 Jan. Report No.: 081cherp. Available from: https://econpapers.repec.org/paper/chyrespap/81cherp.htm

44. Drummond M, Towse A. Orphan drugs policies: a suitable case for treatment. Eur J Health Econ. 2014;15:335–40.

45. Aggarwal A, Fojo T, Chamberlain C, Davis C, Sullivan R. Do patient access schemes for high-cost cancer drugs deliver value to society?—lessons from the NHS Cancer Drugs Fund. Ann Oncol. 2017;28:1738–50.

46. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess Winch Engl. 2015;19:1–503, v–vi.

47. Sculpher M, Claxton K, Pearson SD. Developing a Value Framework: The Need to Reflect the Opportunity Costs of Funding Decisions. Value Health. 2017;20:234–9.

48. Coyle D, Cheung MC, Evans GA. Opportunity Cost of Funding Drugs for Rare Diseases: The Cost-Effectiveness of Eculizumab in Paroxysmal Nocturnal Hemoglobinuria. Med Decis Making. 2014;34:1016–29.

49. Ghabri S, Mauskopf J. The use of budget impact analysis in the economic evaluation of new medicines in Australia, England, France and the United States: relationship to cost-effectiveness analysis and methodological challenges. Eur J Health Econ. 2018;19:173–5.

50. Franken M, Nilsson F, Sandmann F, Boer A de, Koopmanschap M. Unravelling Drug Reimbursement Outcomes: A Comparative Study of the Role of Pharmacoeconomic Evidence in Dutch and Swedish Reimbursement Decision Making. PharmacoEconomics. 2013;31:781–97.

51. Vreman RA, Bouvy JC, Bloem LT, Hövels AM, Mantel-Teeuwisse AK, Leufkens HGM, et al. Weighing of Evidence by Health Technology Assessment Bodies: Retrospective Study of Reimbursement Recommendations for Conditionally Approved Drugs. Clin Pharmacol Ther. 2019;105:684–91.

52. Flume M, Bardou M, Capri S, Sola-Morales O, Cunningham D, Levin L-A, et al. Approaches to manage 'affordability' of high budget impact medicines in key EU countries. J Mark Access Health Policy. 2018;6:1478539.

53. American Society of Clinical Oncology Position Statement on Addressing the Affordability of Cancer Drugs. J Oncol Pract. 2017;JOP.2017.027359.

54. Mauskopf J, Chirila C, Masaquel C, Boye KS, Bowman L, Birt J, et al. RELATIONSHIP BETWEEN FINANCIAL IMPACT AND COVERAGE OF DRUGS IN AUSTRALIA. Int J Technol Assess Health Care. 2013;29:92–100.

55. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2014;17:5–14.

56. Mauskopf JA, Sullivan SD, Annemans L, Caro J, Mullins CD, Nuijten M, et al. Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices—Budget Impact Analysis. Value Health. 2007;10:336–47.

57. Pauwels K, Huys I, Vogler S, Casteels M, Simoens S. Managed Entry Agreements for Oncology Drugs: Lessons from the European Experience to Inform the Future. Front Pharmacol [Internet]. 2017 [cited 2019 Apr 4];8. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2017.00171/full

58. Dakin HA, Devlin NJ, Odeyemi IAO. "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. Health Policy Amst Neth. 2006;77:352–67.

59. Vooren K van de, Duranti S, Curto A, Garattini L. A Critical Systematic Review of Budget Impact Analyses on Drugs in the EU Countries. Appl Health Econ Health Policy. 2014;12:33–40.

60. Cha M, Rifai B, Sarraf P. Pharmaceutical forecasting: throwing darts? [Internet]. Nat. Rev. Drug Discov. 2013 [cited 2018 Mar 7]. Available from: https://www.nature.com/articles/nrd4127

61. Broder MS, Zambrano JM, Lee J, Marken RS. Systematic bias in predictions of new drugs' budget impact: analysis of a sample of recent US drug launches. Curr Med Res Opin. 2018;34:765–73.

62. Keeping S, Deslandes PN, Haines KE, Routledge PA. Estimated Versus Observed Expenditure Associated with Medicines Recommended by the All Wales Medicines Strategy Group. PharmacoEconomics - Open [Internet]. 2019 [cited 2019 Jan 22]; Available from: https://doi.org/10.1007/s41669-019-0116-5

63. Carlson JJ, Sullivan SD, Garrison LP, Neumann PJ, Veenstra DL. Linking payment to health outcomes: A taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. Health Policy. 2010;96:179–90.

64. Carlson JJ, Gries KS, Yeung K, Sullivan SD, Garrison LP. Current Status and Trends in Performance-Based Risk-Sharing Arrangements Between Healthcare Payers and Medical Product Manufacturers. Appl Health Econ Health Policy. 2014;12:231–8.

65. Makady A, van Veelen A, de Boer A, Hillege H, Klungel OH, Goettsch W. Implementing managed entry agreements in practice: The Dutch reality check. Health Policy Amst Neth. 2019;123:267–74.

66. Klemp M, Frønsdal KB, Facey K. What principles should govern the use of managed entry agreements? Int J Technol Assess Health Care. 2011;27:77–83.

67. van de Wetering EJ, van Exel J, Brouwer WBF. The Challenge of Conditional Reimbursement: Stopping Reimbursement Can Be More Difficult Than Not Starting in the First Place! Value Health. 2017;20:118–25.

68. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with Evidence Development, Only in Research, Risk Sharing, or Patient Access Scheme? A Framework for Coverage Decisions. Value Health. 2012;15:570–9.

69. Minnema LA, Giezen TJ, Gardarsdottir H, Egberts TCG, Leufkens HGM, Mantel-Teeuwisse AK. Post-marketing dosing changes in the label of biologicals. Br J Clin Pharmacol. 2019;85:715–21.

70. Favato G, Baio G, Capone A, Marcellusi A, Saverio Mennini F. A Novel Method to Value Real Options in Health Care: The Case of a Multicohort Human Papillomavirus Vaccination Strategy. Clin Ther. 2013;35:904–14.

71. Ross SA. Uses, Abuses, and Alternatives to the Net-Present-Value Rule. Financ Manag. 1995;24:96–102.

72. Akehurst RL, Abadie E, Renaudin N, Sarkozy F. Variation in Health Technology Assessment and Reimbursement Processes in Europe. Value Health. 2017;20:67–76.

73. Ijzerman MJ, Steuten LMG. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. Appl Health Econ Health Policy. 2011;9:331–47.

74. IJzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. PharmacoEconomics. 2017;35:727–40.

75. Maignen F, Osipenko L, Pinilla-Dominguez P, Crowe E. Regulatory watch: Outcomes of early health technology assessment dialogues in medicinal product development. Nat Rev Drug Discov. 2017;16:79.

2

Use of early HTA to inform market access decisions

2.1

Early HTA in pharmacogenomics: a case example in cardiovascular drugs

Joost W. Geenen, Ekaterina V. Baranova, Folkert W. Asselbergs, Anthonius de Boer, Rick A. Vreman, Colin N.A. Palmer, Anke H. Maitland-van der Zee, Anke M. Hövels & on behalf of the PREDICTION-ADR consortium.

Published in Pharmacogenomics.
https://doi.org/10.2217/pgs-2017-0063

ABSTRACT

Background

To assess the required characteristics (cost, sensitivity and specificity) of a pharmacogenomics test for being a cost-effective prevention of ACEi-induced angioedema. Furthermore, we assessed the influence of only testing high risk populations.

Methods

A decision tree was used.

Results

With a willingness-to-pay (WTP) threshold of €20,000 and €80,000 per QALY, a 100% sensitive and specific test may have a maximum cost of €1.30 and €1.95, respectively. When only genotyping high risk populations, the maximum test price would be €5.03 and €7.55, respectively.

Conclusions

This theoretical pharmacogenomic test is only cost-effective at high specificity, high sensitivity and a low price. Only testing high-risk populations yields more realistic maximum test prices for cost-effectiveness of the intervention.

INTRODUCTION

The use of pharmacogenomics is becoming more common in daily clinical practice. In many cases it improves patient outcomes by predicting the response to drugs or adverse events, allowing health care providers to adjust treatment accordingly [1]. Recent literature shows variation in the performance of pharmacogenomics: it varies from a large effect with a large increase in efficiency to a large increase of costs per patient without much benefit [2]. Technology in pharmacogenomics is advancing and the number of known single nucleotide polymorphisms (SNPs) impacting pharmacological treatment is rapidly increasing.

Since both the advancement of technology as well as an ageing population cause an increased pressure on health care budgets, cost-effectiveness of innovations is on the health care policy agenda of many countries. To determine the coverage of innovations from public funds, many countries use a threshold which indicates the maximum costs to be paid for the gain of one extra quality adjusted life year (QALY) by the new intervention. For the UK for example, the threshold is indicated at £30,000 per QALY. For the Netherlands, the discussion on the threshold is ongoing. The current thresholds range from €20,000 to €80,000 per QALY gained, based on disease burden [3].

The price of testing as well as the effect of genetic variation on treatment response or adverse events is often unknown. From a health system perspective, it is therefore important to assess, at an early stage, the impact of a test in daily practice. When price and the sensitivity (true positive rate) and specificity (true negative rate) of the test are still unknown, the estimation of cost-effectiveness is done in a turn-around analysis: investigate the required specifications that would make the test a cost-effective diagnostic. The threshold for costs per QALY is used as the basis of this evaluation. By evaluating an intervention at an early stage, its value becomes clear early and this can inform either further research, an implementation trajectory or an exit strategy.

In this study, we take a case example of an early HTA assessment of the prediction of angioedema caused by the use of Angiotensin Converting Enzyme Inhibitors (ACEi). ACEis are amongst the most frequently prescribed drugs and serve as an important treatment modality for several, highly prevalent cardiovascular indications [4–6]. They are generally well tolerated. Non-productive, persistent cough is the most common adverse drug reaction (ADR) and occurs in approximately 9% of ACEi users [7]. Besides this mild and well-known ADR, ACEis can cause the rare ACEi induced angioedema, a serious and frightening sudden swelling of the upper airways that can be fatal [5,6,8–10].

ACEi induced angioedema is characterised by a transient, localised swelling of the deep reticular dermis, subcutaneous or submucosal tissues of the head and neck region and occasionally the viscera [11]. It frequently affects the face, lips, tongue and upper airways and is usually

accompanied by symptoms such as a lump in the throat, hoarse voice and difficulties in swallowing and breathing [11]. Typically, ACEi induced angioedema develops over 4-6 hours and resolves within 1-2 days [11,12]. Rare lethal cases with severe airway obstruction have also been reported [6,7]. The factors predisposing to ACEi induced angioedema are not fully elucidated. Among clinical risk factors of ACEi induced AE are female sex, age over 65 years, African-American ethnicity, local trauma, smoking, history of drug rash, type 2 diabetes, seasonal allergies and ACEi induced cough. The mechanism of ACEi induced angioedema is thought to involve the accumulation of bradykinin, due to a dysregulation of its inactivation by ACE and alternative enzymes [13]. Genetic variants identified in the membrane metallo-endopeptidase gene (MME) and the X-prolyl aminopeptidase 2 gene (XPNPEP2), belonging to the bradykinin degradation pathway, could contribute to the development of AE in some of the patients [13]. However, the effect of genetic variation on the susceptibility to AE caused by ACEis is yet to be fully uncovered.

The identification of patients at risk of ACEI-induced angioedema using a pharmacogenomic (PG) test prior to treatment initiation could prevent harm caused by this ADR and reduce healthcare expenses.

Hence, the goal of this study is to assess required test characteristics (cost, sensitivity & specificity) in order for the test to be a cost-effective measure for preventing ACEi induced angioedema. In addition, we investigate the benefits of only testing specific populations that are known to have an increased risk of developing this serious ADR.

METHODS

We used a decision tree model to compare genotyping vs no genotyping prior to starting an ACEi. The model reflects the patient pathway and is depicted in figure 1. In constructing the model, we conformed to the ISPOR Modeling Good Research Practices [14]. As angioedema risk is greatest immediately after starting an ACEi and because of scarce data on angioedema risk in long term ACEi use, a decision tree was the preferred model to simulate patient pathways.

Angioedema incidence

ACEi induced angioedema incidence rates (per 1000 person-years) have been reported to be in the 1.97 – 4.38 range in observational studies by Miller et al. and Toh et al [15,16]. The OCTAVE randomised controlled trial (omapatrilat vs enalapril) by Kostis et al. reported 0.68% of patients developing angioedema during 24 weeks of follow-up [17]. Cumulative incidence of 1.79 (1.73-1.85) per 1,000 persons reported by Toh et al. was used for model input, based on 3,301 events in 1,845,138 exposed persons. 326 (9.88%) of these events were classified as 'serious', indicating the need for inpatient care [16].

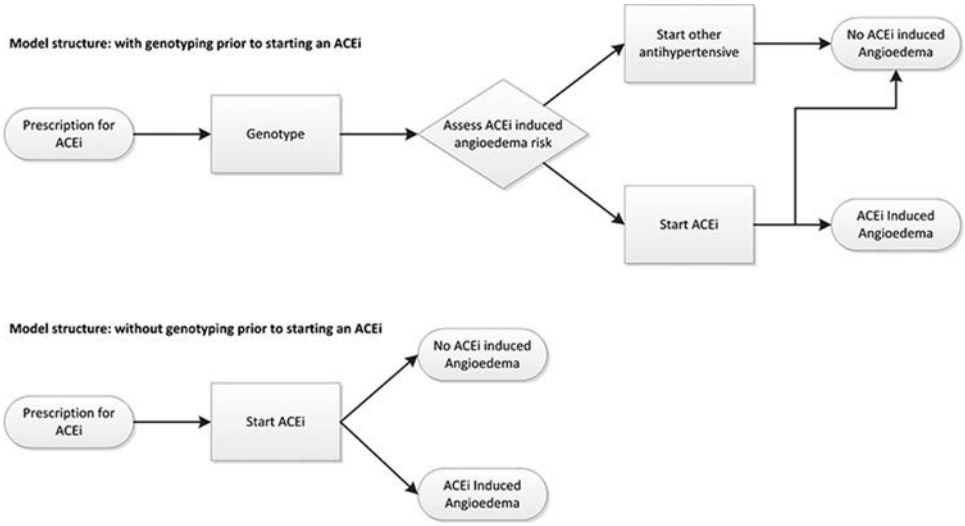


Figure 1. Model structures used. ACEi: Angiotensin-converting enzyme inhibitors.

ACEi treatment characteristics

After the initial ACEi prescription, patients stop and/or switch to another drug class in up to 44% of cases [18,19]. However, it is unlikely that switching and discontinuation patterns are influenced by being genotyped for angioedema prior to starting an ACEi. We therefore assume that all patients stay on the ACEi for one year unless they develop angioedema or receive a positive diagnosis by genomic assay. In these cases, according to guidelines, they are switched to another antihypertensive. The price of 'other antihypertensive' is the weighed per person average of the cost per user*number of users of ATC-classes C03 (diuretics), C08 (calcium antagonists) and C07 (beta-blockers), yielding an average cost per user per year of €23.77. This is higher than the annual per user cost of ACEis at €13.62 [20,21]. The difference between these two treatments (€10.15) is used as model input. Appendix 1 presents the data used for calculating treatment costs.

Subgroups

Subgroups of patients with an increased risk for developing ACEi induced angioedema have been identified by Miller et al [15]. People from African ancestry are at highest risk for developing ACEi induced angioedema, as shown in table 1.

Estimation of QALYs

Mortality due to ACEi related angioedema is extremely rare but, per case, results in a large loss of QALYs. Evidence on mortality is scarce and is mainly available in the form of case reports. To estimate mortality risk, studies that recorded intensive care unit (ICU) admittance or direct

Table 1. Subgroups with increased risk of developing ACEi induced angioedema. Data taken with permission from [15].

Risk Factor	Relative Risk
African ancestry	3.88
Age 65-74	1.42
Female Gender	1.45

mortality due to angioedema, were selected. The selected studies are shown in appendix 2. We assumed that all lethal cases would be admitted to the ICU. Then, lethal cases were divided by the total number of patients with angioedema admitted to the ICU to yield a mortality probability of 0.66% per ICU admittance. The average ACEi starter was 62 years old [22]. QALYs lost by premature mortality were calculated using life expectancy data from Statistics Netherlands and data on quality of life (QoL) per age group, yielding 17.20 QALYs [23,24].

Utilities

By making assumptions regarding answers to the validated EQ5D questionnaire and using the Dutch value set to calculate utility scores, specific health state utilities were generated [25].

Costs – resource use

Banerji et al. assessed the percentage of ACEi induced angioedema among all patients with angioedema presenting to the emergency department and described their healthcare requirements [26]. We combined these results with the data presented by Toh et al. to calculate the fraction of ICU stays of per total inpatient stays [16]. ICU stays were further specified using data from Soo Hoo et al [27]. They investigated ACEi induced angioedema requiring ICU admission, yielding data on hospitalisation duration [27]. Drug utilisation for the treatment of angioedema was not assessed as these costs are included in reference prices for hospital admittance.

Costs – prices

Costs for inpatient stays, GP & ED visits and ambulance use were based on reference prices published by the Dutch Manual for Costing in Economic Evaluations [28]. Drug utilisation and costs were retrieved from The Drug Information System and The Pharmacy Purchase Price database of the Dutch National Health Care Institute [20,21]. All costs are in Euros and, if applicable, indexed to 2016. Because of the one-year time horizon, discounting of future costs and effects was not necessary.

Analysis

The main outcome was the incremental cost effectiveness ratio (ICER) which is the ratio indicating the extra costs per QALY gained. In the OCTAVE-randomised controlled trial,

significantly more patients experienced angioedema in the first month of treatment (3.6/1000 vs 0.4/1000 after 24 weeks of follow-up) [17,29]. Observational studies by Toh et al. and Miller et al. reported that respectively 66% and 55% of events occurred in the first 90 days after ACEi initiation [15,16]. Based on these findings, we assume a one-year timeframe for the development of angioedema and all related healthcare utilisation. Model parameters are shown in table 2. Model parameter sensitivity was assessed by probabilistic and deterministic sensitivity analysis. In a deterministic sensitivity analyses the robustness of the model is tested for variations between the extremes of a plausible range of all parameters. In a probabilistic sensitivity analysis uncertainty of the analysis is examined by first constructing distributions for all parameters in the model. Secondly, the model picks a random value for all parameters from these distributions and the results are recalculated. This is repeated 5,000 times and the results are depicted in a scatterplot. We did not vary the cost components as these are based on reference prices.

Table 2. Model parameters and probability distributions.

Parameter	Value	Distribution	EQ5D input
Prob. of visiting ED*	0.4256	fixed	
Prob .of observational stay at ED*	0.0773	beta	
Prob. of patient stay (regular ward) *	0.0515	beta	
Prob. of ICU stay*	0.0472	beta	
Prob. of ambulance*	0.1141	beta	
Prob. of visiting GP*	0.574	beta	
Incidence rate of angioedema (per 1,000)	1.79	beta	
Prob. of mortality*	0.0004	beta	
Cost of visiting ED (€)	170.59	fixed	
Cost of observational stay at ED (€)	283.56	fixed	
Cost of inpatient stay (regular ward) (€)	737.14	fixed	
Cost of ICU stay (€)	8434.26	fixed	
Cost of requiring ambulance (€)	331.00	fixed	
Cost of visiting GP (€)	28.00	fixed	
Additional cost on other antihypertensive (€)	10.15	fixed	
Utility during ED visit	0.569	fixed	33333
Utility during observational stay at ED	0.569	fixed	33333
Utility during inpatient stay (regular ward)	0.569	fixed	33333
Utility during ICU stay	0.115	fixed	55533
Utility during GP visit	0.638	fixed	22222
Quality of Life lost by fatal angioedema	17.78	fixed	

*Probability is per angioedema event. ED = emergency department, ICU = intensive care unit, GP = general practitioner.

RESULTS

Base-case

The influence of sensitivity, specificity and test price on the ICER are shown in figures 2.1 and 2.2. Data points represent the test price at which the ICER exactly matches the WTP threshold. A grey point reflects a negative test price and a black point reflects a positive test price.

With a willingness-to-pay (WTP) threshold of €20,000 and €80,000 per QALY, a 100% sensitive and 100% specific test has a maximum cost of €1.30 and €1.95, respectively. A free and 100% sensitive test must at least be 87% and 81% specific to be cost effective at aforementioned WTP thresholds. The ICER of a free and 100% specific test is, only in this scenario, not influenced by sensitivity as it is free anyway and does therefore not generate false positives. At 90% specificity, a free test should be at least 79% and 52% sensitive for €20,000 and €80,000 WTP thresholds.

A change in specificity has a 3.5-fold higher impact on the ICER than a change in sensitivity. This is due to the additional cost of switching to another, more expensive, antihypertensive treatment in the case of false positives. False negatives do not cause additional costs; they only lower the maximum but ever positive price, indicated by a black dot at 100% specificity and lowest (50%) sensitivity.

Subgroups

Limiting genotyping to individuals at higher risk for ACEi induced angioedema has a profound influence on test requirements. Figures 3.1 and 3.2 display the relation between test parameters and the maximum price to meet WTP thresholds of €20,000 and €80,000, respectively. In this scenario, a perfect test meets the WTP thresholds at €5.03 and €7.55. The bandwidth for a positive test price has increased dramatically, as well as the spread between the two WTP thresholds. The requirement of a high specificity is no longer present: For a 100% sensitive test costing €3.00, the minimum specificity is 81% and 56% for aforementioned thresholds. Besides, the influence of specificity versus sensitivity lowered from ±3.5:1 to 1:1.

Probabilistic sensitivity analysis

Probabilistic sensitivity analysis (PSA) was based on a 90% sensitive and 90% specific test costing €0.50. Results, shown in figure 4, indicate 100% probability for both QALY gain and higher costs. Furthermore, there is a 10.6% and 55.3% probability of meeting €20,000 and €80,000 WTP thresholds, respectively. The base case ICER at specified parameters is €56,896. The PSA results are higher and lower than this base case in 57% and 43% of cases.

Deterministic sensitivity analysis

Deterministic sensitivity analysis (DSA), shown in figure 5, was also performed with a 90% sensitive and 90% specific test costing €0.50. Incidence rate of angioedema resulted in

the largest effect, followed by the additional cost of 'other antihypertensive'. ICU admission and mortality have a substantial effect on the ICER. The other parameters have a small or negligible influence.

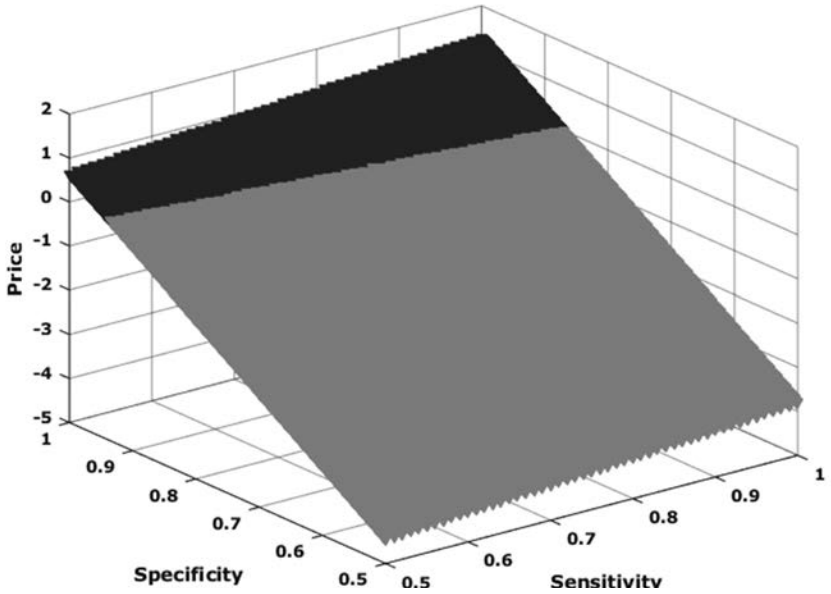


Figure 2a. Base-case results. Maximum test price to meet a willingness-to-pay threshold of €20,000.

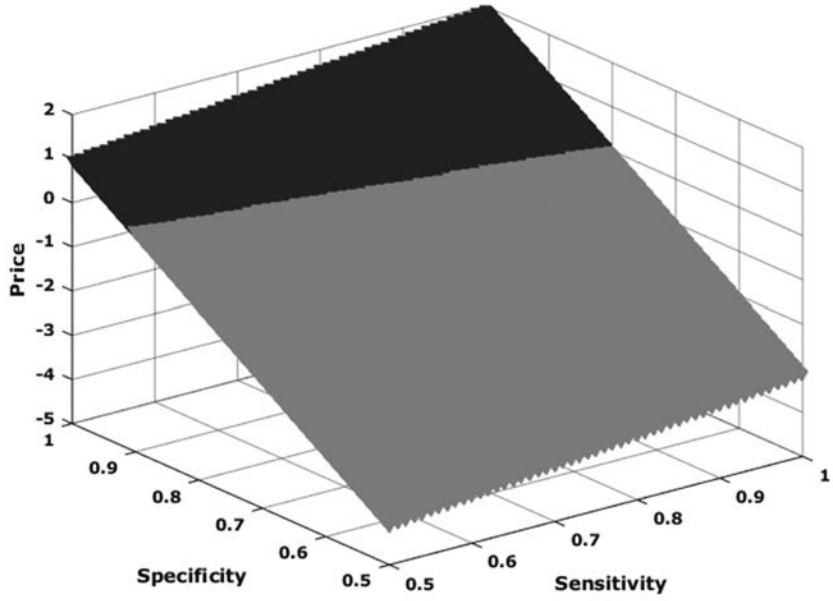


Figure 2b. Base-case results. Maximum test price to meet a willingness-to-pay threshold of €80,000.

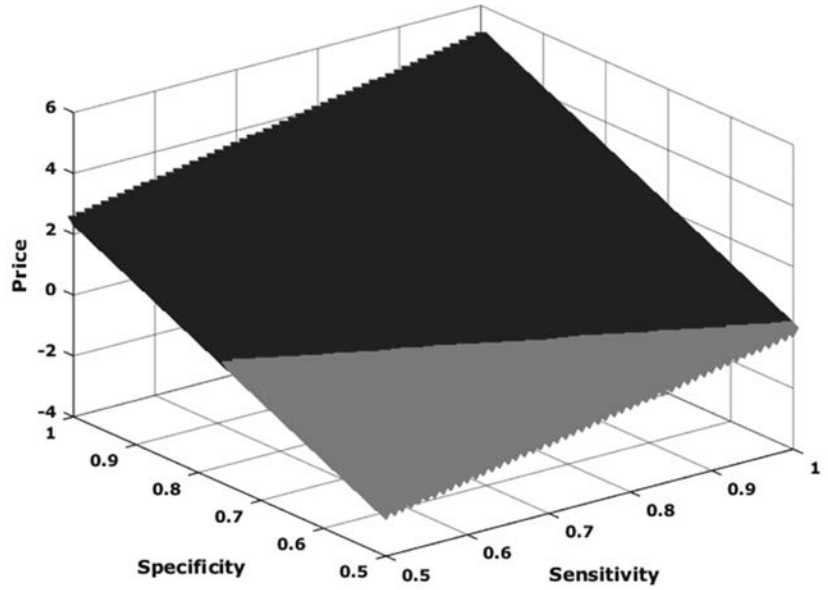


Figure 3a. Subgroup results. Maximum test price when only testing people of African ancestry (HR = 3.88) to meet a willingness-to-pay threshold of €20,000.

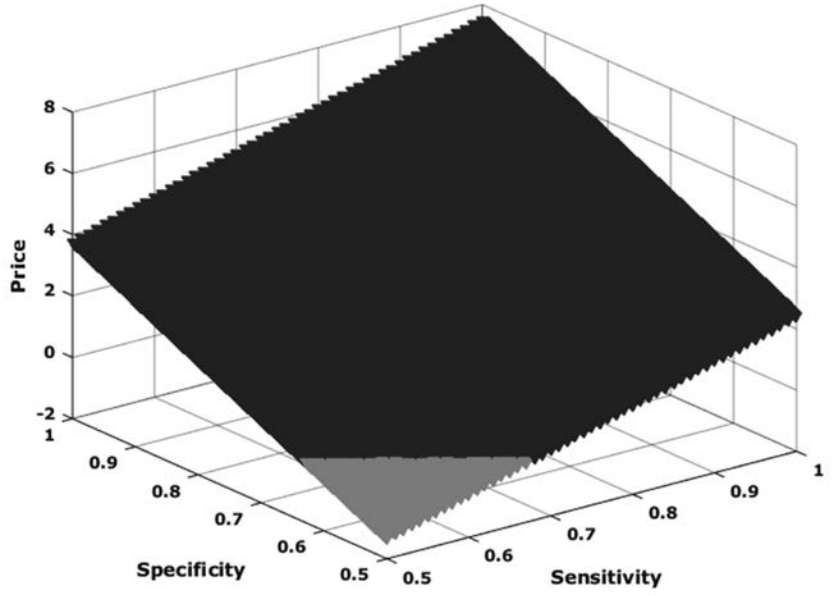


Figure 3b. Subgroup results. Maximum test price when only testing people of African ancestry (HR = 3.88) to meet a willingness-to-pay threshold of €80,000.

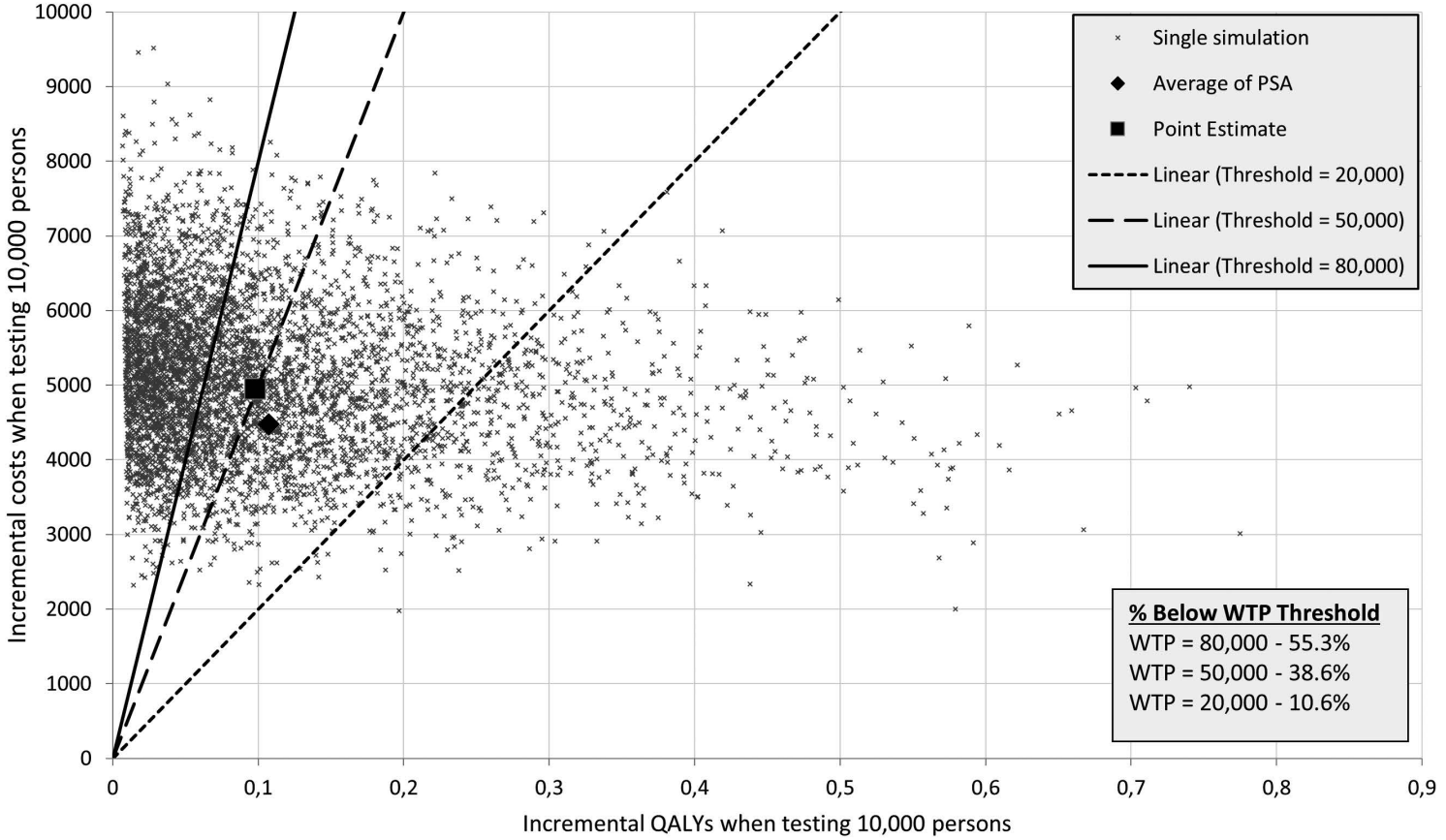


Figure 4. Probabilistic sensitivity analysis. Parameters used: 5000 simulations, fixed test price = €0.50, fixed test sensitivity = 90%, fixed test specificity = 90%.

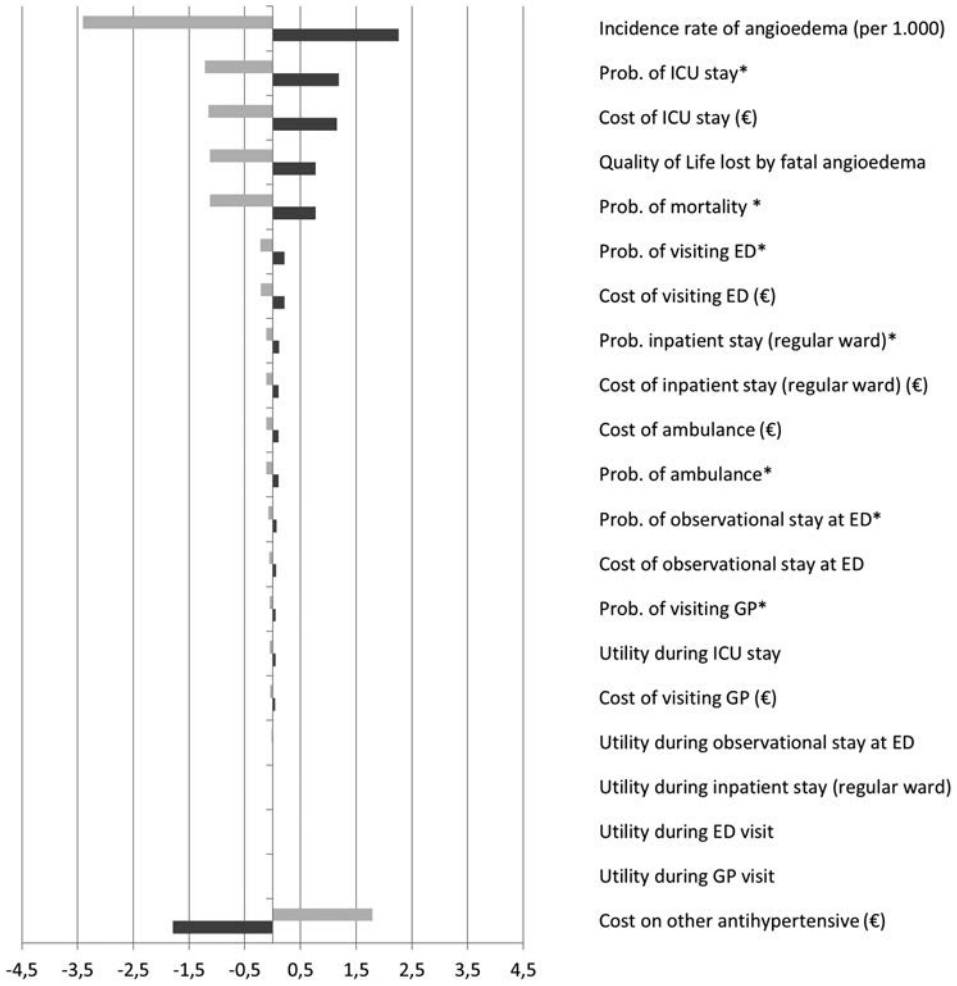


Figure 5. Deterministic sensitivity analysis. X-axis indicates the magnitude of the difference in ICER compared with a parameter. Black indicates a negative parameter change, grey a positive change. Parameters used: fixed test price = €0.50, fixed test sensitivity = 90%, fixed test specificity = 90%. The x-axis indicates the factor of the response versus a change in a parameter. † Probability is per angioedema event. ED: Emergency department; GP: General practitioner; ICU: Intensive care unit

DISCUSSION

We evaluated the specifications of a pharmacogenomic test for preventing ACEi induced angioedema in terms of the required specificity, sensitivity and price for achieving cost effectiveness. Our findings indicate that testing all ACEi starters is unlikely to be cost effective as >90% specificity, >93% sensitivity and a low (<€1.00) price would be required.

Our results highlight that limiting testing to high risk populations can be a fruitful endeavour for increasing cost effectiveness. This statement is further supported by the DSA demonstrating a major influence of angioedema incidence on the ICER. Miller et al. reported a relative risk of 3.88 and 1.45 for people of African-American ethnicity and for women, respectively. In our model this had a profound positive impact on parameter requirements. Further clarification of risk factors, for example for women of African-American ethnicity, could prove to lower diagnostic accuracy and test price to more favourable ranges that could warrant actual development of a PG test for this specific indication.

Nevertheless, individual tests for rare ADRs may not be very efficient. Plumpton and colleagues have shown that single testing is not always cost effective, even when a proper biomarker or SNP is present [1]. Their results indicate that mainly Human Leukocyte Antigen (HLA) polymorphisms are cost-effective single targets. These HLA polymorphisms predispose for hypersensitivity reactions, sometimes leading to very severe ADRs like Stevens-Johnson Syndrome (SJS) and Toxic Epidermal Necrolysis (TEN), induced by carbamazepine, abacavir and allopurinol. Not only are these ADRs more severe with mortality ranging from 10% to 40% for TEN, incidence rates of up to 5% are much higher than incidence rates of ACEi induced angioedema [1,30].

There could be a solution to biomarkers that do have value but are too costly to implement separately: Combine many of these tests into a single package or perform them together with a test that will be performed in routine daily practice. This way, the fixed costs of sampling, transport to a lab and reporting the results are spread and incremental costs per test could decrease dramatically. We can extend the idea of combining tests to whole exome or whole genome sequencing. Currently, these sequencing techniques are considered to be too costly for implementation in routine practice but prices have been falling dramatically [31]. When routine sequencing becomes part of daily clinical practice, all future genomic markers will deliver additional benefit to patients, regardless of the rarity of the predictor. Sadly, the full potential value that innovations may deliver in the future cannot be captured in traditional cost effectiveness analysis.

The two most important limitations of our study need to be addressed. Firstly, the DSA indicates a strong influence of the additional cost of antihypertensive treatment. This is the cost of a false positive case. In Dutch practice, switching to another antihypertensive is more expensive than ACEi treatment. This price difference is likely to be country specific. In other jurisdictions where ACEi treatment is more expensive than other antihypertensive treatment, the genotyping strategy would result in drug-cost savings in the event of a (false) positive diagnosis.

Secondly, model parameters were based on multiple studies with different study designs possibly leading to biased estimates. Especially our assessment of mortality risk was based on suboptimal evidence that required some assumptions. However, the DSA indicates a relatively low influence

of mortality risk on model outcomes. Utility scores were assessed by estimating the answers to the EQ5D questionnaire which is clearly sub-optimal. The DSA indicates that these parameters have a negligible effect on the results.

CONCLUSION

Our study indicates that testing all patients starting an ACEi for developing angioedema is unlikely to be cost effective as the test should have a high diagnostic accuracy combined with a sub €2.00 cost. Selectively testing only populations that have an increased risk of developing ACEi induced angioedema improves test characteristics needed and price for an ICER below €20,000 and €80,000. While separate testing for this variation for all ACEi starters or subgroups is not cost-effective, implementing whole exome or genome sequencing in routine clinical practice will result in economically attractive benefits of finding genetic variations like the one discussed here.

REFERENCES

1. Plumpton CO, Roberts D, Pirmohamed M, Hughes DA. A Systematic Review of Economic Evaluations of Pharmacogenetic Testing for Prevention of Adverse Drug Reactions. PharmacoEconomics. 2016;34:771–93.
2. Deenen MJ, Meulendijks D, Cats A, Sechterberger MK, Severens JL, Boot H, et al. Upfront Genotyping of DPYD*2A to Individualize Fluoropyrimidine Therapy: A Safety and Cost Analysis. J Clin Oncol Off J Am Soc Clin Oncol. 2016;34:227–34.
3. Zorginstituut Nederland. Kosteneffectiviteit in de praktijk [Internet]. 2015. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk/Kosteneffectiviteit+in+de+praktijk.pdf
4. Medicine-Spending-and-Growth_1995-2014.pdf [Internet]. [cited 2016 May 26]. Available from: https://www.imshealth.com/files/web/IMSH%20Institute/Reports/Medicines_Use_and_Spending_Shifts/Medicine-Spending-and-Growth_1995-2014.pdf
5. Baram M, Kommuri A, Sellers SA, Cohn JR. ACE inhibitor-induced angioedema. J Allergy Clin Immunol Pract. 2013;1:442–5.
6. Sánchez-Borges M, González-Aveledo LA. Angiotensin-converting enzyme inhibitors and angioedema. Allergy Asthma Immunol Res. 2010;2:195–8.
7. Powers BJ, Coeytaux RR, Dolor RJ, Hasselblad V, Patel UD, Yancy WS, et al. Updated report on comparative effectiveness of ACE inhibitors, ARBs, and direct renin inhibitors for patients with essential hypertension: much more data, little new information. J Gen Intern Med. 2012;27:716–29.
8. Overlack A. ACE inhibitor-induced cough and bronchospasm. Incidence, mechanisms and management. Drug Saf. 1996;15:72–8.
9. Dean DE, Schultz DL, Powers RH. Asphyxia due to angiotensin converting enzyme (ACE) inhibitor mediated angioedema of the tongue during the treatment of hypertensive heart disease. J Forensic Sci. 2001;46:1239–43.
10. Stauber T, Confino-Cohen R, Goldberg A. Life-threatening angioedema induced by angiotensin-converting enzyme inhibitors: characteristics and risk factors. Am J Rhinol Allergy. 2014;28:54–8.
11. Wadelius M, Marshall SE, Islander G, Nordang L, Karawajczyk M, Yue Q-Y, et al. Phenotype standardization of angioedema in the head and neck region caused by agents acting on the angiotensin system. Clin Pharmacol Ther. 2014;96:477–81.
12. Chan NJ, Soliman AMS. Angiotensin converting enzyme inhibitor-related angioedema: onset, presentation, and management. Ann Otol Rhinol Laryngol. 2015;124:89–96.
13. Mahmoudpour SH, Leusink M, van der Putten L, Terreehorst I, Asselbergs FW, de Boer A, et al. Pharmacogenetics of ACE inhibitor-induced angioedema and cough: a systematic review and meta-analysis. Pharmacogenomics. 2013;14:249–60.
14. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value Health. 2012;15:796–803.
15. Miller DR, Oliveria SA, Berlowitz DR, Fincke BG, Stang P, Lillienfeld DE. Angioedema incidence in US veterans initiating angiotensin-converting enzyme inhibitors. Hypertension. 2008;51:1624–30.
16. Toh S, Reichman ME, Houstoun M, Ross Southworth M, Ding X, Hernandez AF, et al. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. Arch Intern Med. 2012;172:1582–9.
17. Kostis JB, Packer M, Black HR, Schmieder R, Henry D, Levy E. Omapatrilat and enalapril in patients with hypertension: the Omapatrilat Cardiovascular Treatment vs. Enalapril (OCTAVE) trial. Am J Hypertens. 2004;17:103–11.

18. Burke TA, Sturkenboom MC, Lu S, Wentworth CE, Lin Y, Rhoads GG. Discontinuation of antihypertensive drugs among newly diagnosed hypertensive patients in UK general practice. J Hypertens. 2006;24:1193–200.

19. Mahmoudpour SH, Asselbergs FW, de Keyser CE, Souverein PC, Hofman A, Stricker BH, et al. Change in prescription pattern as a potential marker for adverse drug reactions of angiotensin converting enzyme inhibitors. Int J Clin Pharm. 2015;37:1095–103.

20. National Health Care Institute (Zorginstituut Nederland). English information | Medicijnkosten [Internet]. [cited 2016 Jun 3]. Available from: http://www.medicijnkosten.nl/english.asp

21. National Health Care Institute (Zorginstituut Nederland). GIPdatabank. The Drug Information System of National Health Care Institute [Internet]. [cited 2018 Aug 3]. Available from: https://www.gipdatabank.nl/infoPagina.asp?naam=English

22. Makani H, Messerli FH, Romero J, Wever-Pinzon O, Korniyenko A, Berrios RS, et al. Meta-analysis of randomized trials of angioedema as an adverse event of renin-angiotensin system inhibitors. Am J Cardiol. 2012;110:383–91.

23. Centraal Bureau voor de Statistiek. Statline.

24. Nyman JA, Barleen NA, Dowd BE, Russell DW, Coons SJ, Sullivan PW. Quality-of-life weights for the US population: self-reported health status and priority health conditions, by demographic characteristics. Med Care. 2007;45:618–28.

25. van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2012;15:708–15.

26. Banerji A, Clark S, Blanda M, LoVecchio F, Snyder B, Camargo CA. Multicenter study of patients with angiotensin-converting enzyme inhibitor-induced angioedema who present to the emergency department. Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol. 2008;100:327–32.

27. Soo Hoo GW, Lin HK, Junaid I, Klaustermeyer WB. Angiotensin-converting Enzyme Inhibitor Angioedema Requiring Admission to an Intensive Care Unit. Am J Med. 2015;128:785–9.

28. Tan SS, Bouwmans CAM, Rutten FFH, Hakkaart-van Roijen L. Update of the Dutch Manual for Costing in Economic Evaluations. Int J Technol Assess Health Care. 2012;28:152–8.

29. Bezalel S, Mahlab-Guri K, Asher I, Werner B, Sthoeger ZM. Angiotensin-converting enzyme inhibitor-induced angioedema. Am J Med. 2015;128:120–5.

30. Yang C-Y, Dao R-L, Lee T-J, Lu C-W, Yang C-H, Hung S-I, et al. Severe cutaneous adverse reactions to antiepileptic drugs in Asians. Neurology. 2011;77:2025–33.

31. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30:418–26.

32. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA. 2002;288:2981–97.

33. Grant NN, Deeb ZE, Chia SH. Clinical experience with angiotensin-converting enzyme inhibitor-induced angioedema. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 2007;137:931–5.

34. Kyrmizakis DE, Papadakis CE, Liolios AD, Karatzanis AD, Malandrakis S, Skoulakis CE, et al. Angiotensin-converting enzyme inhibitors and angiotensin II receptor antagonists. Arch Otolaryngol Head Neck Surg. 2004;130:1416–9.

APPENDICES

Appendix 1. Cost of switching to another antihypertensive.

Other Antihypertensive	ATC	No. users	Cost/user (2014)	Total costs
Diuretics	C03	1,143,000	€22.48	€25,694,640
Calcium antagonists	C08	831,430	€39.09	€32,500,599
Beta blockers	C07	1,642,000	€16.91	€27,766,220
	Sum:	3,616,430	Sum:	€85,961,459
Average cost per user per year:	€23.77			
ACEi (cost/user)	€13.62			
Difference:	€10.15			

Appendix 2. Studies included in mortality assessment.

Study	No. Angioedema	No. ICU	No. mortality
OCTAVE [17]	86	0	0
ALLHAT [32]	38	1 (assumed)	1
Grant et al. [33]	228	0	0
Soo Hoo et al. [27]	50	50	0
Banerji et al. [26]	220	24	0
Kyrmizakis et al. [34]	31	1	0
Chan et al. [12]	88	75	0

2.2

Phase I/II based early economic evaluation of acalabrutinib for relapsed chronic lymphocytic leukemia

Rick A. Vreman, Joost W. Geenen, Anke M. Hövels, Wim G. Goettsch, Hubert G.M. Leufkens, Maiwenn J. Al

Accepted for publication in Applied Health Economics and Health Policy.

ABSTRACT

Objectives

The objective of this study was to construct an early economic evaluation for acalabrutinib for relapsed chronic lymphocytic leukaemia (CLL) in order to assist early reimbursement decision making. Scenarios were assessed to find the relative impact of critical parameters on incremental costs and quality-adjusted life years (QALYs).

Methods

A partitioned survival model was constructed comparing acalabrutinib to ibrutinib with a National Health Service (United Kingdom) perspective. This model included states for progression free survival (PFS), post-progression survival (PPS) and death. PFS and overall survival (OS) were parametrically extrapolated from ibrutinib publications and a preliminary hazard ratio based on phase I/II data was applied for acalabrutinib. Deterministic and probabilistic sensitivity analyses were performed, and 1296 scenarios were assessed.

Results

The base case ICER is 61,941 £/QALY, with 3.44 incremental QALYs and incremental costs of £ 213,339. Deterministic sensitivity analysis indicated that survival estimates, utilities and treatment costs of ibrutinib and acalabrutinib and resource use during PFS have the greatest influence on the ICER. Probabilistic results under different development scenarios indicated that greater efficacy of acalabrutinib will decrease the likelihood of cost-effectiveness (from 63% at no effect to 2% at maximum efficacy). Scenario analyses showed that a reduction in PFS did not lead to great QALY differences (-8 to -14% incremental QALYs) although it did greatly impact costs (-47 to -122% incremental pounds). For OS, the opposite is true (-89 to -93% QALYs and -7 to -39% pounds).

Conclusions

Acalabrutinib is not likely to be cost-effective compared to ibrutinib under current development scenarios. The conflicting effects of OS, PFS, drug costs and utility during PFS show that determining cost-effectiveness of acalabrutinib without insight into all parameters complicates HTA decision making. Early assessment of cost-effectiveness of new products can support development choices and reimbursement processes through effective early dialogues between stakeholders.

INTRODUCTION

Bruton's tyrosine kinase (BTK) inhibitors represent a new line of treatment for chronic lymphocytic leukaemia (CLL), with drugs in development and one already on the market: ibrutinib. Ibrutinib has now been approved for previously treated and untreated CLL patients and has shown to be clinically effective with a durable response [1–3]. However, ibrutinib is not entirely specific for BTK. It may also inhibit epidermal growth factor receptor (EGFR), interleukin-2–inducible T-cell kinase (ITK), T-cell X chromosome kinase (TXK), and tyrosine kinase expressed in hepatocellular carcinoma (TEC) family proteins leading to side effects such as bleeding, atrial fibrillation, rash and diarrhoea [4–6].

A more specific BTK inhibitor showing promise in preclinical and early clinical trials is acalabrutinib (ACP-196). In preclinical research acalabrutinib did not inhibit EGFR, TEC, ITK, or other agents [7–10]. Similar to ibrutinib, acalabrutinib binds covalently to Cys481 in the ATP binding pocket of BTK. It shows a rapid oral absorption with a short plasma half-life theoretically leading to less toxicity [11,12]. Early clinical studies with acalabrutinib have shown overall response rates of 95% at median follow-up of 14.3 months and mostly grade 1 or 2 adverse events without dose-limiting toxicity [13].

Based on these findings, acalabrutinib would be a valuable addition to the therapeutic options for CLL. However, patient access also relies on the decisions of health technology assessment (HTA) bodies. In the United Kingdom the National Institute for Health and Care Excellence (NICE) reported that ibrutinib's initial price led to a base case incremental cost-effectiveness ratio (ICER) of £45,486 per quality-adjusted life year (QALY) when compared with treatment with physicians' choice [14]. They advised to reimburse ibrutinib only if the negotiated (confidential) discount would be upheld. Such evaluations by HTA bodies and concomitant price negotiations can take up to a year, delaying patient access.

To reduce delays in patient access, it is possible to start early with the assessment of added value of a new therapy. This can start during preclinical development but is more regularly executed during phase I or II clinical trials. One method is the use of early cost-utility analyses, which can clarify the relative impact of the parameters that drive cost-effectiveness. Early cost-utility analyses allow decision makers and manufacturers to streamline clinical development and reimbursement processes [15–18].

The objective of this study was to construct, based on published phase I/II data, an early cost-utility analysis comparing acalabrutinib to ibrutinib for chronic lymphocytic leukaemia in order to assist early reimbursement decision making. Sensitivity analyses are performed and possible development scenarios are assessed, identifying critical parameters and quantifying their relative impact on incremental costs and quality-adjusted life years.

METHODS

General

An effectively lifetime partitioned survival model comparing acalabrutinib to ibrutinib was constructed in Microsoft Excel (Microsoft, Redmond, WA) from an NHS (UK) perspective. As portrayed in figure 1, included health states were progression free survival (PFS), post-progression survival (PPS), and death. PPS is split into two sub-states, i.e. subsequent treatment (PPS-ST) and best supportive care (PPS-BSC). The modelled population is based on the only available phase I/II trial for acalabrutinib and assumed representative for the UK relapsed CLL patient population [13]. The model is based on the NICE assessment of the manufacturer's submission for ibrutinib, appraisal number TA429 [14]. Patients move between health states in cycles of 28 days with a time horizon of 30 years (effectively lifetime). Half cycle corrections are applied. Costs and outcomes are both discounted by 3.5%. The model is constructed according to ISPOR Good Modelling Practice and method reporting follows the CHEERS statement for reporting standards [19].

Treatment, comparator and subsequent treatment

Ibrutinib is administered as 420 mg/day (3 capsules) until disease progression or until no longer tolerated by the patient. Acalabrutinib is given as 200 mg/day (2 capsules). After progression, 41.9% of patients receive subsequent treatment. Subsequent treatment consists of rituximab and idelalisib. Rituximab is given during six cycles of four weeks, with an initial dose of 375 mg/m2 and subsequent doses of 500 mg/m2, according to the NICE guideline for CLL [20]. Idelalisib is administered until disease progression or death in a dose of 150 mg twice daily. A dose intensity of 94.8% was applied for acalabrutinib, ibrutinib and idelalisib, in accordance with findings

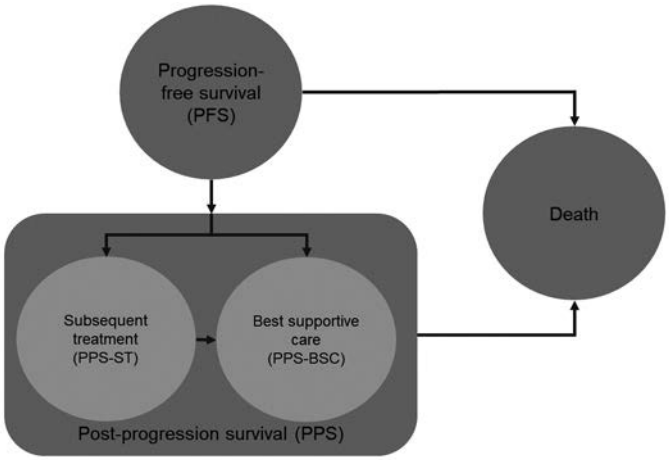


Figure 1. Model structure.

from the RESONATE trial [2,14]. In this study, acalabrutinib is assessed relative only to its primary comparator from the same class within the same indication (ibrutinib), as it is expected this will be the main competitor in practice. Ofatumumab, physician's choice or other treatment regimens are not assessed in this study.

Survival data

Efficacy of ibrutinib has been established in a phase III multicentre, open-label, randomised clinical trial comparing it to ofatumumab [2]. Preliminary efficacy of acalabrutinib was established in a multicentre, open-label, single-arm phase I/II trial [13].

PFS and OS individual patient data for ibrutinib was reconstructed from the reported Kaplan-Meier curves [21]. Multiple parametric survival curves were tested: an exponential, Weibull, log-logistic and lognormal distribution. The exponential curve showed physiological plausibility and overall best fit for OS as well as PFS, corresponding with the Expert Review Group (ERG) comments on the ibrutinib submission for NICE [14]. For acalabrutinib, efficacy compared to ibrutinib was established through an indirect treatment comparison based on the extracted individual patient data, providing a hazard ratio (HR) of 0.479 for PFS (95% confidence interval (CI) 0.230 – 0.998) and 0.391 for OS (95% CI 0.141 – 1.081). Because these are based on very limited data, we set the range of variation in sensitivity and scenario analyses for these HR's from 0.479 and 0.391 to 1.00, representing the full range up until no benefit for acalabrutinib. Furthermore, no assumptions are made about the distribution of this effect. The base case (which equals the maximum HR) is tested (0.479 and 0.391 for PFS and OS respectively) and five uniform steps up until no benefit, resulting in six scenarios for the HR's (base case/maximum, 80%, 60%, 40%, 20%, and no benefit). In the probabilistic sensitivity analysis, the base assumption is that OS and PFS are independent. To test the effect of this assumption, the six scenarios for the PSA are also implemented with the OS and PFS sharing the same random number when sampled, creating dependence. The full survival calculations, fitting criteria and ranges for all parametric models are provided in appendix 1.

Post-progression survival is defined as OS minus PFS and comprises patients on subsequent treatment as well as patients on best supportive care. Post-progression survival on subsequent treatment (PPS-ST) is implemented by plotting a Weibull curve from the progression free survival given in the Kaplan-Meier graph provided by Furman et al [22]. The Weibull curve was chosen because this was evaluated as the most suited curve for this treatment by NICE. In this multicentre, randomised, double-blind, placebo-controlled, phase 3 study, the efficacy of rituximab and idelalisib combination therapy was assessed in patients with relapsed CLL [22]. After 80 cycles (75 months), Weibull plotted survival is less than 0.01% and therefore assumed to be 0. Because transition probabilities to the PPS state are not explicitly modelled in a partitioned survival model, entry into the PPS state each cycle was calculated by subtracting a specific background mortality from the proportion of patients leaving PFS. This background

mortality was retrieved from the Life Expectancy Tables from the Office for National Statistics [23]. This method was also used in the ibrutinib submission though the background mortality was considered fixed whereas ours increases with increasing age.

Costs and resource use

Unit costs for the drug treatments are provided by the British National Formulary [24]. All costs are reported in 2018 UK pound sterling. When cost inputs were based on different years, they were inflated with the Hospital and Community Health Services Pay and Price Index and, after discontinuation of this index in 2017, with the health-specific subset of the consumer price inflation index [25]. Daily costs for acalabrutinib treatment are assumed equal to ibrutinib in the base case and sensitivity analyses and varied through scenario analyses, testing for 30% pricing premiums and reductions, see table 1. Costs for rituximab are only inflicted in the first six cycles of subsequent treatment, in accordance with its approved indication, and are based on an average body surface area of 1.9 m2 [14]. Full calculations for costs per cycle of treatment with acalabrutinib, ibrutinib and rituximab + idelalisib are provided in appendix 2.

Costs of grade 3 and 4 adverse events (AEs) were included according to the UK national schedule of reference costs 2015-2016 [26]. Incidences were implemented from clinical trials for acalabrutinib and from the NICE ibrutinib assessment [13,14]. Adverse event costs are inflicted once, in the first cycle. This matches the approach used in the NICE ibrutinib submission and is supported by the fact that onset of side effects was generally within the first half year and the duration of side effects was short [2,14].

Annual healthcare resource use such as hospital visits or blood tests associated with routine follow-up care was included. Resource use is based on expert elicitation reported by the manufacturer in the ibrutinib submission and differs per model state (PFS, PPS-ST & PPS-BSC) and whether the patient in the PFS state is a complete responder (CR), partial responder (PR) or non-responder (NR, including stable disease and progressive disease). Treatment responses for ibrutinib were reported to be 84% PR, 6% CR and 10% NR. For acalabrutinib, 95% were PR and 5% were NR [13,14]. Full calculations of adverse event costs and resource use per treatment are provided in appendix 3.

Costs for the death state are inflicted once in the cycle when death happens, and equal the per patient costs of health care utilisation during the last 30 days of life for patients of age 65+ with any cancer reported by Bekelman et al [27]. Costs and ranges for sensitivity analyses are stated in table 1.

Utilities

Utility for acalabrutinib was not available and is therefore assumed equal to the ibrutinib utility of 0.799 reported in the RESONATE trial, as measured by EQ-5D-5L. UK weights were used to

Table 1. Input parameters and their ranges. Calculations are presented in appendices 1 to 4. PFS: progression free survival, OS: overall survival, PPS-ST: post-progression survival on subsequent treatment, PPS-BSC: post-progression survival on best supportive care.

Parameter	Base	Min	Max	Distrib.	Source
General					
Mean age	62			Fixed	[13]
Mean body surface area (m2)	1.9			Fixed	
Time horizon (years)	30			Fixed	N/A
Discount rates					
Costs	0.035			Fixed	[25]
Effects	0.035			Fixed	[25]
Utilities					
Progression-free survival					
Acalabrutinib	0.799	0.799	0.837	Beta	[2,14]
Ibrutinib	0.799	0.799	0.837	Beta	[2,14]
Post-progression survival	0.701	0.631	0.771	Beta	[2,14]
Adverse event disutility					
Acalabrutinib	0.065	0.058	0.071	Beta	[13,14]
Ibrutinib	0.091	0..082	0.100	Beta	[1,2,14]
Costs (£)					
Treatment during progression-free survival					
Acalabrutinib	4279	2996	5563	Fixed	[14]
Ibrutinib	4279	2996	5563	Fixed	[14]
Adverse events					
Acalabrutinib	639	319	958	Gamma	[13,14]
Ibrutinib	829	414	1243	Gamma	[1,2,14]
Progression-free survival state					
Acalabrutinib	244	122	367	Gamma	[14]
Ibrutinib	245	122	367	Gamma	[14]
Post-progression survival state					
Rituximab + idelalisib cycle 1-6	5428	1206	7137	Gamma	[20,23]
Rituximab + idelalisib cycle 7+	3298	780	4368	Gamma	[20,23]
Best supportive care	177	88	265	Gamma	[14]
Death	3051	1525	4576	Gamma	[26]
Survival parameters					
Ibrutinib (hazard rates)					
Progression free survival	0.013	0.018	0.010	Normal	[2]
Overall survival	0.008	0.010	0.006	Normal	[2]
Acalabrutinib (hazard ratios)					
Progression-free survival	0.479	1.000	0.479	6 steps	[13]
Overall survival	0.391	1.000	0.391	6 steps	[13]
Subsequent treatment (ST) (Weibull)					
Scale	0.008	0.008	0,008	Fixed	[22]
Shape	1.582	1.758	1.439	Fixed	[22]
Percentage receiving ST	0.419	0.219	0.619	Normal	[14]

generate patient utilities [14]. As an optimum utility for sensitivity and scenario analyses, utility was calculated according to utilities awarded to the response states [28]. Adverse event disutility was calculated according to the incidence and utility decrement of AEs reported in clinical trials for ibrutinib and acalabrutinib [1,2,13]. As with AE costs, disutility according to adverse events was inflicted once, in the first cycle. The full calculations for disutility due to adverse events are provided in appendix 4. To get the post-progression utility, the baseline utility is corrected for the reported utility decrement of 0.098 associated with progression [14]. Base case utilities and ranges are provided in table 1.

Sensitivity and scenario analyses

Uncertainties were assessed through sensitivity and scenario analyses. In a one-way sensitivity analysis, the impact of each model input parameter was assessed individually according to their minimum and maximum value provided in Table 1. This deterministic sensitivity analysis shows the impact of the minimum and maximum values for each separate parameter on the ICER. Additionally, probabilistic sensitivity analyses were performed for each of the six HR steps, thus testing cost-effectiveness for different acalabrutinib efficacy scenarios. Body surface area and age were not varied in the PSA, in line with the ibrutinib submission.

From the deterministic analysis, important parameters were selected that had a profound influence on the ICER, defined by variations >5% from the base case ICER for the minimum and/or maximum scenario. For these critical parameters, all possible combinations of parameter values were tested in scenarios. This means that for each value for each important parameter (the base case, minimum and maximum values), all combinations of values for the other parameters are tested. This results in an overview of the impact of each parameter on incremental costs and QALYs. The calculation for this relative impact is given in appendix 5.

RESULTS

The base case ICER is 61,941 £/QALY, with 3.44 incremental QALYs with incremental costs of £ 213,339. Absolute costs and QALYs are £317,853 and 5.88 for ibrutinib and £531,192 and 9.33 for acalabrutinib, respectively. The one-way sensitivity analysis shown in figure 2 indicates that survival estimates, utilities and treatment costs of ibrutinib and acalabrutinib and resource use during PFS have a distinct influence on the ICER. As figure 2 also shows, OS and PFS have opposite effects, i.e. when OS for acalabrutinib is reduced, it increases the ICER, whilst reducing PFS leads to a smaller ICER. Higher utility and lower treatment and resource costs during PFS reduce the ICER. For ibrutinib, the opposite is true for all variables.

Results of the probabilistic sensitivity analysis are shown in figure 3. When the efficacy of acalabrutinib grows (HR further from 1.00), the incremental costs and incremental QALYs both increase, but not simultaneously. With no effect (HRs PFS & OS = 1.00), the probability

of cost-effectiveness is 63% with a WTP threshold of 50,000 £/QALY. This declines gradually from 42%, to 25%, 10%, 3% and 2% (HRs: 20, 40, 60, 80% and HR maximum, respectively). Thus, higher efficacy and the resulting higher QALYs lead to disproportionately higher costs for acalabrutinib, when the price is equal to ibrutinib. Assuming dependence between PFS and OS did not lead to very different results. Mean ICERs are within +/- 3% of mean ICERs without dependence. Probabilities of cost-effectiveness with a WTP threshold of 50,000 £/QALY are 60%, 42%, 26%, 12%, 5% and 1% for HRs 1.00, 20%, 40%, 60%, 80% and maximum respectively.

Scenarios

The deterministic analysis provided ten parameters that explained the majority of the variation. Of those ten, resource use costs and treatment costs during PFS are perfectly correlated with each other. Therefore, these were combined into one parameter (called costs acalabrutinib/ibrutinib) in order to reduce the number of scenarios. Eight parameters remain: four have base case, minimum and maximum values (costs during PFS for both treatments and PFS & OS survival parameters for ibrutinib) and 4 have only 2 values (base case and minimum (hazard

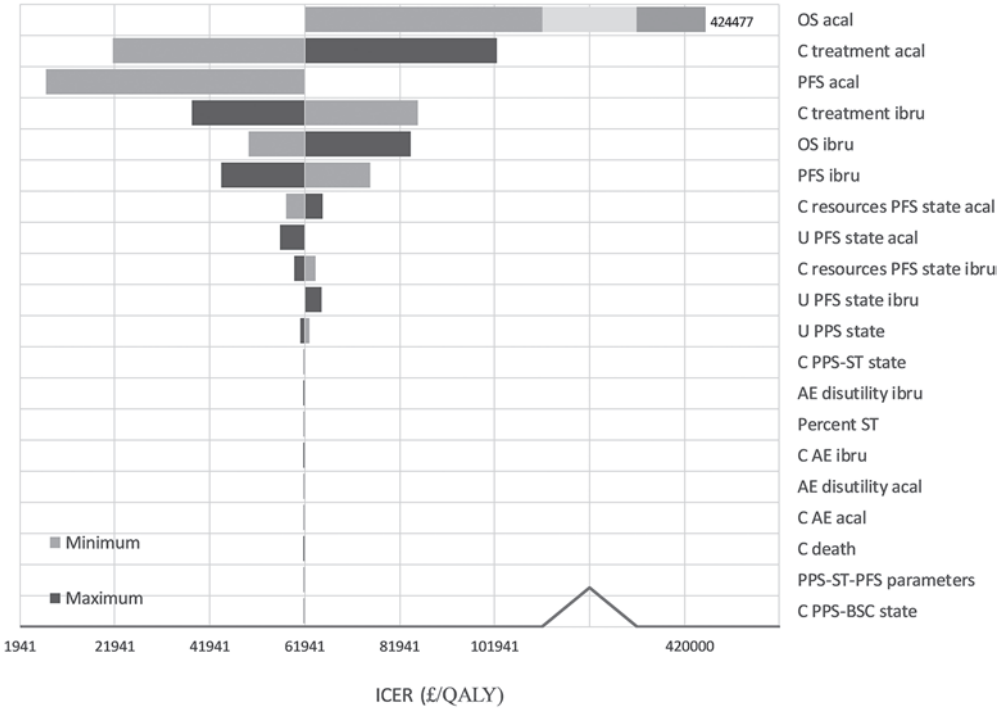


Figure 2. Relative effects of individual parameters in comparison to the base case ICER of 61,941 £/QALY. Note that some parameters are only varied one way because the base case represents the maximum or minimum. PFS = progression-free survival; OS = overall survival; PPS = post-progression survival; C = costs; U = utility; acal = acalabrutinib; ibru = ibrutinib; ST = subsequent treatment; BSC = best supportive care; AE = adverse events.

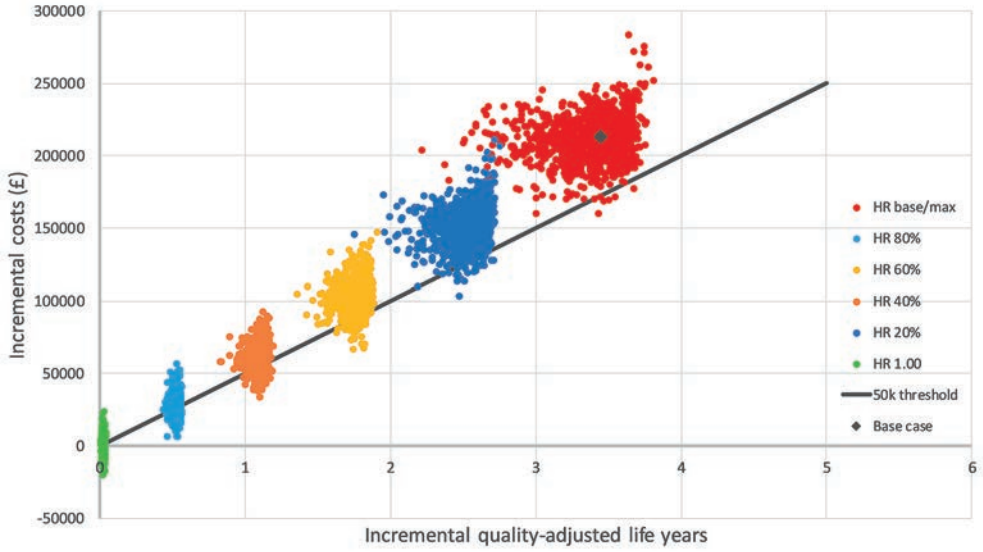


Figure 3. Cost-effectiveness planes for different hazard ratios. The dark grey dot indicates the base case.

ratios for acalabrutinib) or base case and maximum (utility during PFS for both treatments)). This led to a total of 34*24 = 1296 scenarios, which were all tested. For each parameter, the effect on incremental QALYs and incremental costs in the minimum and/or maximum scenario versus the base case scenario was calculated for all scenarios. Figure 4 shows these effects for each of the included eight parameters. The size of the impact a parameter has on incremental costs and QALYs depends on the scenario, i.e. the values of the other input parameters. Figure 4 shows all distinctive values for each parameter. The results indicate that overall survival is a main driver of incremental QALYs throughout all scenarios, however, it does not impact costs proportionally. The inverse is the case for progression-free survival, which greatly impacts costs but does not impact QALYs proportionally.

A minimal OS of acalabrutinib leads to a reduction in incremental QALYs (89 – 93% of base case QALYs) whilst not greatly influencing incremental cost (reduction of 7 – 39% of base case costs), leading to a higher ICER. A minimal PFS leads to smaller incremental costs (reduction of 47 – 122% of base case costs) but does not significantly affect QALYs (reduction of 8 – 14% of base case QALYs), leading to a smaller ICER. Effects of these parameters for ibrutinib are similar but have an opposite direction of effect (i.e. small PFS leads to a greater ICER). Utility has a relatively minor effect. Incremental QALY benefits due to greater utility during treatment (based on response rates as described in the methods section) are relatively small throughout all scenarios (4.5 – 9.7% of base case). QALY and cost benefits due to fewer side effects were even smaller (and thus not included in scenario analyses).

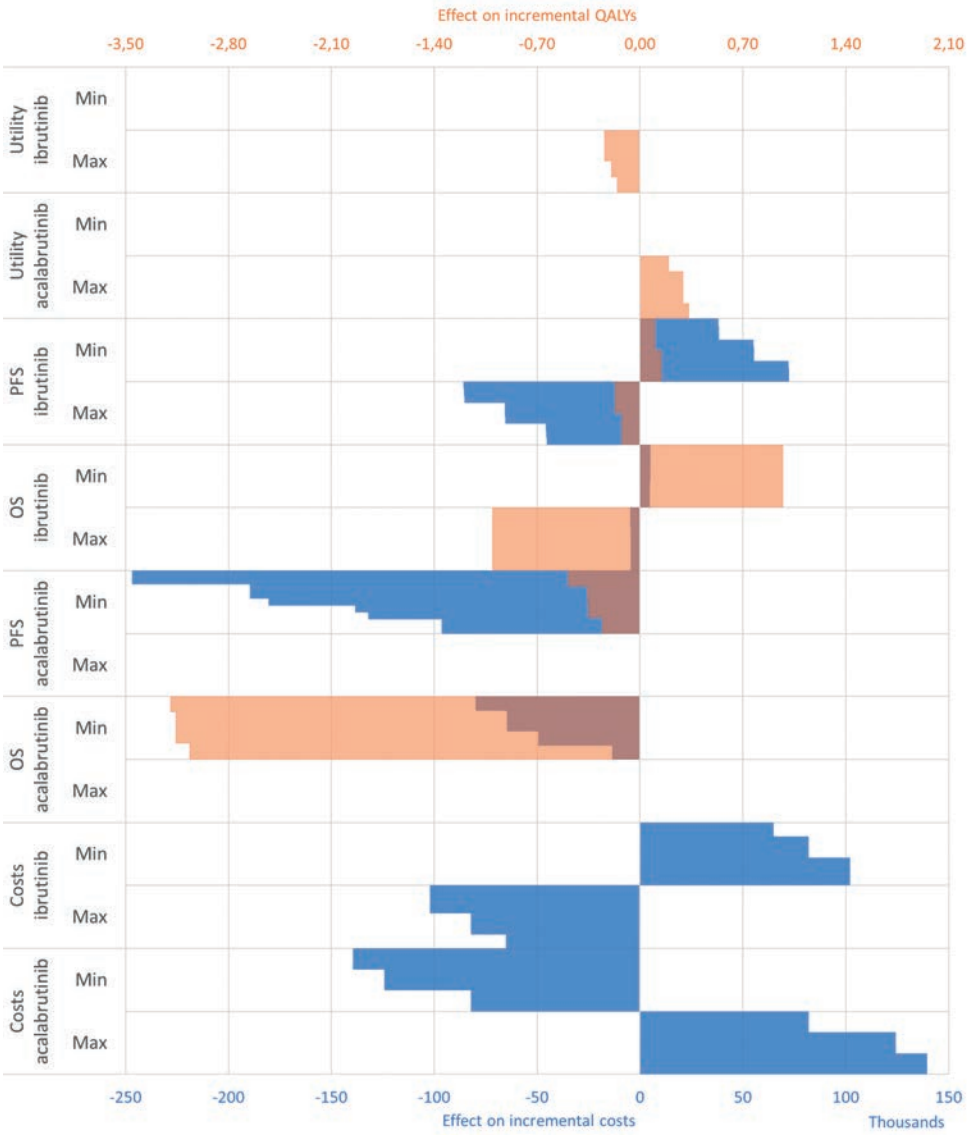


Figure 4. Range of variation in incremental costs and QALYs due to each critical parameter throughout all scenarios.

DISCUSSION

In this model, the base case ICER for acalabrutinib compared to ibrutinib was 61,941 £/QALY, with 3.44 incremental QALYs and incremental costs of £213,339. This indicates that even with a price equal to ibrutinib, acalabrutinib is not cost-effective with a willingness-to-pay threshold of £50,000. The probability for acalabrutinib to be cost-effective declines with greater efficacy. This finding is explained by the fact that longer progression-free survival leads to disproportionally higher costs, even though overall survival is prolonged as well.

In the deterministic analysis, all parameters associated with PFS and OS had significant impact on the ICER. Parameters that had little impact were all one-off parameters (adverse events, death costs) and parameters associated with the PPS state. Apparently, subsequent treatment choices do not greatly affect cost-effectiveness of acalabrutinib.

A price for acalabrutinib higher than ibrutinib, with a threshold of £50,000/QALY, would not lead to a cost-effective scenario. This indicates that even though acalabrutinib would show good survival benefits, reimbursement for a higher price is still unlikely, impeding patient access. With treatment costs set at the base case, ibrutinib nor acalabrutinib is cost-effective in the PFS state. However, the DSA clarifies that treatment costs have a large impact on the ICER. Thus, a cost reduction may potentially lead to time spent the PFS state being cost-effective, which would greatly alter cost-effectiveness of both treatments. Indeed, in the tested scenario where both drug costs are minimal (with the rest of the parameters at base case), the ICER is £44,000 per QALY.

For decision purposes, the cost-effectiveness of an expensive treatment in a certain health state can be roughly estimated from its treatment costs and the utility in that state. If this estimate greatly exceeds the threshold, a modelling exercise may be redundant. However, as our analysis shows, modelling may still be very useful to provide insight in the relative effects of all parameters and their relevance to the ICER. When varied between their plausible bounds, improvements in PFS and OS led to opposite effects on the ICER. The relationship between PFS, OS, and the ICER is often not straightforward within the context of an incremental analysis. For example, when costs occur during PFS that are higher than the willingness-to-pay threshold, the moderate QALY improvement associated with prolonged PFS may not offset these costs if prolonged PFS does not translate to prolonged OS. A positive correlation may exist but previous publications have highlighted that these correlations are very inconsistent between and within different cancer types [29]. In this NICE decision support unit publication, it was furthermore deemed unclear how evidence supporting a correlation should be quantitatively implemented in a cost-effectiveness model. Thus, our primary assumption was independence of PFS and OS, but we ran scenario analyses assuming dependence through sampling from a shared random number. It should be noted that this was possible because an exponential curve was implemented for both survival curves. If one of the curves would have been parameterised differently (e.g.

Weibull), this approach would not have been viable. Correlating PFS and OS changed the shape of the cost-effectiveness plane but it did not greatly impact the probability of the treatment being cost-effective. However, the impact of correlation between PFS and OS may be greater for other drugs or in different disease areas. For early cost-effectiveness models, when there is fairly little information on the relation between PFS and OS, we strongly advise to test the effects of correlation between survival curves in scenario analyses.

Recent research has shown that overall survival was included as a primary outcome in studies in only 18/68 (26%) of drug indications, whereas PFS accounted for another 31 (46%) and response rates for 11 (16%) [30]. For drug indications that lacked data on OS at time of approval, after a median follow-up of 5.4 years after market entry, only 7% were subsequently shown to extend life. Our findings emphasise that this lack of demonstrated OS benefit induces problems in reimbursement processes.

Acalabrutinib has been approved by the FDA for mantle cell lymphoma (MCL) via the accelerated approval pathway based on benefit in overall response rate. Though treatments for MCL and CLL are different, of interest is that our analysis shows that the manufacturer would get the best price in CLL when they solely show benefit through better response rates and do not prove PFS benefit. Acalabrutinib is currently being investigated in several phase II and III clinical trials for first-line and subsequent treatment in CLL, MCL and at least eight other indications varying from rheumatoid arthritis to urothelial carcinoma [10]. Our analysis has shown that perverse incentives might be present in reimbursement processes. Therefore, it is essential that stakeholders engage early and discuss adequate evidence generation plans prospectively based on scenario analyses such as the one presented here.

Strengths and limitations

A strength of this research effort is that it establishes an indication of cost-effectiveness well in advance of any reimbursement considerations for acalabrutinib. Additionally, our model is based on a previous submission to NICE and assesses the influence of each parameter in sensitivity and scenario analyses, leading to well-founded conclusions on each parameters' relevance.

However, relying on a previous NICE submission has its caveats. The use of input parameter values provided for ibrutinib may lead to biased estimates. In the ibrutinib submissions, resource use was estimated through expert opinion. Furthermore, the public report of the NICE appraisal is redacted in many places, which made it hard to implement some of the features and numbers. For example, we had to estimate the survival curves from the published data because the parameters for the curves were redacted in the report. Additionally, the utility reported was established in patients in clinical trials different than those for acalabrutinib. The lack of mature data specifically for acalabrutinib leads to larger uncertainties in cost-effectiveness estimates but is also an inherent limitation to early modelling. We provide extensive sensitivity and scenario analyses to limit these risks.

Additionally, as mentioned, survival benefit was extracted from published phase I/II (acalabrutinib) and phase III studies (ibrutinib). Several valid methods exist to estimate individual patient data from published Kaplan-Meier curves which all vary slightly [21,31,32]. We have chosen the method developed by Hoyle & Henley but others may also have been appropriate [33]. All of them represent an approximation of individual patient data (IPD) and thus have limitations. Unfortunately, IPD is not shared by the company.

While naive comparisons between trials have limitations, they are also common in the economic evaluation of pharmaceutical products. Additionally, a previous study investigated effect sizes between phase II and phase III and found that for solid malignancies, phase III studies yielded on average a 12.9% lower objective response rate [34]. Though our analysis is not in solid malignancies and the endpoints used from the trials are survival endpoints, it should be noted from this previous research that comparing a phase II with a phase III trial may not be appropriate. However, we have performed extensive sensitivity and scenario analyses on the hazard ratios provided by this comparison. To represent all possible outcomes for the survival benefit of acalabrutinib in comparison to ibrutinib, we chose the lower value for sensitivity analyses as no benefit (HR = 1.00).

Partitioned survival modelling itself has limitations, because modelling PFS and OS without modelling the underlying events may lead to over or underestimation of long-term survival. However, partitioned survival modelling is a common approach in oncology and is usually accepted by HTA bodies, as it was in the case of ibrutinib. Because survival in the PPS state was time-dependent, we required the proportion of patients entering PPS from PFS. Our method to retrieve these events was similar to the ibrutinib submission in that it included correcting for background mortality. Still, the lack of actual information on progression of patients is a limitation to partitioned survival models.

We also did not include subsequent treatments other than rituximab + idelalisib, but results show that the nature and costs of subsequent treatment are practically irrelevant for cost-effectiveness estimates. Last, we also did not include ofatumumab as a comparator. Though the benefits of acalabrutinib over ofatumumab may be different, it is likely that ibrutinib will be the primary comparator because it belongs to the same class.

Further research

It was impossible to assess all scenarios when including all parameters (>650 million scenarios). Automated analyses might provide additional insight into parameters that we excluded from scenario analysis. Finally, combining multiple disease and treatment models leads to more insight into product dynamics and lifetime cost-effectiveness. Such interactive models can accommodate the complexity of value-based pricing within different indications for multiple drugs, leading to more appropriate reimbursement mechanisms.

CONCLUSION

In this early cost-utility analysis, survival benefits of acalabrutinib do not result in a cost-effective scenario compared to ibrutinib. The relative and conflicting effects of OS, PFS, drug costs and utility during PFS show that determining cost-effectiveness of acalabrutinib without insight into all parameters complicates HTA decision making. Early assessment of cost-effectiveness of new products can support development choices and reimbursement processes through effective early dialogues between stakeholders, ultimately improving patient access.

REFERENCES

1. Byrd JC, Furman RR, Coutre SE, Burger JA, Blum KA, Coleman M, et al. Three-year follow-up of treatment-naïve and previously treated patients with CLL and SLL receiving single-agent ibrutinib. Blood. 2015;125:2497-506.
2. Byrd JC, Brown JR, O'Brien S, Barrientos JC, Kay NE, Reddy NM, et al. Ibrutinib versus Ofatumumab in Previously Treated Chronic Lymphoid Leukemia. N Engl J Med. 2014;371:213–23.
3. Claro RA de, McGinn KM, Verdun N, Lee S-L, Chiu H-J, Saber H, et al. FDA Approval: Ibrutinib for Patients with Previously Treated Mantle Cell Lymphoma and Previously Treated Chronic Lymphocytic Leukemia. Clin Cancer Res. 2015;21:3586–90.
4. Levade M, David E, Garcia C, Laurent P-A, Cadot S, Michallet A-S, et al. Ibrutinib treatment affects collagen and von Willebrand factor-dependent platelet functions. Blood. 2014;124:3991–5.
5. Fabbro SK, Smith SM, Dubovsky JA, Gru AA, Jones JA. Panniculitis in Patients Undergoing Treatment With the Bruton Tyrosine Kinase Inhibitor Ibrutinib for Lymphoid Leukemias. JAMA Oncol. 2015;1:684–6.
6. McMullen JR, Boey EJH, Ooi JYY, Seymour JF, Keating MJ, Tam CS. Ibrutinib increases the risk of atrial fibrillation, potentially through inhibition of cardiac PI3K-Akt signaling. Blood. 2014;124:3829–30.
7. Burki TK. Acalabrutinib for relapsed chronic lymphocytic leukaemia. Lancet Oncol. 2016;17:e48.
8. Patel V, Balakrishnan K, Bibikova E, Ayres M, Keating MJ, Wierda WG, et al. Comparison of Acalabrutinib, A Selective Bruton Tyrosine Kinase Inhibitor, with Ibrutinib in Chronic Lymphocytic Leukemia Cells. Clin Cancer Res. 2017;23:3734–43.
9. Wu J, Liu C, Tsui ST, Liu D. Second-generation inhibitors of Bruton tyrosine kinase. J Hematol OncolJ Hematol Oncol. 2016;9:80.
10. Wu J, Zhang M, Liu D. Acalabrutinib (ACP-196): a selective second-generation BTK inhibitor. J Hematol OncolJ Hematol Oncol. 2016;9:21.
11. Herman SEM, Montraveta A, Niemann CU, Mora-Jensen H, Gulrajani M, Krantz F, et al. The Bruton Tyrosine Kinase (BTK) Inhibitor Acalabrutinib Demonstrates Potent On-Target Effects and Efficacy in Two Mouse Models of Chronic Lymphocytic Leukemia. Clin Cancer Res. 2017;23:2831–41.
12. Smith CIE. From identification of the BTK kinase to effective management of leukemia. Oncogene. 2017;36:2045–53.
13. Byrd JC, Harrington B, O'Brien S, Jones JA, Schuh A, Devereux S, et al. Acalabrutinib (ACP-196) in Relapsed Chronic Lymphocytic Leukemia. N Engl J Med. 2016;374:323–32.
14. National Institute for Health and Care Excellence. Ibrutinib final appraisal determination. Committee papers. United Kingdom. 2017.
15. Ijzerman MJ, Steuten LMG. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. Appl Health Econ Health Policy. 2011;9:331–47.
16. IJzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. PharmacoEconomics. 2017;1–14.
17. Maignen F, Osipenko L, Pinilla-Dominguez P, Crowe E. Regulatory watch: Outcomes of early health technology assessment dialogues in medicinal product development. Nat Rev Drug Discov. 2017;16:79–79.
18. Geenen JW, Baranova EV, Asselbergs FW, de Boer A, Vreman RA, Palmer CN, et al. Early health technology assessments in pharmacogenomics: a case example in cardiovascular drugs. Pharmacogenomics. 2017;18:1143–53.
19. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation

Reporting Standards (CHEERS) statement. BMC Med. 2013;11:80.

20. National Institute for Health and Care Excellence. Blood and bone marrow cancers pathway. 2017. Accessed at https://pathways.nice.org.uk/pathways/blood-and-bone-marrow-cancers/leukaemia#path=view%3A/pathways/blood-and-bone-marrow-cancers/lymphoid-leukaemia.xml&content=view-node%3Anodes-treatment-for-relapsed-or-refractory-chronic-lymphocytic-leukaemia on Jul 25, 2017.
21. Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. BMC Med Res Methodol. 2011;11:139.
22. Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P, et al. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. N Engl J Med. 2014;370:997–1007.
23. Office for National Statistics. National life tables. 2017. Accessed at https://www.ons.gov.uk/peoplepopulation andcommunity /birthsdeaths andmarriages/lifeexpectancies/datasets/nationalli fetablesunitedkingdomreferencetables on Jul 25, 2017. [Internet]. [cited 2019 Apr 4]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables
24. British Medical Association and the Royal Pharmaceutical Society of Great Britain. The British National Formulary (BNF). Accessed at: https://www.medicinescomplete.com/mc/login.htm on Jul 25 2017. London, UK.
25. Department of Health (2016). Hospital and Community Health Services (HCHS) Pay and Price Index. Department of Health, London.
26. Department of Health. National Health Service. Reference Cost Collection: National Schedule of Reference Costs - Year 2015-16 - NHS trust and NHS foundation trusts. United Kingdom.
27. Bekelman JE, Halpern SD, Blankart CR, Bynum JP, Cohen J, Fowler R, et al. Comparison of Site of Death, Health Care Utilization, and Hospital Expenditures for Patients Dying With Cancer in 7 Developed Countries. JAMA. 2016;315:272–83.
28. Beusterien KM, Davies J, Leach M, Meiklejohn D, Grinspan JL, O'Toole A, et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. Health Qual Life Outcomes. 2010;8:50.
29. Davis S, Tappenden P, Cantrell A. A Review of Studies Examining the Relationship between Progression-Free Survival and Overall Survival in Advanced or Metastatic Cancer. [Internet]. London: National Institute for Health and Care Excellence (NICE); 2012 [cited 2019 Apr 4]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK425826/
30. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. BMJ. 2017;359:j4530.
31. Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol. 2012;12:9.
32. Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan–Meier curves. Stata J. 2017;17:786–802.
33. Wan X, Peng L, Li Y. A Review and Comparison of Methods for Recreating Individual Patient Data from Published Kaplan-Meier Survival Curves for Economic Evaluations: A Simulation Study. PLoS ONE [Internet]. 2015 [cited 2018 May 4];10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4372344/
34. Zia MI, Siu LL, Pond GR, Chen EX. Comparison of Outcomes of Phase II Studies and Subsequent Randomized Control Studies Using Identical Chemotherapeutic Regimens. J Clin Oncol. 2005;23:6982–91.
35. National Institute for Health and Care Excellence. Ibrutinib final appraisal determination. Committee papers. United Kingdom. 2017.

36. National Institute for Health and Care Excellence. Blood and bone marrow cancers pathway. 2017. Accessed at https://pathways.nice.org.uk/pathways/blood-and-bone-marrow-cancers/leukaemia#path=view%3A/pathways/blood-and-bone-marrow-cancers/lymphoid-leukaemia.xml&content=view-node%3Anodes-treatment-for-relapsed-or-refractory-chronic-lymphocytic-leukaemia on Jul 25, 2017.

37. British Medical Association and the Royal Pharmaceutical Society of Great Britain. The British National Formulary (BNF). Accessed at: https://www.medicinescomplete.com/mc/login.htm on Jul 25 2017. London, UK.

38. Department of Health. National Health Service. Reference Cost Collection: National Schedule of Reference Costs - Year 2015-16 - NHS trust and NHS foundation trusts. United Kingdom.

39. Byrd JC, Furman RR, Coutre SE, Burger JA, Blum KA, Coleman M, et al. Three-year follow-up of treatment-naïve and previously treated patients with CLL and SLL receiving single-agent ibrutinib. Blood. 2015;125:2497-506.

APPENDICES

Appendix 1: Survival estimation

For the estimation of the OS and PFS survival curves for ibrutinib, the Hoyle & Henley method to recreate individual patient data was used as described in Hoyle *et al* [21]. First, we extracted survival data from the Kaplan-Meier curve reported in Byrd *et al* [2]. We can input the number at risk and survival at each time point into the Excel file provided by Hoyle & Henley. This Excel file then approximates data on censoring and event times. Via the supplied R code of Hoyle & Henley we can then fit parametric survival models to the recreated data.

These models include an exponential, Weibull, lognormal and a loglogistic curve. Table 1 shows the intercept and ln(scale) for each of these models.

With these data, we can calculate survival at each time point, thus also allowing for extrapolation beyond the observed data. The relative goodness-of-fit for each curve is expressed in the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), where lower values indicate better fit. Table 2 shows the AIC and BIC for the plotted ibrutinib curves, showing that for PFS the lognormal distribution shows the best fit to the KM curve, whereas for OS this is the Weibull distribution.

Table 1. Derived parameters for ibrutinib survival curves.

Parameter		Exponential	Weibull	Log normal	Log-logistic
Ibrutinib					
PFS	Intercept	4.096	4.307	3.929	3.830
	Ln(scale)		0.198	0.619	0.003
OS	Intercept	4.599	5.032	4.492	4.385
	Ln(scale)		0.263	0.634	0.001

Table 2. AIC and BIC for extrapolated ibrutinib curves.

		Exponential	Weibull	Lognormal	Loglogistic
PFS	AIC	316.7630	315.0908	313.0206	315.0640
	BIC	321.2884	319.6161	317.5460	319.5893
OS	AIC	217.7135	215.8702	216.0610	217.7807
	BIC	222.1525	220.3092	220.5000	222.2198

For acalabrutinib, survival was extracted from the phase I/II study. It showed a Kaplan-Meier curve for progression free survival (PFS) and the text described overall survival (OS) (one person died during follow-up, at 13 months) [13]. According to Hoyle & Henley, survival at different time points was extracted. The data on time points for events and censoring as output of the Hoyle & Henley method were then inputted into SPSS and a Cox regression was performed for acalabrutinib OS and PFS in comparison to ibrutinib OS and PFS.

The outcomes of the Cox regression are presented in table 3 and 4.

Hazard ratios (HR's) of 0.479 (95% confidence interval (CI) 0.230 – 0.998) and 0.391 (95% CI 0.141 – 1.081) were found for PFS and OS, respectively. Note that this is not a valid final measure of PFS and OS, as data are preliminary and incomparable. Therefore, we only use these estimates to define a range, with the maximum benefit representing these HR's and the minimum benefit representing no effect (HR = 1.00).

These hazard ratios were then applied to find the survival for acalabrutinib simply by calculating $S(t)_{acalabrutinib} = S(t)_{ibrutinib}$^(hazard ratio). The survival curves that were derived via these methods are provided in figure 1 and 2.

To select a curve, AIC and BIC criteria were assessed and physiological plausibility was investigated.

Looking at AIC and BIC, for PFS the lognormal distribution shows the best fit while for OS this is the Weibull distribution, though the differences between the goodness-of-fit of the curves is relatively small. When looking at the extrapolated part of the curve, beyond the observed data, the exponential curves show the best fit, because they have the least people surviving after 30 years (400 cycles). This is physiologically the most plausible scenario. Considering that

Table 3. SPSS results for the Cox regression on PFS between acalabrutinib and ibrutinib.

							95% CI for Exp(B)	
PFS	**B**	**SE**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Treatment	-0.735	0.374	3.865	1	0.049	0.479	0.230	0.998

Table 4, SPSS results for the Cox regression on OS between acalabrutinib and ibrutinib.

							95% CI for Exp(B)	
OS	**B**	**SE**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Treatment	-0.939	0.519	3.278	1	0.070	0.391	0.141	1.081

the exponential curves were also used in the ibrutinib submission, we are confident in selecting these for further analysis [35].

For easy sensitivity and scenario analysis, acalabrutinib intercept values were calculated for the exponential curve from the survival found by applying the hazard ratio. The found intercept values were 5.045 for PFS and 5.808 for OS.

Figures 3 – 6 show all the curves, including the published curves, for acalabrutinib and ibrutinib.

It is clear that the tail of the curves for acalabrutinib OS do not match the observed data very well, however, the observed curve is based on very limited data, and hence, the parametric curves may still be reasonable estimates given the overall uncertainty about OS.

For the subsequent treatment, a similar approach was followed. Via the Hoyle & Henley method, a Weibull curve was estimates from published data for survival on rituximab + idelalisib [22]. The parameters for this curve are presented in table 5 and the curve is presented in figure 7. Figure 7 includes two curves for sensitivity analysis where both the intercept and the scale are varied by 10%.

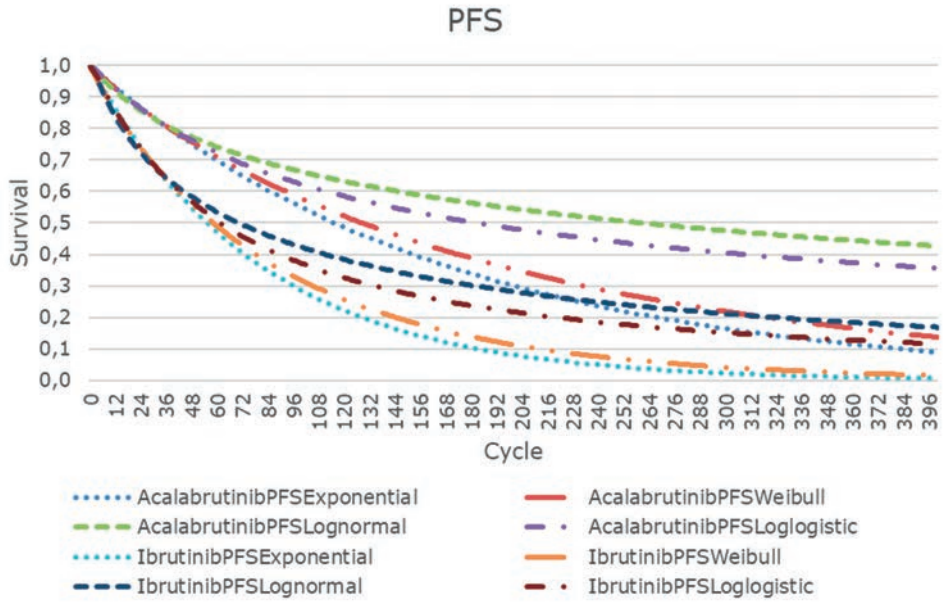


Figure 1. Plotted survival curves for acalabrutinib and ibrutinib progression-free survival (PFS). The lower curves are for ibrutinib.

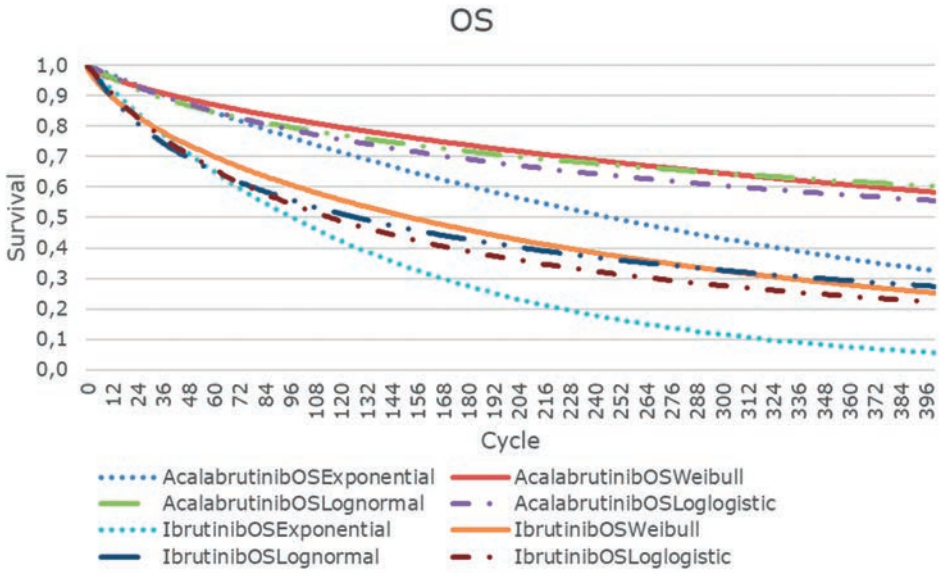


Figure 2. Plotted survival curves for acalabrutinib and ibrutinib overall survival (OS). The lower curves are for ibrutinib.

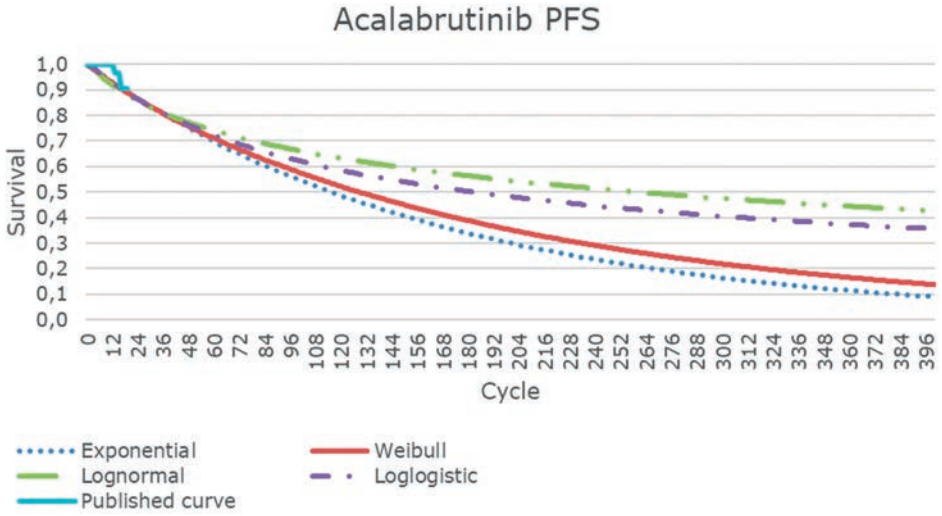


Figure 3. Curves for acalabrutinib progression-free survival (PFS), including the published curve.

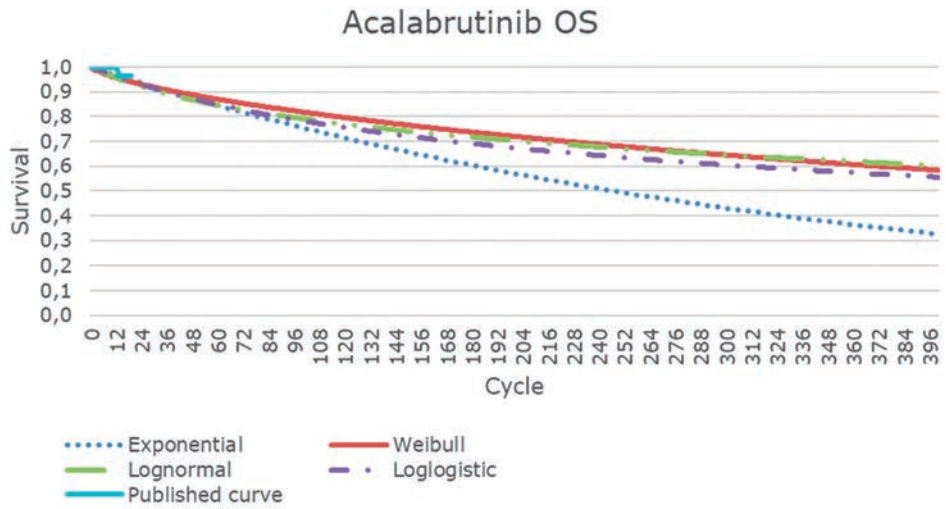


Figure 4. Curves for acalabrutinib overall survival (OS), including a curve based on published data.

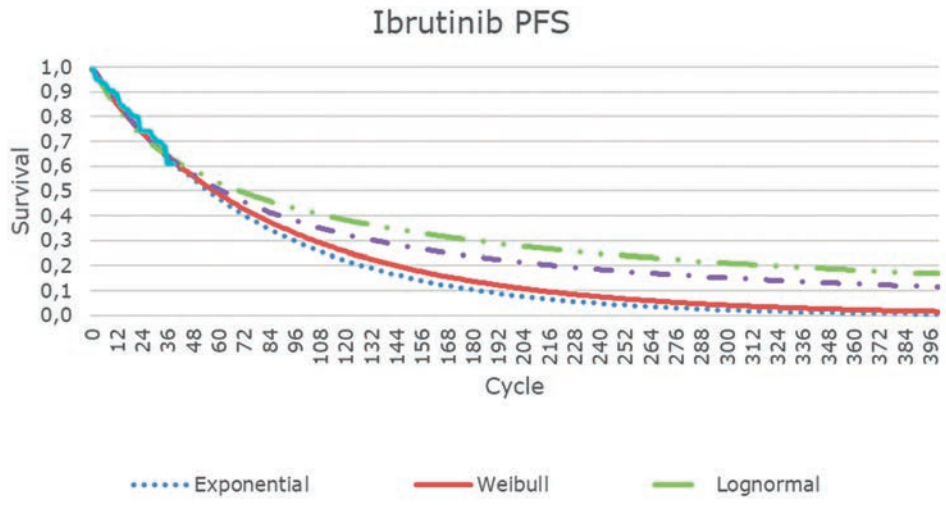


Figure 5. Curves for ibrutinib progression-free survival (PFS), including the published curve.

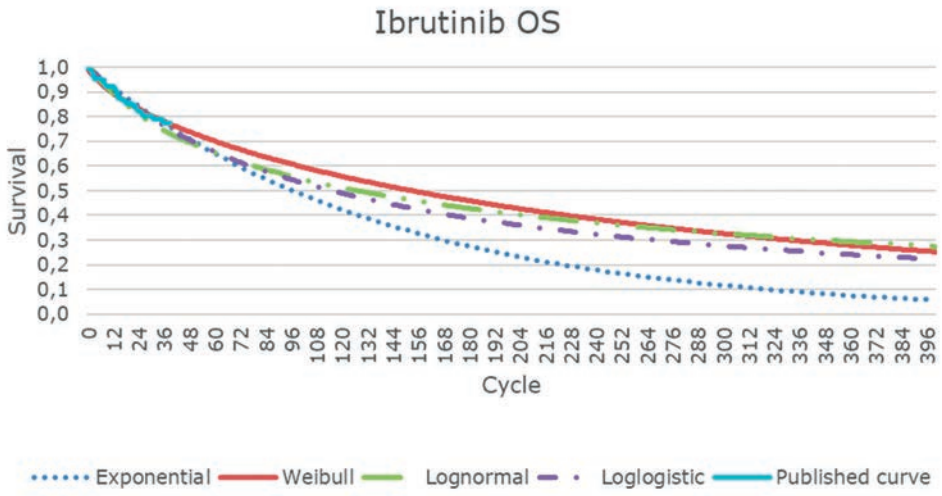


Figure 6. Curves for ibrutinib overall survival (OS), including the published curve.

Table 5. Parameters for the Weibull curve of subsequent treatment.

Subsequent treatment	Base Case	Minimum	Maximum
Intercept	3.05	2.74	3.35
Ln(scale)	-0.46	-0.56	-0.36

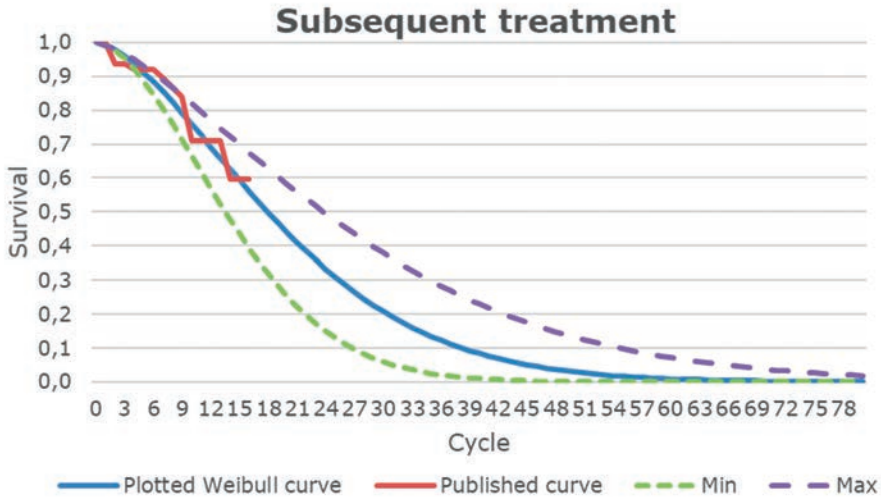


Figure 7. Base case and minimum and maximum curves for subsequent treatment with rituximab + idelalisib.

Appendix 2: Drug treatment specifications and costs calculations

Ibrutinib is administered as 420mg/day (3 capsules) until disease progression or until no longer tolerated by the patient. Acalabrutinib is given as 200 mg/day (2 capsules). Subsequent treatment exists of rituximab and idelalisib. Rituximab is given during six cycles of four weeks according to the NICE guideline for CLL, with an initial dose of 375 mg/m^2 and subsequent doses of 500 mg/m^2. Idelalisib is administered until disease progression or death in a dose of 150 mg twice daily [36]. Dosing intensity for all chronic treatments is assumed equal at 94.8%. No correction was applied for rituximab.

An overview of treatment costs is provided in table 6. Acalabrutinib unit costs are assumed equal to ibrutinib in the base case and calculated by multiplying the unit costs of ibrutinib treatment to the use per day and dividing this by the use of acalabrutinib units per day. Drug costs come from the British National Formulary [35,37].

Costs per cycle for acalabrutinib and ibrutinib is calculated by multiplying the unit costs with the unit size, the use per day, the dosing intensity and the days per cycle. This gives a treatment cycle costs of £4069.20 for both treatments.

For idelalisib and rituximab, the calculations are a bit more complicated. Idelalisib costs per day are calculated by multiplying use per day with costs per unit and the dose intensity, giving a cost of £98.43 per day. For the first 6 cycles (one cycle in the model is exactly four weeks, thus this matches the treatment with rituximab), rituximab is added to idelalisib. Rituximab costs are calculated by multiplying the square meters body surface (1.9m^2) with the indicated dose per m^2 [35]. This is 375 mg in the first administration and 500 mg in the subsequent five administrations. For each dose, administration costs are added. These are found in the UK National Schedule of Reference costs 2015-2016 and include £383.13 for the 'Delivery of Complex Chemotherapy, Including Prolonged Infusional Treatment, at First Attendance and £328.10 for the 'Delivery of Subsequent Elements of a Chemotherapy Cycle, for first and subsequent administrations, respectively [38]. We assumed no vial sharing.

Table 6. Unit costs and sizes for modelled treatments.

Drug	Dose/concentration	Tablet or vial size	Costs per unit	Use per day
Acalabrutinib	100 mg	1	£76.65	2
Ibrutinib	140 mg	1	£51.10	3
Idelalisib	150 mg	1	£51.91	2
Rituximab	10 mg/ml	10	£174.63	N/A
Rituximab	50 mg/ml	10	£873.15	N/A

For example, the first dose includes 375*1.9 = 712.50 mg. This means one vial of 500 mg and 3 vials of 100 mg are needed. The total costs for those vials is £1397.04. Including administration costs of £383.13 gives a total cost of £1780.17.

This was repeated for all six treatment cycles giving a total cost of £12,152.17. Including the treatment of idelalisib means that average costs for each cycle during the first six cycles totalled £4,781.29. Starting at cycle 7, only costs for idelalisib are included, totalling £2,755.93 per cycle. In sensitivity analyses, the variation of the costs by -80% to +30% was only applied to treatment costs, not administration costs.

APPENDIX 3: RESOURCE USE AND STATE COSTS

Resource use was derived directly from the ibrutinib submission. The submission included an overview of the use of certain types of resources per disease state and response rate, as specified in table 7. These resources were determined through an expert panel by the manufacturer and were accepted by NICE [35].

Costs per resource unit were informed by the UK National Schedule of Reference Costs 2015-2016 [38]. The exact terminology for which the costs were applied is specified in table 8.

To calculate total costs per response and per disease state, the units were multiplied by the price, and then summed. These annual costs were then corrected for cycle duration, as is presented in table 9.

To calculate resource costs per cycle per treatment, these costs per disease and response state were multiplied by treatment response known from literature. Treatment response for ibrutinib was reported to be 84% PR, 6% CR and 10% SD. For acalabrutinib, 95% had PR and 5% had SD [2,13,39]. Resource costs per disease and response state are presented in table 10.

Other costs included are costs for adverse events and costs for death. Both are inflicted once, in the cycle they happen. For adverse events this is assumed to be the first cycle of the model. Costs for the death state are inflicted once in the cycle (when death happens), and equal the per patient costs of health care utilisation (£2,900.98) during the last 30 days of life for patients of age 65+ with any cancer reported by Bekelman *et al* [27]. Adverse event costs are calculated by

Table 7. Resource unit use as defined by an expert panel in the ibrutinib submission. PFS-CR = progression free survival, complete response; PFS-PR = partial response; PFS-SD = stable disease; PPS-ST = post progression state, subsequent treatment; PPS-BSC = best supportive care.

Resources	PFS-CR	PFS-PR	PFS-SD	PPS-ST	PPS-BSC
Full blood count	2	4	4	4	4
LDH	2	2.26	2	2	0
Lymphocyte counts	3.5	7	3.5	3.2	0
Chest X-Ray	0	1	2	2	0
Bone marrow exam	0	1	1	0	0
Haematologist visit	2.26	3	4.5	4	4.9
Inpatient visit	0.66	2	2	2	1
Nurse Home visit	1.5	2.64	3	2	4
Full blood transfusion	0	1	2	2	2
Platelet transfusion	0	1	0	0	0
Biopsy	0	0	2	2	0

Table 8. Overview of costs and their sources within the National Schedule for Reference Costs.

Resources	Code	Source	Costs (£)
Full blood count	DAPS05	Other Currencies Data	3.10
LDH	DAPS04	DIRECTLY ACCESSED PATHOLOGY SERVICES	1.18
Lymphocyte counts	DAPS05	Other Currencies Data	3.10
Chest X-Ray	DAPS02	DIRECTLY ACCESSED PATHOLOGY SERVICES	30.77
Bone marrow exam	SA33Z	OUTPATIENT PROCEDURES	266.83
Haematologist visit	WF01A	Outpatient CL - Clinical Haematologist - Non-Admitted Face to Face Attendance, Follow-Up	166.03
Inpatient visit	WH53B	Follow-Up Examination for Other Conditions, without Interventions	763.42
Nurse Home visit	NURS	COMMUNITY HEALTH SERVICES - District Nurse, Adult, Face to face - Nursing	37.98
Full blood transfusion	SA13A	OUTPATIENT PROCEDURES	225.11
Platelet transfusion	SA13A	OUTPATIENT PROCEDURES	225.11
Biopsy	SA33Z	ELECTIVE INPATIENT	1078.29

Table 9. Resource costs as calculated by multiplying unit costs with unit use. PFS-CR = progression free survival, complete response; PFS-PR = partial response; PFS-SD = stable disease; PPS-ST = post progression state, subsequent treatment; PPS-BSC = best supportive care.

Resources	PFS-CR (£)	PFS-PR (£)	PFS-SD (£)	PPS-ST (£)	PPS-BSC (£)
Full blood count	6.20	12.41	12.41	12.41	12.41
LDH	2.36	2.67	2.36	2.36	0.00
Lymphocyte counts	10.86	21.72	10.86	9.93	0.00
Chest X-Ray	0.00	30.77	61.55	61.55	0.00
Bone marrow exam	0.00	266.83	266.83	0.00	0.00
Haematologist visit	375.23	498.09	747.13	664.12	813.55
Inpatient visit	503.86	1526.85	1526.85	1526.85	763.42
Nurse Home visit	56.97	100.26	113.93	75.95	151.91
Full blood transfusion	0.00	225.11	450.22	450.22	450.22
Platelet transfusion	0.00	225.11	0.00	0.00	0.00
Biopsy	0.00	0.00	2156.58	2156.58	0.00
Annual costs	955.47	2909.81	5348.71	4959.96	2191.51
Cycle costs	73.25	223.07	410.03	380.23	168.00

Table 10. Resource costs per disease state for acalabrutinib and ibrutinib.

Resources	PFS-CR (£)	PFS-PR (£)	PFS-SD (£)	PFS total (£)	PPS-ST (£)	PPS-BSC (£)
Acalabrutinib	0.00	211.91	20.50	232.41	380.23	168.00
Ibrutinib	4.39	187.37	41.00	232.77	380.23	168.00

multiplying adverse event incidence with their costs. Grade 3 and 4 adverse event incidences are specified in table 11 and were based on clinical trials for ibrutinib and acalabrutinib and on the ibrutinib submission to NICE [2,13,35,39]. Adverse event average costs are specified in table 12. For each adverse event, multiple costs are provided in the National Schedule for Reference Costs [38], depending on disease severity or score. Per adverse event, the average is calculated by multiplying the incidence of each score of the adverse event with the costs for that type, as is shown in table 13. The codes in table 12 indicate all severity types that were included per adverse event.

The costs for adverse events per treatment are then calculated by multiplying the average costs with the incidence and summing those, as is shown in table 13.

Table 11. Incidences for adverse events for ibrutinib and acalabrutinib.

Adverse event	Ibrutinib incidence	Acalabrutinib incidence
Anaemia	5.60%	5.60%
Atrial fibrillation	6.00%	0.00%
Diarrhoea	4.60%	2.00%
Hypertension	6.20%	7.00%
Neutropenia	18.50%	15.00%
Pneumonia	10.80%	10.80%
Sepsis	1.50%	1.50%
Thrombocytopenia	5.60%	0.00%

Table 12. Overview of costs for adverse events and their sources within the National Schedule for Reference Costs.

Adverse event	Code	Name	Average costs (£)
Anaemia	SA03G-H	Haemolytic Anaemia	1,129.17
Atrial fibrillation	EB07A-E	Arrhythmia or Conduction Disorders	996.67
Diarrhoea	FZ91A-M	Non-Malignant Gastrointestinal Tract Disorders	1,492.69
Hypertension	EB04Z	Hypertension	729.87
Neutropenia	SA01G-K	Acquired Pure Red Cell Aplasia or Other Aplastic Anaemia	1,498.86
Pneumonia	DZ11K-V	Lobar, Atypical or Viral Pneumonia, with Multiple Interventions	1,904.86
Sepsis	WJ06A-J	Sepsis	2,163.51
Thrombocytopenia	SA12G-K	Thrombocytopenia	636.19

Table 13. Total adverse event costs per treatment as is calculated by multiplying the incidence with the costs per adverse event.

Adverse event	Ibru inc	Acal inc	Ibru costs (£)	Acal costs (£)
Anemia	5.60%	5.60%	63.23	63.23
Atrial fibrillation	6.00%	0.00%	59.80	0
Diarrhea	4.60%	2.00%	68.66	29.85
Hypertension	6.20%	7.00%	45.25	51.09
Neutropenia	18.50%	15.00%	277.29	224.83
Pneumonia	10.80%	10.80%	205.72	205.72
Sepsis	1.50%	1.50%	32.45	32.45
Thrombocytopenia	5.60%	0.00%	35.63	0
		Total	788.04	607.18

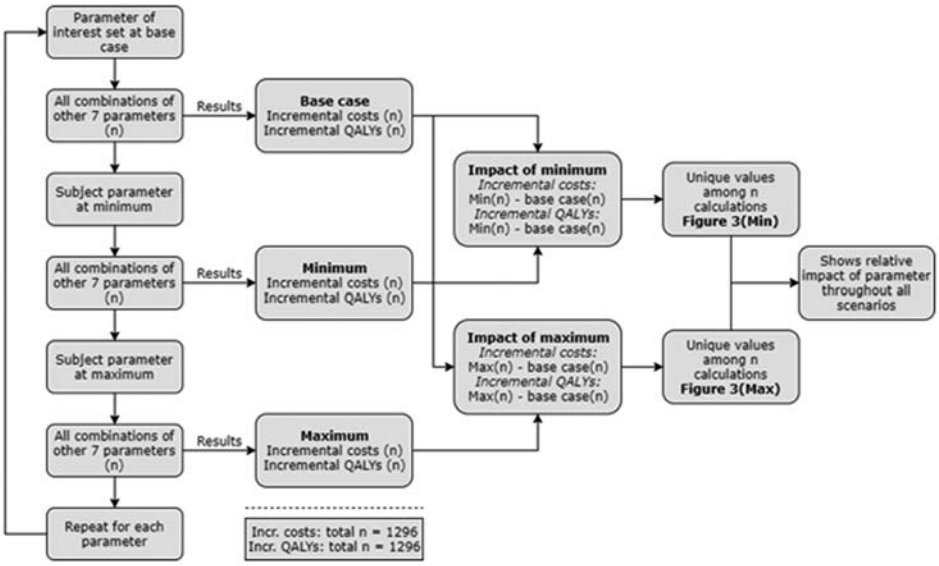
APPENDIX 4: CALCULATIONS FOR UTILITY DECREMENT

The utility decrement associated with adverse events is calculated as the utility lost per adverse event as reported in the ibrutinib submission times the incidence of that adverse event, as shown in table 14 [35]. Utility decrement for each treatment is inflicted once in the first cycle of the model.

Table 14. Incidences and utility decrements for included adverse events.

Adverse event	Ibru	Acal	Utility decrement	Ibru product	Acal product
Anemia	5.6%	5.6%	0.088	0.0049	0.0049
Atrial fibrillation	6.0%	0.0%	0.195	0.0117	0.0000
Diarrhea	4.6%	2.0%	0.088	0.0040	0.0018
Hypertension	6.2%	7.0%	0.088	0.0055	0.0062
Neutropenia	18.5%	15.0%	0.185	0.0342	0.0278
Pneumonia	10.8%	10.8%	0.195	0.0211	0.0211
Sepsis	1.5%	1.5%	0.195	0.0029	0.0029
Thrombocytopenia	5.6%	0.0%	0.123	0.0069	0.0000
			Total	0.0912	0.0646
				Difference	0.0266

APPENDIX 5: CALCULATIONS FOR THE RELATIVE IMPACT OF EACH PARAMETER VALUE.



3

Quantification of budget impact uncertainty

3.1

Accuracy of budget impact estimations and impact on patient access: a hepatitis C case study

Joost W. Geenen, Cornelis Boersma, Olaf H. Klungel, Anke M. Hövels

Published in The European Journal of Health Economics
https://doi.org/10.1007/s10198-019-01048-z

ABSTRACT

Objectives

High budget impact (BI) estimates of new drugs limit access to patients due to concerns regarding affordability and displacement effects. The accuracy and methodological quality of BI analyses are often low, potentially mis-informing reimbursement decision making. Using hepatitis C as a case study, we aim to quantify the accuracy of the BI predictions used in Dutch reimbursement decision-making and to characterise the influence of market-dynamics on actual BI.

Methods

We selected hepatitis C direct-acting antivirals (DAAs) that were introduced in the Netherlands between January 2014 and March 2018. Dutch National Health Care Institute (ZIN) BI estimates were derived from the reimbursement dossiers. Actual Dutch BI data were provided by FarmInform. BI prediction accuracy was assessed by comparing the ZIN BI estimates with the actual BI data.

Results

Actual BI, from 1 Jan 2014 to 1 March 2018, was €248 million whilst the BI estimates ranged from €388–€510 million. The latter figure represents the estimated BI for the reimbursement scenario that was adopted, implying a €275 million overestimation. Absent incorporation of timing of regulatory decisions and inadequate correction for the introduction of new products were main drivers of BI overestimation, as well as uncertainty regarding the patient population size and the impact of the final reimbursement decision.

Discussion

BI in reimbursement dossiers largely overestimated actual BI of hepatitis C DAAs. When BI analysis is performed according to existing guidelines, the resulting more accurate BI estimates may lead to better informed reimbursement decisions.

INTRODUCTION

The role of budget impact (BI) in healthcare decision-making varies across different jurisdictions as recent reviews indicate [1–4]. Germany and the USA are examples of jurisdictions that do not have a formal or informal role for budget impact in decision-making. Other countries, for example the Netherlands, France and Australia, do have guidance or even legislation on BI, but the actual role of BI or the impact on decision-making remains rather informal and moreover politically driven [1–4]. On the other end of the spectrum, England has one of the best-defined systems with a clear role for BI in healthcare decision making [2,3]. In general, however, there is an informal role for BI and its contribution to reimbursement decisions often remains unclear. As a result of that, the role of BI in decision-making remains an important topic for debate [1,5,6]. In particular, the growing attention for healthcare and pharmaceutical expenditures in combination with price negotiation mechanisms increasingly raises questions about the role of cost-effectiveness (CE).

Whilst the role of BI is often unclear in reimbursement decision-making, there are ample examples where BI did play a significant role in either the reimbursement decisions or where high BI estimates resulted in restricted reimbursement for a specific patient population [6–11]. Recently, the introduction of new, very effective but high priced Direct-Acting Antivirals (DAA) in Hepatitis C sparked worldwide affordability concerns and access restrictions [7,8]. Also in oncology, patients have limited access to many high-priced products due to concerns regarding affordability as a result of high BI [9–11].

Especially in the hepatitis C case, the cost-effectiveness of the innovations were generally regarded as positive and medical need was high [12–16]. The future will bring new products with potentially high short-term BI, which could spark further BI-guided restrictions and will call for further deliberation of the role of affordability in the political and societal debate and as such of relevant meaning in healthcare decision-making [17,18]. Therefore, clarity on the role and hierarchy of CE vs BI will not only be of interest but seems to become very important in informing reimbursement decisions.

Unfortunately, the (methodological) quality and accuracy of BI analysis does not seem to match the proven scientific rigor of CEAs [6,19,20]. A review by Van de Vooren et al. reports that (methodological) quality of many published BI analyses is poor [21]. Furthermore, Broder et al. and Cha et al. illustrate that the accuracy of BI predictions is regarded as low [22,23]. These observations, in light of the increased debate on drug prices, growing interest in price negotiation, BI of pharmaceuticals as part of healthcare budgets (e.g. Hospital), and therefore burden to societies, warrant questions about whether BI is being used properly and what the extent of influence is on patient access.

In this paper, the accuracy and role of BI in reimbursement decisions is investigated by assessing the life cycle of Hepatitis C DAAs in the Netherlands. This case was selected as there were concerns for an extremely high BI (up to €1.78 billion). The final reimbursement decision of Sofosbuvir (Sovaldi), the first DAA, resulted in restricting treatment to the most critically ill whilst this seems rather irrational from a cost-effectiveness perspective and was likely triggered by other elements like price, BI considerations and affordability discussions [12,13,24–28].

The aim of this hepatitis C case study is twofold: First, we aim to quantify the accuracy of the BI predictions used for informing the Dutch reimbursement decisions. Second, we attempt to characterise the influence of market-dynamics on actual BI and the way these are implemented in the BI predictions. This includes, for example, timing of regulatory decisions, influence of introductions of new hepatitis C products and the influence of a restricted reimbursement decision that limits the product's indication.

METHODS

Product inclusion

We included Hepatitis C DAAs that were mainly designated a standalone option for treatment of hepatitis C according to the EASL guidelines, thereby not considering co-treatment with ribavirin and/or pegylated interferon [29–32]. We subsequently excluded products that were not introduced or not used in the Netherlands in the period from 1 Jan 2014 to 1 March 2018. Lack of use or introduction was based on a publicly available national drug information system (GIP), which has national coverage and is maintained by the National Health Care Institute (ZIN) [33].

Daclatasvir (Daklinza) and simeprevir (Olysio) were excluded as these products are mainly used in combination with sofosbuvir (Sovaldi) but not as monotherapy. The sofosbuvir/velpatasvir/voxilaprevir (Vosevi) combination was not introduced and is thus excluded. Sovaldi, sofosbuvir/ledipasvir (Harvoni), ombitasvir/paritaprevir/ritonavir (Viekirax) + dasabuvir (Exviera), sofosbuvir/velpatasvir (Epclusa), elbasvir/grazoprevir (Zepatier) and glecaprevir/pibrentasvir (Maviret) were included.

BI Data and BI estimation accuracy

The actual Dutch BI data was provided by FarmInform [34]. The population-level data of FarmInform comprises of monthly volume of all prescription drugs in the in- and outpatient setting multiplied by the respective monthly list price in the Netherlands [35]. Validity of the data is ensured as the data is crosschecked with patient-level data that is representative of the Netherlands (PHARMO) [36,37]. As DAAs target specific hepatitis C viral proteins, off-label

use of DAAs is highly unlikely and we therefore assume that all DAA BI is used for treatment of hepatitis C.

The BI estimates used to inform the reimbursement decisions of hepatitis C therapy in the Netherlands were collected from the published and publicly available ZIN reimbursement dossiers [12,13,24–27,38]. These dossiers typically project the BI for the 3 years after publication of the dossier. The ZIN BI estimation format and methodology are based on the most recent ISPOR guidelines for conducting BI analysis [39,40]. BI is based on market potential: it accounts for expected patient populations and one or more treatment regimens and associated costs [39,40]. Correction should be performed for 1st in class vs subsequent introductions by making assumptions regarding the market penetration [39]. It is also recommended to include the effects of restrictions in indication due to the eventual reimbursement decision [39].

The treatment regimens or subpopulations that are mentioned in the reimbursement dossiers are based on (combinations of) METAVIR score, genotype, IFN or ribavirin co-medication and prior treatment experience. From the dossiers, estimated BI, population size and average treatment costs were recorded, as well as the subpopulations and the aforementioned characteristics these estimations were based on. The BI prediction accuracy was then assessed by comparing the ZIN BI estimates with the real world actual BI data for all included products.

Treatment indication and resulting access

As there was a potential for significant budget impact, the Sovaldi reimbursement decision stated treatment was to be restricted to more severely ill patients [28]. In order to investigate the effect and extent of this reimbursement restriction and the development of access when DAAs without restrictions were introduced, we aimed to quantify the amount of DAA access by translating actual BI to a number of patients treated.

Number of patients treated was calculated as follows:

$$Number\ of\ patients\ treated = \frac{Budget\ Impact}{Average\ treatment\ cost\ per\ patient}$$

As BI is known from the actual BI data, average treatment cost per patient had to be established. Each product has a standard treatment duration (12 weeks for most products, 8 weeks for Maviret) that can be multiplied by the known list price to obtain the cost of treating one patient. Some subpopulations however require a longer treatment duration:

- Genotype: The hepatis C virus is classified in 6 genotypes. They differ in susceptibility to (DAA) treatment as GT 3 typically requires longer treatment [29,31,32,41,42].
- Severity of disease: More severe disease evidently warrants not only (more) immediate treatment but also longer treatment [29,31,32].

- Prior treatment: Treatment experienced patients in some cases require longer treatment [31,32].

Chronic hepatitis C disease severity is frequently categorised using the well-validated METAVIR scoring system [43,44]. This 5-point scale distinguishes between various stages of liver fibrosis where F0 = no fibrosis, F1 = portal fibrosis without septa, F2 = portal fibrosis with rare septa, F3 = numerous septa without cirrhosis, F4 = cirrhosis [43,44]. For clarity, we do not consider extrahepatic complications of hepatitis C and thus solely reflect disease severity by means of METAVIR score.

EASL guidelines on treatment of particular METAVIR scores changed particularly:

- EASL 2014 & 2015: All patients with chronic liver disease related to HCV should be considered for therapy. Treatment should be prioritised in patients with METAVIR score F3 and F4. Treatment is justified in patients with METAVIR score F2. The timing and nature of therapy for patients with METAVIR score F0 + F1 debatable, and informed deferral can be considered. [29,30].
- EASL 2016: All patients with chronic liver disease related to HCV must be considered for therapy. Treatment must be considered without delay in patients with METAVIR score F2 – F4 [31].
- EASL 2018: All patients with HCV infection should be treated. Treatment must be considered without delay in patients with METAVIR score F2 – F4 [32].

The influence of these factors on treatment duration changed over time as the leading European Association for the study of the Liver (EASL) hepatitis C guidelines changed and new DAAs were introduced [29–32]. Appendix 1 summarises the major exceptions regarding treatment duration for various subpopulations.

Mean treatment costs per product

As, in for example the case of Harvoni, METAVIR stage F4 indicates a longer treatment duration, a larger proportion of patients with stage F4 would increase average treatment costs. This is of particular interest as the Harvoni and Viekirax + Exviera, dossiers specifically address 3 national BI scenarios based on only treating patients with F4 + F3 (scenario A), F4 – F2 (scenario B) and F4 – F0 (scenario C) [12,26]. Average treatment costs thus differ per scenario. In order to be able to adjust average treatment costs to different scenarios, the ZIN BI calculations had to be recreated so that the influence of different populations could be assessed. Average treatment costs per patient were solely recreated using the assumptions and data from the respective reimbursement dossiers.

In the reimbursement dossiers, the following assumptions for the Dutch setting were made: The genotype distribution is 49% GT1, 10% GT2, 29% GT3 and 11% GT4, GT5 and GT6 are very rare in the Netherlands [13,41,45]. Sovaldi treatment regimens for GT2 and GT3 are IFN

free whilst for GT1, GT4 – 6 30% of patient will be treated with an IFN free regimen [13]. The METAVIR distribution is assumed to be 24.9% F0, 26% F1, 16.1% F2, 16.8% F3, 16.2% F4 [12]. ZIN states that the total number of chronic HCV patients, thus from F0 – F4 and GT1 – 4, in The Netherlands is between 2000 and 3000 (29). With a population of 16.9 million at the time of publication, this implies a prevalence of 0.012% – 0.018% [46].

The reimbursement dossiers of Sovaldi, Harvoni and Viekirax + Exviera provided detailed insights into the types of patients receiving specific treatment durations, costs and various calculations [12,13,26]. For these products we were therefore able to calculate scenario/ population dependent average treatment costs.

For Epclusa, Zepatier and Maviret, only a short reimbursement report with rudimentary budget impact prediction was published [24,25,27]. These BI estimations lacked detailed assessments of treatment duration per subpopulation. In our analysis, we therefore assumed the following:

- Maviret: We assume that all patients are treatment naïve as we have no valid data regarding the distribution of treatment experienced vs treatment naïve patients. That implies that, according to the ZIN reimbursement dossier, all F0-F3 patients receive 8 weeks of treatment and F4 patients are treated for 12 weeks [25].
- Zepatier: The only exceptions to the standard 12-week treatment are in cases where HCZ RNA >800,000 IU/ML [27,31]. As we have no clear data on the number of patients in the Dutch setting, we disregard this exception and assume all patients are treated for 12 weeks.
- Epclusa: There are no exceptions to the standard 12-week treatment duration so we assume that all patients are treated for 12 weeks [24,31].

For our base-case analysis, we take the average of the estimated patient population size at 2500 (range 2000 – 3000). We furthermore use treatment costs of the F4 – F2 (B) scenario. Changes in list-price over time were corrected using the G-standard, a database that contains the monthly list-prices of all Dutch prescription drugs so that the correct amount of patients are calculated [35]. We did not incorporate EASL guideline changes in our analyses.

Sensitivity analyses

By means of sensitivity analysis, we investigate the scenarios proposed in the reimbursement dossier that we did not use as base-case scenario. We thus investigate the influence of a different population size and treatment cost. For population size, we take the minimum (2000) and maximum (3000) values of the range that was estimated in the reimbursement dossiers. Furthermore, we investigate the influence of average treatment cost per patient by using the F4 + F3 (A) and F4 – F0 (C) scenarios. Additionally, we perform an analysis with the absolute minimum treatment cost where we assume that no patients get extended treatment regimens for any of the included products. Finally, we assess the influence of GT3 prevalence on average treatment costs as this genotype generally warrants a longer and thus a more costly treatment.

We therefore increase GT3 prevalence from 30% to 50%, which is higher than the reported GT3 prevalence in any European country, decrease GT1 prevalence from 50% to 30% and recalculate the average treatment costs [42].

RESULTS

Accuracy of BI estimates

We compared the estimated BI from reimbursement reports with the actual BI. The estimated BI timeline starts at the date at which national reimbursement was granted except when an explicit period was mentioned. As mentioned before, the Sovaldi reimbursement decision restricted treatment to F4 + F3. The eventual reimbursement decisions for Harvoni and Viekirax + Exviera were without restrictions, meaning that scenario C (F4 – F0) had been adopted. Maviret, Epclusa and Zepatier were also reimbursed without restriction but for these products, no a priori scenarios were made. Figure 1 displays the actual BI and estimated BI for the only four products with a reported estimated BI. BI overestimation is apparent for all products with respect to their eventual reimbursement decision. For Harvoni and Viekirax + Exviera, the lower F4 + F3 scenario is closer to the actual BI than the adopted scenario.

Figure 2 combines the monthly BI of the individual products and shows the total BI of these four products. BI initially peaks with the introduction of Sovaldi to a monthly BI of about €8 million in the first quarter of 2015. Then, with the introduction of Viekirax + Exviera and Harvoni, a monthly BI of €14 million is reached and sustained for 4 months. In figure 3 we display the relative market share of the four products with an estimated BI, including the cumulative BI. The cumulative BI shows that in about 4 years, €250 million was spent on the four DAAs. The assumed market share of Harvoni (35%) and Viekirax + Exviera (35%) was, in reality, between 40-60% and <10%, respectively.

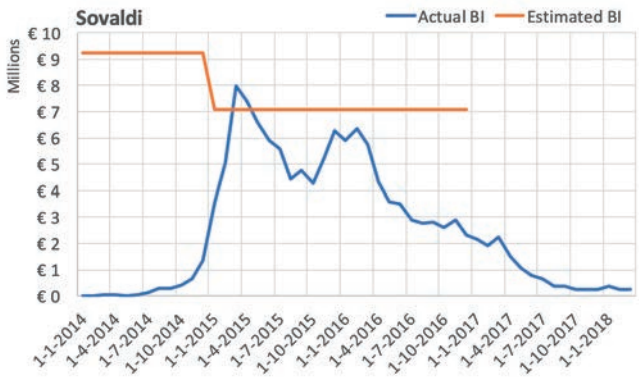


Figure 1a. Sovaldi estimated BI vs Actual BI. Values are in € millions and per month

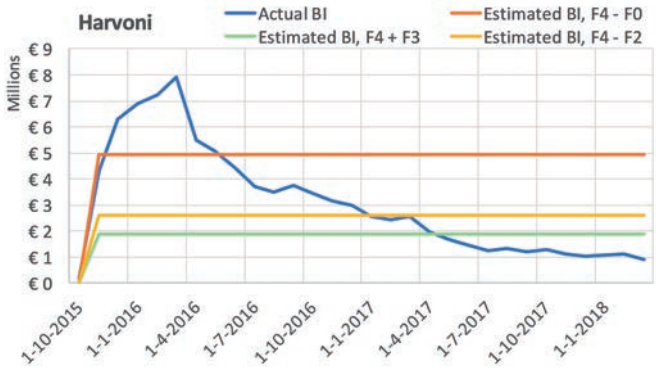


Figure 1b. Harvoni estimated BI vs Actual BI. Values are in € millions and per month

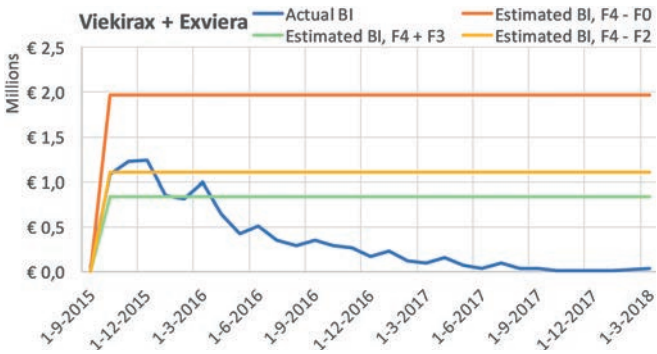


Figure 1c. Viekirax + Exviera estimated BI vs Actual BI. Values are in € millions and per month

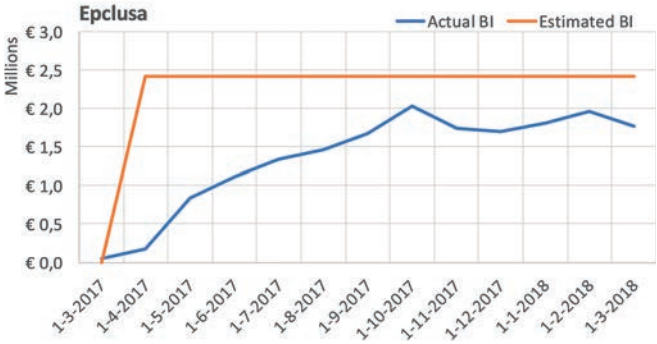


Figure 1d. Epclusa estimated BI vs Actual BI. Values are in € millions and per month

The total actual BI, estimated BI and total absolute deviation per individual product and for the total cohort are denoted in table 1. Table 2 displays the standard treatment duration per product, the eventual average treatment costs per product and, if applicable, per scenario. Even with the most modest treatment scenario (F4 + F3), treatment costs were overestimated at €153 million. When extending to the adopted and most inclusive treatment regimen (F4 – F0), total overestimation increases to €275 million. As the time between introduction of Sovaldi (1 Jan. 2014) until the last data-point (1 Mar. 2018) is slightly over four years, the annual overestimation of hepatitis C treatment costs are around €38 – €69 million.

Analysis of market dynamics

The actual number of patients treated per month is visualised in figure 4. The theoretical number of patients in different METAVIR categories indicate the extent of treatment availability to various degrees of disease severity where we assume that treatment is prioritised according to METAVIR score. On the x-axis, date of granting European Medicines Agency (EMA) Marketing Authorisation (MA) and the date of the formal initiation of national reimbursement are displayed. Note that a formal reimbursement status decision comes from the Minister of Health following an advice from ZIN.

It is evident that Sovaldi and Harvoni, at least until 2017, were most frequently used. Interestingly, for almost entire 2014, access was very limited due to absent reimbursement whilst EMA MA was granted in January. It seems that at least patients with METAVIR F4 and F3 were treated from 2015 onwards. For 2015 and 2016, treatment appears to have been extended to F2. The increase to F0 at the start of (unrestricted) reimbursement of Harvoni could be explained by the fact that treatment then became available for F2 – F0 patients.

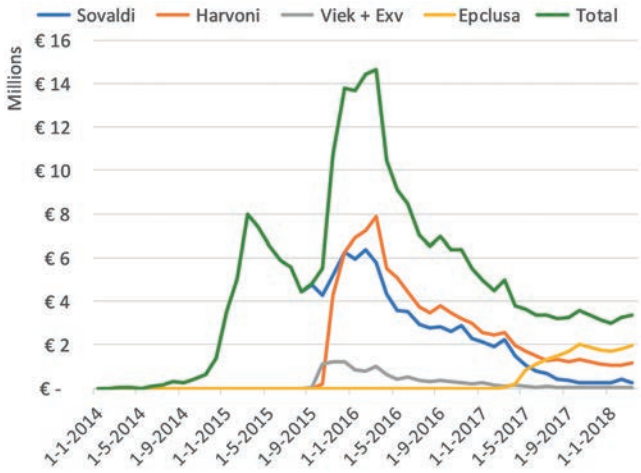


Figure 2. Total monthly BI. Total BI and Sovaldi BI overlap until 1 Sep 2015 as Sovaldi is then the only product

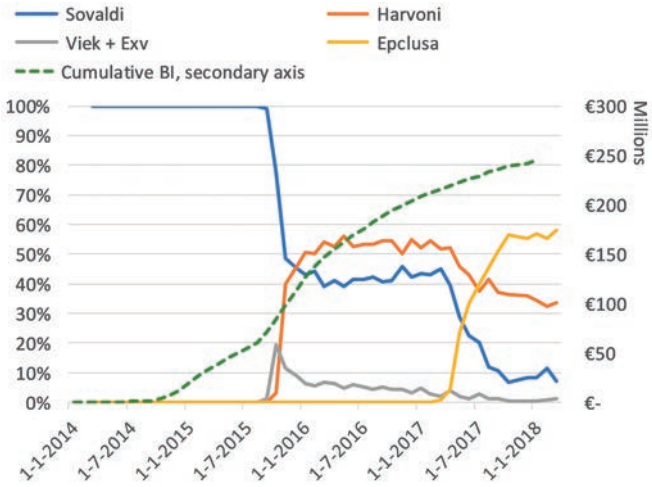


Figure 3. share of BI per product on the left vertical axis. Cumulative BI (in millions) of all four products is shown in the right vertical axis

Table 1. Overview of Actual BI, Estimated BI and the difference between actual- and estimated BI. A negative difference implies an overestimation of BI

Product (METAVIR score)	Actual BI (€)	Estimated BI (€)	Difference (€)
Sovaldi	128,692,991	281,166,336	-165,151,871
Harvoni (F4 + F3)	91,461,345	54,495,833	36,796,849
Harvoni (F4 - F2)	91,461,345	76,004,167	15,457,178
Harvoni (F4 - F0) [a]	91,461,345	143,550,000	-52,257,318
Viekirax + Exviera (F4 + F3)	10,557,104	24,166,667	-13,609,563
Viekirax + Exviera (F4 - F2)	10,557,104	32,020,833	-21,463,730
Viekirax + Exviera (F4 - F0) [a]	10,557,104	57,033,333	-46,476,230
Epclusa	17,632,950	29,000,000	-11,413,049
Total (F4 - F3)	248,344,389	388,828,836	-153,377,635
Total (F4 – F2)	248,344,389	418,191,336	-182,571,472
Total (F4 - F0)	248,344,389	510,749,669	-275,298,468

[a] The eventual reimbursement decision.

Apart from the initial peak of Harvoni and Sovaldi, broadening of the treatment population over time, as is recommended by the EASL guidelines and as is permitted by the reimbursement decisions of all products but Sovaldi, seems to be absent. This is apparent as from 2017 onwards, treated patient numbers remain stable at a level only encompassing the F3 and F4 patients. The rise in treated patients around the reimbursement date clearly confirm that in the Netherlands, access is governed by national coverage decisions and not by EMA MA.

Table 2. Average treatment cost per patient per reimbursement scenario

Product	Treatment costs (F3+F4) (€)	Treatment costs (F2-F4) (€)	Treatment costs (F0-F4) (€)	Standard treatment duration costs	Standard treatment duration (weeks)
Sovaldi	73,153	73,153	73,153	48,000	12
Harvoni	85,936	80,383	74,589 [a]	51,750	12
Viekirax + Exviera	57,959	51,444	45,172 [a]	39,400	12
Epclusa	45,999	45,999	45,999	45,999	12
Zepatier	41,397	41,397	41,397	41,397	12
Maviret	30,666	30,666	30,666	30,666	8

[a] The eventual reimbursement decision.

We compared our patient estimates with the publicly available national GIP drug information system to ensure validity of our approach [33]. In appendix 2 we extracted the annual number of users per product and they are reasonably comparable with our monthly estimates as displayed in figure 4.

Sensitivity analyses

The different treatment costs used are displayed in table 2. The sensitivity analyses, shown in appendices 3 – 8, show that the estimated size of the patient population has some influence with larger total populations yielding less access for more favourable METAVIR scores as a smaller

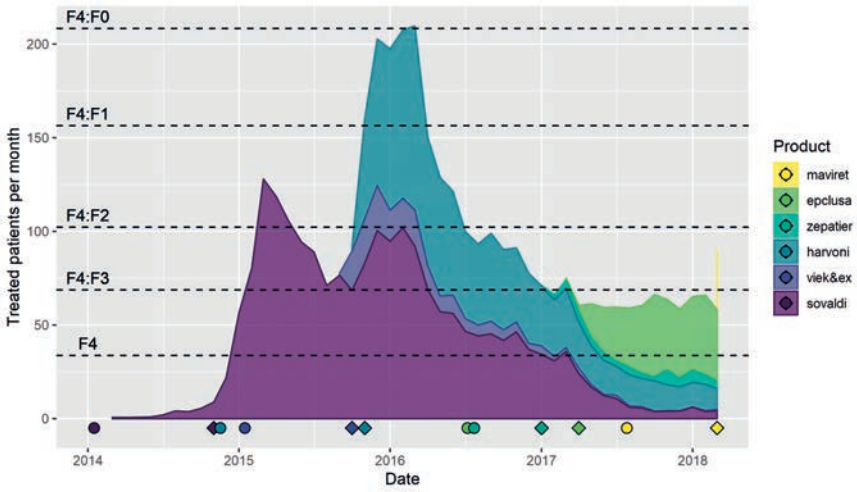


Figure 4. Treated patients over time, with monthly data. Dotted lines indicate the assumed number of patients with a specific METAVIR score. Circles indicate the date of EMA Marketing Authorisation, diamonds indicate the date of positive reimbursement decision

fraction of the population appears to be treated. Appendix 8 (treatment scenario C and base-case population size) shows the data based on the eventual reimbursement decisions for Harvoni and Viekirax & Exviera. Treatment scenarios A and C vary little compared to scenario B that was used as base-case. The minimum treatment cost scenario disregards any possibility for extended treatment durations for various subpopulations. This scenario thus results in a higher number of patients treated as treatment costs per patient were lower.

We explored the influence of GT3 prevalence on average treatment costs. For Harvoni and the various scenarios, average treatment costs increased with 7.5% - 13% whereas average Sovaldi treatment costs increased with 12-15%. Viekirax and Exviera are not recommended for GT3 and the reported treatment costs of Epclusa, Zepatier and Maviret are not influenced by genotype. Given that the GT3 prevalence increase from 30% to 50% is a 67% increase, we can conclude that average treatment costs are relatively insensitive to GT3 prevalence.

DISCUSSION

In a Dutch setting, we showed that BI estimates reported in ZIN reimbursement dossiers largely overestimated the actual BI for hepatitis C DAAs. Although the most severely ill patients did get access to the innovative hepatitis C therapies, access was initially not granted to the extent of the recommendations in the then prevailing EASL guidelines.

In the EU, the crude hepatitis C incidence is estimated to be 7.4 per 100,000 persons but with a very large spread (0.1 – 73.3) between countries, at least partly driven by varying quality of surveillance systems and data completeness [47–49]. ZIN dossiers as well as studies by Iyengar et al., Cornberg et al., and Saraswat et al. report larger potential eligible populations (22,000 – 28,000, 0.14% - 0.17%) than those that ZIN actually used for the BI calculations (2000 – 3000, 0.012% - 0.018%) [12,48–50]. The former population, all those with chronic hepatitis C, should be treated according to the most recent EASL guideline. Interestingly, the estimates of 28,000 stated in the reimbursement dossiers are denoted as a scenario where the 'indication is broadened' to all hepatitis C patients. No further notion is given as to whether the presumable non-symptomatic patients are actually F0 patients. We can safely conclude that 1) current patient volumes have not been near this level and 2) the estimates of 2000 – 3000 are likely to be a conservative estimate.

In 2015, when Sovaldi was reimbursed, the BI and estimates of number of patients appear to be rather accurate as we see that, according to the reimbursed indication, treated patients are within the F4 – F3/F2 range. Then, with the unrestricted reimbursement of Harvoni and Viekirax + Exviera, treated patients plateau to the predicted F4 – F0 population for approximately 4 months. These observations would suggest that the ZIN estimate of 2000 – 3000 patients per year is quite accurate. From June 2016 onwards however, patient numbers decline to an F4 +

F3 level that is below the expected F4 – F0 range. If we, for now, assume that the estimate of number of patients was indeed reasonably accurate, other factors must have been responsible for the large deviations between estimated and actual BI.

A first reason for this deviation could be the inadequate implementation of timing of regulatory decisions. The Sovaldi reimbursement dossier was published on 20 May 2014 based on which ZIN formally advised the minister of health on 23 May 2014. The final reimbursement status was granted per 1 Nov 2014. The delay between advice and reimbursement could have been unforeseen. The manufacturer and/or ZIN could however have assumed that reimbursement of Sovaldi during entire 2014 (MA was 16 Jan 2014) was highly unlikely. Still, the BI estimate assumed access during the entire year. This alone contributed to a €46 million overestimation which could have been prevented. Of course, the relevancy of this overestimation can be questioned as it is common-practice to start the period of estimation from the initiation of reimbursement. Applying this logic would shift the Sovaldi 'Estimated BI' line in figure 1 to the right and would cause a nearly equal overestimation from 1-1-2017 onwards due to declined market share.

In line with the overestimation due to a declined market share, inadequate correction for the introduction of new products could be a second reason. The Sovaldi BI estimations did not at all account for the introduction of new products in the same class whilst this was nearly inevitable as various manufacturers were in advanced stages of clinical development or regulatory approval [30,51,52]. The ISPOR BI guideline, to which ZIN refers in their own guidance on BI, states that an attempt should be made to forecast introduction of new interventions for the chosen time horizon [39]. Harvoni and Viekirax + Exviera were introduced nearly simultaneously and both were estimated to reach a market share of 35%. As our results show, the latter product only reached a small fraction of the estimated 35% whilst the former outperformed the expectations.

Of course, the patient estimates or the distribution over METAVIR scores could have been inaccurate. Also, as DAA BI was in general overestimated and patient estimates are on the low side of estimations, we should consider the possibility that other forms of access restrictions were present. Stringent reimbursement criteria or volume caps issued to hospitals by payers might have been a factor in this regard as payers in the Netherlands have gained influence [53].

Transitioning towards the role BI played in governing access for this specific case study, we like to reiterate that actual BI stayed well below the BI estimations ZIN deemed realistic. Consequently, the actual BI was also considerably lower than alternative scenarios postulated by ZIN where a 'broader DAA indication' would be adopted. This, according to the reimbursement dossiers, could theoretically lead to a BI of €1.78 billion [12]. Such a wide call for treating all viraemic hepatitis C patients, culminating in the EASL's 2018 guideline recommendation, has not been apparent in our data. Our data did however show that access is strictly governed by a positive reimbursement decision as a products' BI is very low before it is reimbursed.

The reimbursement decision was, on average, taken 258 days after MA whilst the reimbursement dossier was published after on average 117 days. Price negotiations were conducted for all products for which the HTA report was used as guidance. There is however no report on the role that BI played in this process and whether BI estimations, which are known and proven to be uncertain, are necessary for either the reimbursement decision or the price negotiation. Additionally, one can extend this way of thinking with a debate on whether the 258 days of access restriction is worthwhile. This especially in light of the fact that DAAs are generally considered cost-effective [12–16].

The implications of over- and underestimations differ for various stakeholders. Patients and manufacturers of the specified products, would probably incur no real loss due to an initial underestimation. It could potentially even facilitate the reimbursement process whereas the contrary could be true for an initial overestimation. For payers however, an underestimation could be more troublesome than an overestimation as the former could cause direct and measurable budgetary deficits. We have no evidence to support or quantify these statements, but it is known that payers and manufacturers have confidential price negotiations where their BI estimates together with other parameters serve some purpose for bargaining [54].

If there is a more aligned and agreed estimate of patient numbers, as the basis of a BIA, we believe that more accurate BI estimates are achievable. As is stated in the most recent BIA guideline, it is important to include treatment dynamics including new introductions and displacement effects as well as pricing dynamics to provide more accurate and meaningful BI estimates. This would allow for a more prominent role and value of BI in the decision-making process for reimbursement of different types of treatments (e.g. chronic and one-off).

Our study has various strengths. First, we used real world data to assess access using validated and monthly updated data. Our data covers in- and outpatient dispensing data, is irrespective of the healthcare provider or insurance company and therefore captures the entire Dutch DAA access data.

Second, we made an accurate representation of average treatment cost per patient using the distribution of several subpopulations. As our sensitivity analysis shows, not including patients with longer than typical treatment durations have a profound influence on outcomes.

Our study also has some limitations. First, several assumptions underlie population size estimates and the distribution of subgroup characteristics. We therefore crosschecked our estimates with the GIP-database which reports comparable figures [33]. We furthermore aimed to illustrate the influence of population size and treatment costs on outcomes by means of sensitivity analysis.

Second, our Dutch scenario induces limitations regarding generalisability. Of course, the Dutch healthcare and reimbursement system is not directly comparable to others but the, albeit imperfect, method of BI estimation is rather similar [21–23,39,39].

Third, confidential discounts or rebates are not included in this analysis. In the Netherlands, the general outcome of price negotiations is published but the actual discount remains confidential. A study by Morgan et al. indicated that, in ten high-income countries (including the Netherlands), discounts are common and confidential discounts are most frequently in the 20 -29% range but can also be substantially higher (>60%) [55]. We thus know that the actual BI, as costs to society, are lower than presented here. Yet, we based our analyses on BI estimations that disregard potential discounts and rebates. Lack of inclusion of pricing agreements is therefore not a concern.

CONCLUSION

The BI estimates published by ZIN provide a substantial overestimation of the actual BI with a deviation of between €153 - €275 million. The number of treated patients remains low, especially in light of the much higher incidence of viraemic hepatitis C and the most recent EASL guideline recommending treatment for this entire population. Underlying patient number that were used for BI estimates seem to be at least somewhat overestimated but are probably not the sole cause of BI overestimations. Differences could potentially be caused by inadequate correction for (timing of) regulatory decisions, reimbursement for a limited indication and insufficient incorporation of the introduction of new products. These market dynamics are, to varying extent, unanticipated but could and should at least partly be corrected for. When BIA is performed according to existing guidelines, the resulting more accurate BI estimates can lead to better informed reimbursement decisions. Currently, it is unclear how the BI estimates informed the reimbursement decision and if different decisions would have been if more accurate BI estimates been had available. In light of increasing debate on prices, the (uncertain) role of the reimbursement dossier in confidential price negotiations and an increasing pressure on healthcare budgets in general, it is important to further develop an approach to use BI as a more integrated part of healthcare decision-making processes.

REFERENCES

1. Cohen JP, Stolk E, Niezen M. Role of Budget Impact in Drug Reimbursement Decisions. J Health Polit Policy Law. 2008;33:225–47.
2. Flume M, Bardou M, Capri S, Sola-Morales O, Cunningham D, Levin L-A, et al. Approaches to manage 'affordability' of high budget impact medicines in key EU countries. J Mark Access Health Policy [Internet]. 2018 [cited 2018 Jul 10];6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5998770/
3. Ghabri S, Mauskopf J. The use of budget impact analysis in the economic evaluation of new medicines in Australia, England, France and the United States: relationship to cost-effectiveness analysis and methodological challenges. Eur J Health Econ. 2018;19:173–5.
4. Mauskopf J, Chirila C, Masaquel C, Boye KS, Bowman L, Birt J, et al. Relationship between financial impact and coverage of drugs in australia. International Journal of Technology Assessment in Health Care. 2013;29:92–100.
5. Cohen J, Stolk E, Niezen M. The Increasingly Complex Fourth Hurdle for Pharmaceuticals. Pharmacoeconomics. 2007;25:727–34.
6. Niezen MGH, Bont A de, Busschbach JJV, Cohen JP, Stolk EA. Finding legitimacy for the role of budget impact in drug reimbursement decisions. International Journal of Technology Assessment in Health Care. 2009;25:49–55.
7. Barua S, Greenwald R, Grebely J, Dore GJ, Swan T, Taylor LE. Restrictions for Medicaid Reimbursement of Sofosbuvir for the Treatment of Hepatitis C Virus Infection in the United States. Annals of Internal Medicine. 2015;163:215.
8. Re VL, Gowda C, Urick PN, Halladay JT, Binkley A, Carbonari DM, et al. Disparities in Absolute Denial of Modern Hepatitis C Therapy by Type of Insurance. Clinical Gastroenterology and Hepatology. 2016;14:1035–43.
9. Niezen MGH, Stolk EA, Steenhoek A, Uyl-De Groot CA. Inequalities in oncology care: Economic consequences of high cost drugs. European Journal of Cancer. 2006;42:2887–92.
10. American Society of Clinical Oncology Position Statement on Addressing the Affordability of Cancer Drugs. JOP. 2017;JOP.2017.027359.
11. The Organisation for Economic Co-operation and Development (OECD). New Health Technologies: Managing Access, Value and Sustainability [Internet]. [cited 2018 Feb 7]. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/managing-new-technologies-in-health-care_9789264266438-en
12. Ministerie van Volksgezondheid W en S. GVS-advies sofosbuvir/ledipasvir (Harvoni®) bij chronische hepatitis C - Advies - Zorginstituut Nederland [Internet]. 2015 [cited 2018 Mar 26]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2015/05/18/gvs-advies-sofosbuvir-ledipasvir-harvoni-bij-chronische-hepatitis-c
13. Ministerie van Volksgezondheid W en S. GVS-advies sofosbuvir (Sovaldi®) bij chronische hepatitis C - Advies - Zorginstituut Nederland [Internet]. 2014 [cited 2018 May 3]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2014/05/20/gvs-advies-sofosbuvir-sovaldi-bij-chronische-hepatitis-c
14. Sofosbuvir for treating chronic hepatitis C | Guidance and guidelines | NICE [Internet]. [cited 2018 Sep 26]. Available from: https://www.nice.org.uk/guidance/ta330
15. Ledipasvir–sofosbuvir for treating chronic hepatitis C | Guidance and guidelines | NICE [Internet]. [cited 2018 Sep 26]. Available from: https://www.nice.org.uk/guidance/ta363/chapter/1-Guidance
16. Elbasvir–grazoprevir for treating chronic hepatitis C | Guidance and guidelines | NICE [Internet]. [cited 2018 Sep 26]. Available from: https://www.nice.org.uk/guidance/ta413/chapter/1-Recommendations
17. Schulman KA, Balu S, Reed SD. Specialty Pharmaceuticals for Hyperlipidemia — Impact

on Insurance Premiums. New England Journal of Medicine. 2015;373:1591–3.

18. Horizonscan geneesmiddelen [Internet]. [cited 2018 Jul 23]. Available from: https://www.horizonscangeneesmiddelen.nl/geneesmiddelen/galcanezumab-neurologische-aandoeningen-hoofdpijn/versie2
19. Faleiros DR, Álvares J, Almeida AM, Araújo VE de, Andrade EIG, Godman BB, et al. Budget impact analysis of medicines: updated systematic review and implications. Expert Review of Pharmacoeconomics & Outcomes Research. 2016;16:257-66.
20. Garattini L, Vooren K van de. Budget impact analysis in economic evaluation: a proposal for a clearer definition. Eur J Health Econ. 2011;12:499.
21. Vooren K van de, Duranti S, Curto A, Garattini L. A Critical Systematic Review of Budget Impact Analyses on Drugs in the EU Countries. Appl Health Econ Health Policy. 2014;12:33–40.
22. Cha M, Rifai B, Sarraf P. Pharmaceutical forecasting: throwing darts? Nat Rev Drug Discov. 2013;12:737–8.
23. Broder MS, Zambrano JM, Lee J, Marken RS. Systematic bias in predictions of new drugs' budget impact: analysis of a sample of recent US drug launches. Current Medical Research and Opinion. 2018;34:765–73.
24. Ministerie van Volksgezondheid W en S. GVS-advies sofosbuvir/velpatasvir (Epclusa®) bij chronische hepatitis C - Advies - Zorginstituut Nederland [Internet]. 2016 [cited 2018 Aug 6]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2016/11/07/gvs-advies-sofosbuvir-velpatasvir-epclusa-bij-chronische-hepatitis-c
25. Ministerie van Volksgezondheid W en S. GVS-advies glecaprevir/pibrentasvir (Maviret®) bij chronische hepatitis C - Advies - Zorginstituut Nederland [Internet]. 2017 [cited 2018 Aug 6]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2017/10/17/gvs-advies-glecaprevir-pibrentasvir-maviret-bij-chronische-hepatitis-c
26. Ministerie van Volksgezondheid W en S. GVS-advies ombitasvir/paritaprevir/ritonavir (Viekirax®) en dasabuvir (Exviera®) bij chronische hepatitis C - Advies - Zorginstituut Nederland [Internet]. 2015 [cited 2018 May 1]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2015/06/22/gvs-advies-ombitasvir-paritaprevir-ritonavir-en-dasabuvir-viekirax-en-exviera-bij-chronische-hepatitis-c
27. Ministerie van Volksgezondheid W en S. GVS-advies elbasvir/grazoprevir (Zepatier®) bij chronische hepatitis C - Advies - Zorginstituut Nederland [Internet]. 2016 [cited 2018 May 2]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2016/11/07/gvs-advies-elbasvir-grazoprevir-zepatier-bij-chronische-hepatitis-c
28. Staten-Generaal TK der. Geneesmiddelenbeleid; Brief regering; Vergoeding Hepatitis C middel Sovaldi met financieel arrangement [Internet]. 2014 [cited 2018 Oct 23]. Available from: https://zoek.officielebekendmakingen.nl/dossier/29477/kst-29477-296?resultIndex=252&sorttype=1&sortorder=4
29. EASL Recommendations on Treatment of Hepatitis C 2014. Journal of Hepatology. 2014;61:373–95.
30. European Association for Study of Liver. EASL Recommendations on Treatment of Hepatitis C 2015. J Hepatol. 2015;63:199–236.
31. EASL Recommendations on Treatment of Hepatitis C 2016. Journal of Hepatology. 2017;66:153–94.
32. EASL Recommendations on Treatment of Hepatitis C 2018 - Journal of Hepatology [Internet]. [cited 2018 May 2]. Available from: https://www.journal-of-hepatology.eu/article/S0168-8278(18)31968-8/fulltext
33. National Health Care Institute (Zorginstituut Nederland). GIPdatabank. The Drug Information System of National Health Care Institute [Internet]. [cited 2018 Aug 3].

Available from: https://www.gipdatabank.nl/infoPagina.asp?naam=English

34. English - Farminform [Internet]. [cited 2018 Apr 26]. Available from: https://www.farminform.nl/English

35. G-Standaard — Z-Index [Internet]. [cited 2018 Aug 6]. Available from: https://www.z-index.nl/g-standaard

36. PHARMO Institute for Drug Outcomes Research [Internet]. Pharmo. [cited 2018 Apr 26]. Available from: https://www.pharmo.nl/

37. Mantel-Teeuwisse AK, Goettsch WG, Klungel OH, Boer A de, Herings RMC. Long term persistence with statin treatment in daily medical practice. Heart. 2004;90:1065–6.

38. Ministerie van Volksgezondheid W en S. GVS-advies daclatasvir (Daklinza®) bij chronische hepatitis C - Advies - Zorginstituut Nederland [Internet]. 2014 [cited 2018 Aug 8]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2014/10/27/gvs-advies-daclatasvir-daklinza-bij-chronische-hepatitis-c

39. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value Health. 2014;17:5–14.

40. Ministerie van Volksgezondheid W en S. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg - Publicatie - Zorginstituut Nederland [Internet]. 2016 [cited 2018 Oct 23]. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg

41. de Vries MJ, te Rijdt B, van Nieuwkerk CMJ. Genotype distribution amongst hepatitis C patients in The Netherlands. Neth J Med. 2006;64:109–13.

42. Petruzziello A, Marigliano S, Loquercio G, Cozzolino A, Cacciapuoti C. Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. World J Gastroenterol. 2016;22:7824–40.

43. Thein H-H, Yi Q, Dore GJ, Krahn MD. Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: A meta-analysis and meta-regression. Hepatology. 2008;48:418–31.

44. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. Hepatology. 1996;24:289–93.

45. Vriend HJ, Veen MGV, Prins M, Urbanus AT, Boot HJ, Coul ELMOD. Hepatitis C virus prevalence in The Netherlands: migrants account for most infections. Epidemiology & Infection. 2013;141:1310–7.

46. StatLine - Population and population dynamics; month, quarter and year [Internet]. [cited 2018 Aug 9]. Available from: https://opendata.cbs.nl/statline/#/CBS/en/dataset/37943eng/table?ts=1533041760011

47. Hepatitis C - Annual Epidemiological Report for 2016 [Internet]. European Centre for Disease Prevention and Control. 2018 [cited 2018 Aug 8]. Available from: http://ecdc.europa.eu/en/publications-data/hepatitis-c-annual-epidemiological-report-2016

48. Saraswat V, Norris S, de Knegt RJ, Sanchez Avila JF, Sonderup M, Zuckerman E, et al. Historical epidemiology of hepatitis C virus (HCV) in select countries - volume 2. J Viral Hepat. 2015;22 Suppl 1:6–25.

49. Cornberg M, Razavi HA, Alberti A, Bernasconi E, Buti M, Cooper C, et al. A systematic review of hepatitis C virus epidemiology in Europe, Canada and Israel. Liver International. 31:30–60.

50. Iyengar S, Tay-Teo K, Vogler S, Beyer P, Wiktor S, Joncheere K de, et al. Prices, Costs, and Affordability of New Medicines for Hepatitis C in 30 Countries: An Economic Analysis. PLOS Medicine. 2016;13:e1002032.

51. Poordad F, Dieterich D. Treating hepatitis C: current standard of care and emerging direct-acting antiviral agents. Journal of Viral Hepatitis. 2012;19:449–64.

52. Welsch C, Jesudian A, Zeuzem S, Jacobson I. New direct-acting antiviral agents for the treatment of hepatitis C virus infection and perspectives. Gut. 2012;61:i36–46.

53. Maarse H, Jeurissen P, Ruwaard D. Results of the market-oriented reform in the Netherlands: a review. Health Economics, Policy and Law. 2016;11:161–78.

54. Morgan S, Daw J, Thomson P. International Best Practices For Negotiating 'Reimbursement Contracts' With Price Rebates From Pharmaceutical Companies. Health Affairs. 2013;32:771–7.

55. Morgan SG, Vogler S, Wagner AK. Payers' experiences with confidential pharmaceutical price discounts: A survey of public and statutory health systems in North America, Europe, and Australasia. Health Policy. 2017;121:354–62.

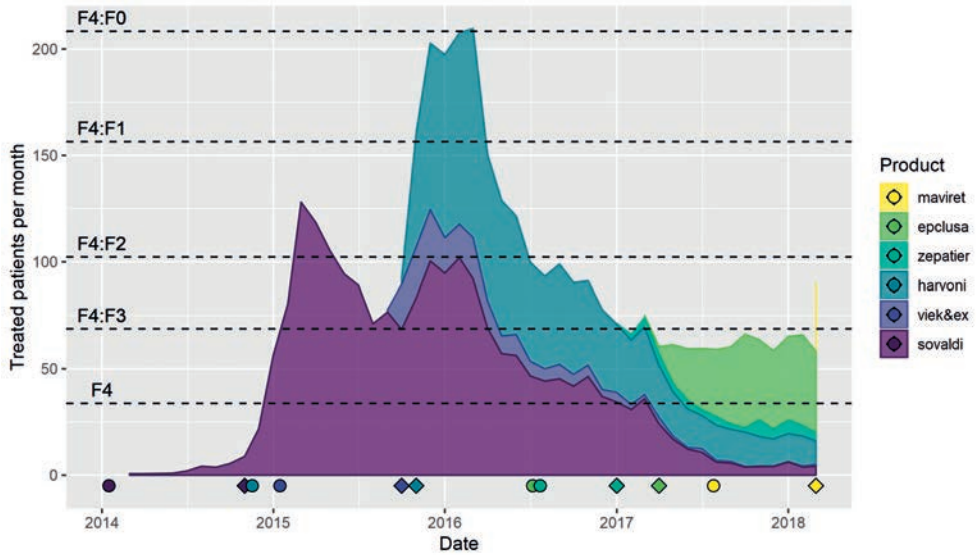
APPENDICES

Appendix 1. Overview of extended treatment durations for specific subpopulations according to various EASL guidelines

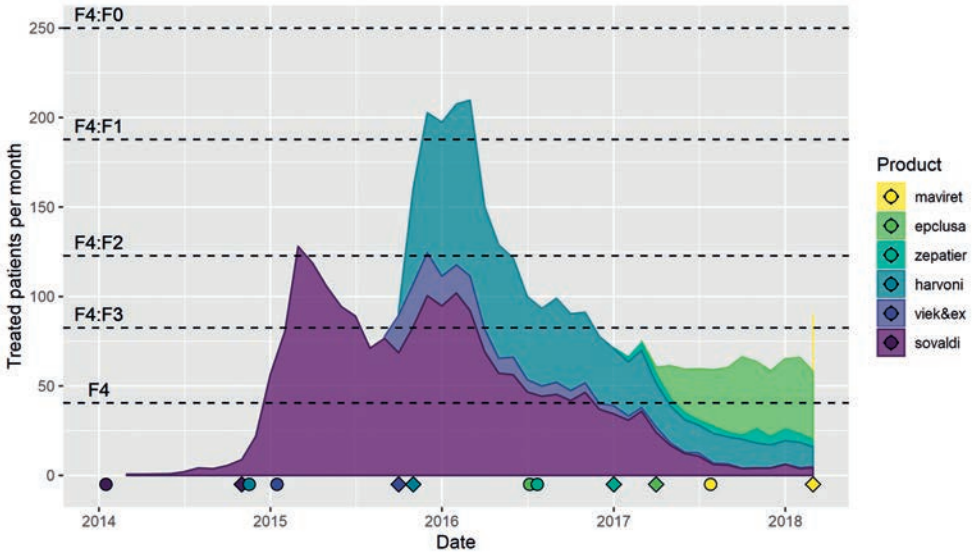
Product	Subpopulation	Treatment duration (weeks)
EASL 2014		
Sovaldi	GT1, GT4, IFN intolerant or ineligible	24 vs 12.
Sovaldi + Daklinza	GT1, Tx experienced	24 vs 12.
Sovaldi	GT2, F4, especially in Tx experienced	16 – 20 vs 12.
Sovaldi	GT3, IFN-free	24 vs 12.
Sovaldi + Daklinza	GT3, GT4, IFN-free in Tx experienced	24 vs 12.
EASL 2015		
Sovaldi	GT3, IFN-free, F0-F3	24 vs 12.
Harvoni	GT1, F0-F3	8-12 vs 12
Sovaldi	GT2, IFN-free, F4	16 – 20 vs 12.
Viekirax + Exviera	GT1a, F4	24 vs 12
Harvoni	GT1, GT4, F4, without RBV or with RBV if negative predictors of response	24 vs 12
Sovaldi + Olysio	GT1, GT4, F4, RBV-free	24 vs 12
Sovaldi + Daklinza	GT1 and GT4 RBV-free or GT3 with RBV, F4,	24 vs 12
EASL 2016		
Harvoni	GT1, Tx naïve, F0-F3	8-12 vs 12
Harvoni	GT1a, GT4, Tx experienced, F0-F3, RBV free	24 vs 12
Epclusa	GT3, Tx experienced, F0-F3, RBV free	24 vs 12
Viekirax + Exviera	GT1b, Tx naive, F0-F3	8 -12 vs 12
Zepatier	GT1, HCV RNA >800,000	16 vs 12
Zepatier	GT4, Tx experienced, HCV RNA >800,000	16 vs 12
Sovaldi + Daklinza	GT1a, GT4, Tx experienced, RBV free	24 vs 12
Harvoni	GT1a, GT4, F4, Tx experienced, RBV-free	24 vs 12
Epclusa	GT3, F4, RBV-free	24 vs 12
Viekirax + Exviera	Gt1a, F4	24 vs 12
Sovaldi + Daklinza	GT3, F4	24 vs 12
EASL 2018		
Maviret	GT3, F0-F3, Tx experienced	12 vs 8
Harvoni	GT1, F0-F3, Tx naive	8-12 vs 12
Zepatier	GT1b, F0-F2, Tx naive	8 vs 12
Viekirax + Exviera	GT1b, F0-F2, Tx naive	8 vs 12
Maviret	GT1, GT2, GT4, F4	12 vs 8
Maviret	GT3, Tx naive, F4	12 vs 8
Maviret	GT3, Tx experienced, F4	16 vs 8

Appendix 2. The Drug Information System of the National Health Care Institute data on Direct-Acting Antivirals in the Netherlands [33]. Numbers denote number of users (DDDs)

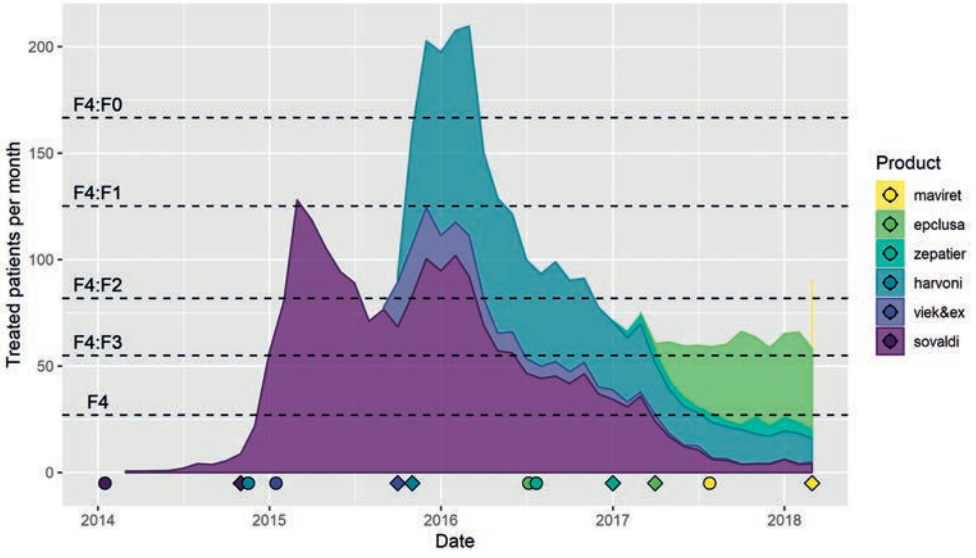
Product	2013	2014	2015	2016	2017
Sovaldi	0	77 (3826)	1359 (128,560)	1102 (90,269)	292 (22,895)
Exviera	0	0	84 (5584)	149 (11,397)	23 (1598)
Harvoni	0	0	323 (18,813)	1397 (104,540)	524 (35,338)
Viekirax	0	0	102 (6899)	176 (13,240)	28 (1906)
Zepatier	0	0	0	0	61 (4532)
Epclusa	0	0	0	0	274 (20,634)



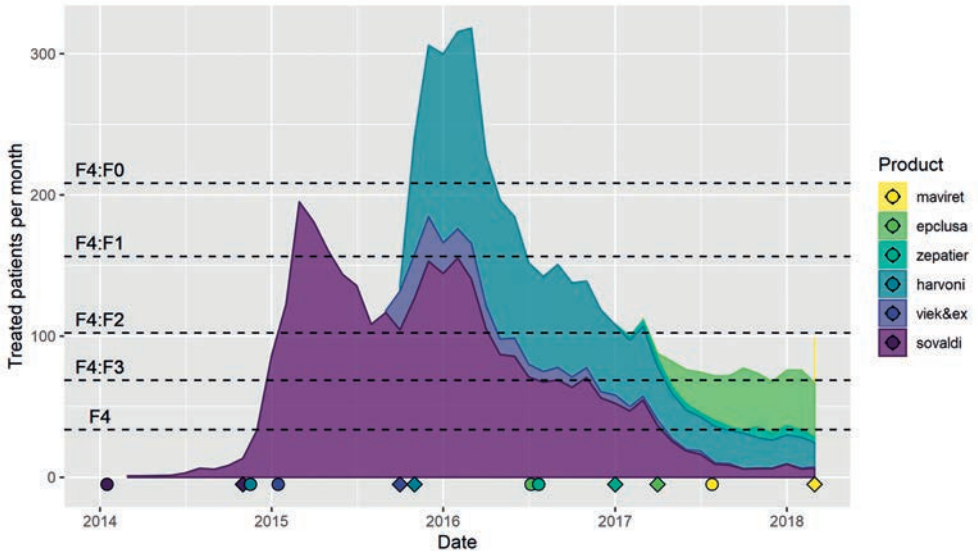
Appendix 3. Base-case (treatment costs & population size of 2500) as reference



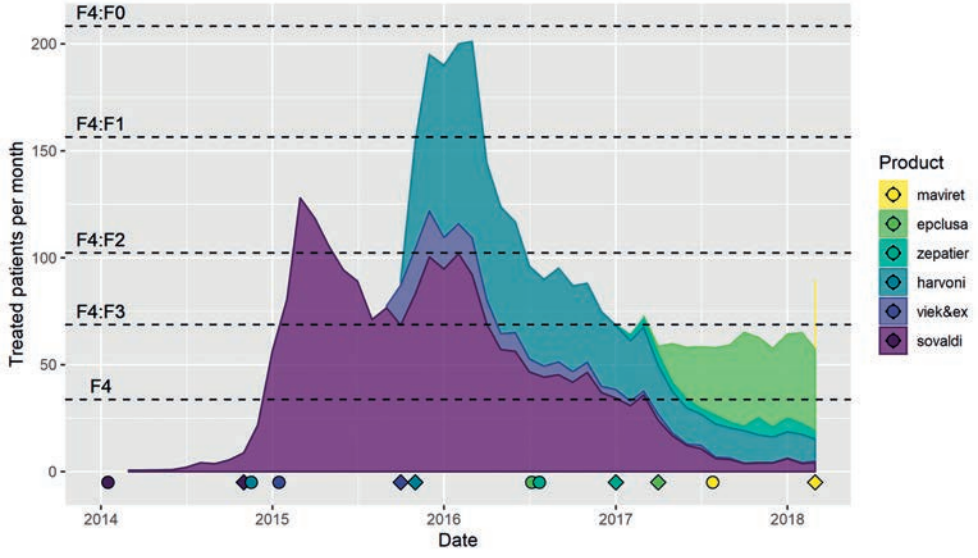
Appendix 4. Base-case treatment costs, maximum population size of 3000



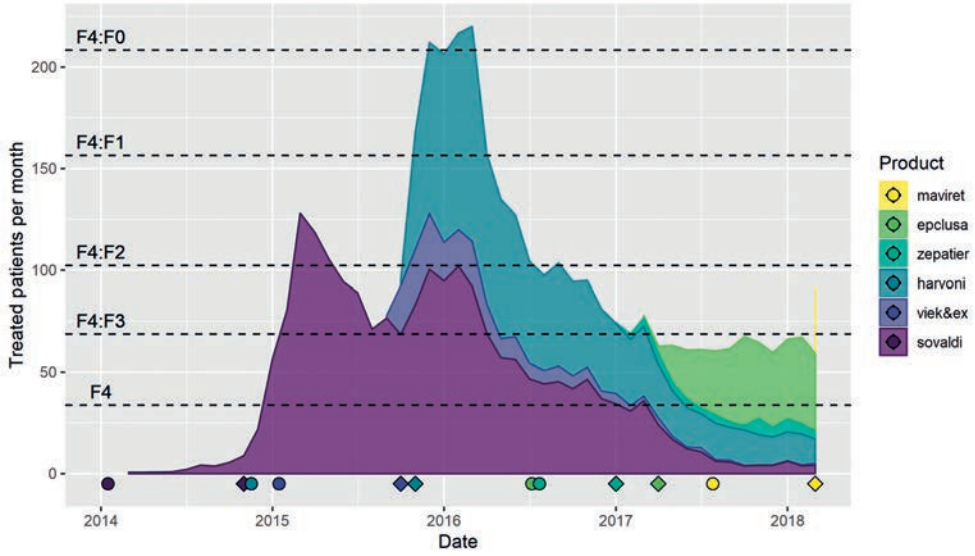
Appendix 5. Base-case treatment costs, minimum population size of 2000



Appendix 6. Minimum treatment costs, base-case population



Appendix 7. Scenario A treatment costs, base-case population



Appendix 8. Scenario C treatment costs, base-case population

3.2

Affordability of oncology drugs: accuracy of budget impact estimations

Joost W. Geenen, Mark Jut, Cornelis Boersma, Olaf H. Klungel, Anke M. Hövels

Manuscript in preparation for submission.

ABSTRACT

Objectives

In many countries, Budget Impact (BI) plays a role in assessing affordability and in informing reimbursement decisions. Various cases have shown that decision-makers have restricted access to new drugs based on high BI estimates. Evidence shows that BI estimates are often inaccurate. This is especially relevant in oncology, where high BI has led to access restrictions to potentially life-saving drugs. We aim to assess the accuracy of BI estimations used for informing access decisions on oncology drugs in the Netherlands.

Methods

Oncology products for which European Medicines Agency (EMA) Marketing Authorisation was granted between 1-1-2000 and 1-10-2017 and with a 'New Active Substance' designation, were selected. Observed BI data was provided by FarmInform. BI estimates were extracted from the reimbursement dossiers of the Dutch Healthcare Institute. Products without an estimated BI in the reimbursement dossier were excluded. We compared observed BI with BI estimates of the third (and final) year of the Dutch BI estimation period. Products with ≤6 months between publication of the reimbursement dossier and first BI record were included in the base-case analyses. Accuracy is defined as the ratio observed BI / estimated BI.

Results

Ten products were included in the base case analysis. Mean accuracy was 0.64 and observed BI deviated by more than 40% and 100% from the estimated BI for 4 and 5 products, respectively. For all products together, €141 million BI was estimated and €82 million BI was observed, resulting in a €59 million difference.

Conclusions

The findings indicate that BI estimates for oncology drugs in the Netherlands are inaccurate. They are thus an inevitable source of uncertainty to the decision-maker impacting reimbursement decisions and therefore potentially leading to restricted access for patients. The role and use of BI in reimbursement decisions for these potentially life-saving drugs should therefore be considered carefully, as well as BI estimation methodology.

INTRODUCTION

Increased spending on oncology drugs is a global problem of major concern [1,2]. The global annual spending on oncology drugs was around US$ 35 billion (€28.5 billion) in 2006, compared to around US $100 billion (€81.5 billion) in 2017. This figure is expected to rise to $150 billion (€122 billion) in 2020 [3]. The increasing number of oncology approvals for new and existing drugs, higher prices for personalised medicine and potential off-label use causes an increasing burden to health care budgets and therefore results in growing affordability [4–6].

Budget impact analyses (BIA) are performed and submitted as part of reimbursement applications in many countries to quantify the potential budget-impact (BI) [6–8]. Next to clinical evidence and cost-effectiveness, BI has an implicit – in some countries explicit – role in assessing affordability and therefore a role in informing reimbursement decisions (e.g. for orphan and specialty drugs). Various recent cases have shown that decision-makers have restricted the access to new drugs based on high BI estimates [3,7,9–12].

BI estimates used by decision-makers for informing these reimbursement decisions are however often inaccurate [10,13]. Keeping et al described that BI estimates used by the All Wales Medicines Strategy Group (AWMSG), the Welsh Health Technology Assessment (HTA) agency, deviated from observed BI with more than 40% in 80% of the cases [13]. Similar or even higher deviations were reported by others [14,15].

In the Netherlands, the estimated BI of hepatitis C drugs was nearly twice as high as the observed BI [10]. Also, among other types of information (e.g. cost-effectiveness ratios) BI estimates have contributed to decisions for initial access restrictions as well as price negotiations [10,16]. For oncology drugs, similar access restrictions are currently imposed [12,17]. The accuracy of BI estimates for oncology is however unknown, potentially hampering decision-makers in considering the uncertainty in BI outcomes. Especially in the oncology field, various schemes have been implemented by decision-makers to manage affordability and BI, such as the Cancer Drug Fund in the UK, and various managed entry schemes frequently used for oncology drugs in for example Sweden, Belgium and Italy [18–20]. In order to provide more insight in this source of uncertainty and to allow for better appraisal of BI, this study aims to quantify the accuracy of BI estimations using Dutch reimbursement decisions on oncology products as an example.

METHODS

Data source for Observed and Estimated BI

The observed BI data was provided by FarmInform [21]. This Dutch population-level data source contains the monthly volume of all in- and outpatient prescription drugs. BI is calculated by multiplying the volume by the Dutch list price including monthly price updates [22]. This

BI data is available from 1 Jan 2000 to 1 March 2018. Validity of the data was assessed by crosschecking with a patient-level data source that is representative of the Netherlands [23].

BI estimates were extracted from the publicly available reimbursement dossiers of the Dutch Healthcare Institute (ZIN) [24]. The Dutch Healthcare Institute (ZIN) is the Dutch HTA agency. ZIN conducts BIA for the first three years after product introduction based on assuming the number of eligible patients as well as volume uptake figures. The BI in the third year is typically regarded as the maximum BI. The date of publication of the dossier is denoted as index date. The estimated BI in the entire third year after the index date is thus used as estimated BI.

Product inclusion

Products were selected that obtained a European Medicines Agency (EMA) Marketing Authorization (MA) between 1-1-2000 and 1-10-2017. To select oncology products, only products were included that belonged to the L01 (antineoplastic agents) ATC category and had an initial oncology indication as specified by the European Public Assessment Report (EPAR) [25]. Biosimilars, generics and products that were not designated as a 'New Active Substance' by the EMA, were excluded. Furthermore, a ZIN BIA should be available. Finally, products without the full market data of the third year available were also excluded.

Assessment of BI estimates from reimbursement dossiers

In some cases, a BI range is given in the reimbursement dossiers. In these instances, the average of the minimum and maximum BI estimate was included. In our analysis, BI is defined as the drug cost of the new treatment so that it can be compared to the observed BI, which also solely consists of drug costs. Therefore, costs of co-medication and substitution costs were subtracted from the ZIN reimbursement dossiers to distil the BI of a specific product.

Oncology products can have multiple indications and Dutch reimbursement decisions are typically on indication level instead of on product level. Therefore, a BIA is typically tailored to a specific indication. To properly assess BI estimation accuracy, the observed BI should be solely generated by the indication for which BI is estimated.

BI can be generated (and recorded) some time before the index date as a form of early access or compassionate use scheme. Similarly, actual availability of a product (and consequent BI) can lag behind the index date. It was therefore assumed that if the first BI record in the dataset for a specific product is at most 6 months from the index date, all observed BI is attributed to the indication for which BI was estimated. Only the products that meet this criterium are included in the primary analysis. The products that did not meet this criterium were included in a secondary analysis.

When an additional reimbursement dossier (including BIA) is published within the 3-years after the index date, estimated BI is adapted to reflect this. From the date of publication of additional

indication's dossier, estimated BI (in the third year after the publication of the additional indication) is added to the estimated BI reported in the initial dossier.

Accuracy Definition

Estimation accuracy is calculated as:

$$1. \quad Accuracy = \frac{Observed\ BI}{Estimated\ BI}$$

We use this description of accuracy as primary outcome as it is frequently used and easy to interpret. When however using equation 1, under-estimations yield ratios from 1 to infinity whilst over-estimations yield ratios from 0 - 1 so averaging these ratios gives biased results [26,27]. To overcome this, a symmetric accuracy in the form of equation 2 is used as secondary outcome [26,27].

$$2. \quad Symmetric\ accuracy = e^{\left|\ln\frac{Observed\ BI}{Estimated\ BI}\right|}$$

In order to allow for a comparison with other literature, estimation accuracy is also calculated as percentage difference, defined in equation 3.

$$3. \quad percentage\ difference = \frac{Estimated\ BI - observed\ BI}{observed\ BI} * 100$$

To illustrate the uptake and estimation accuracy over time, we plot accuracy against the index date. For improved interpretability, loess smoothing is applied and is performed using R version 3.5.2 and ggplot2 version 3.1.1 [28,29].

RESULTS

Nineteen products were included of which 10 are included in the primary analysis and 9 in the secondary analysis. Table 1 displays the characteristics and outcomes of all included products. In table 2, the aggregated outcomes are presented.

Primary Analysis

Table 1 shows that the BI estimate for Nexavar was least accurate with an accuracy of 0.14 and a €24 million net over-estimation. As can be seen in table 2, the mean accuracy is 0.64, the mean percentage difference is 142% and the mean symmetric accuracy is 2,50. These results illustrate that, on average, for each €1 estimated €0.64 is observed. For each individual product,

an estimated €1 compared to a mean observed BI of €0.40 or €2.50, dependent on whether BI is over- or underestimated, respectively. The total observed BI of €82 million was over-estimated by €58.5 million.

In figure 1, the development of accuracy over time is presented. Time is defined as the months from the index date. Accuracy of 1 implies that observed BI exactly matches estimated BI. As the ZIN BI estimates are for the whole third year, the average ratio in months 24 – 36 should ideally be 1. Figure 1 shows that BI uptake is rather gradual and that products that are over-estimated at around 12 months, generally do not reach their estimated BI.

Secondary analysis

In the secondary analysis, individual accuracy (see table 1) is lower than for the products included in the primary analysis. This is however at least partly due to previously described difference in indication. For Herceptin for example, no reimbursement dossier was available for the initial (and very large) indication of HER2-positive breast cancer [30]. The Herceptin BIA, based on which the product was included in this study, only included the potential of out-patient Herceptin use [30]. The relative BI of this subgroup was relatively small and therefore impacts the results. For Avastin, high off-label use likely caused a very high observed BI relative to the indication on which the BI estimate was based.

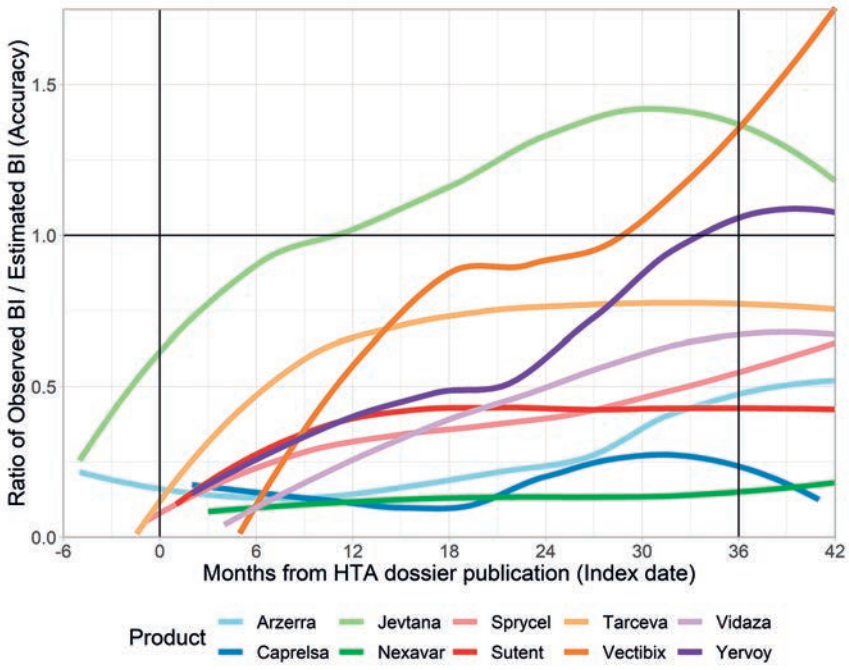


Figure 1. Estimation accuracy of the primary analysis. Primary analysis indicates that a product's first BI record was within 6 months of the index date.

Table 1. Characteristics and budget impact of 19 oncology products.

Product	Active Substance	Analysis	Estimated BI (€)	Observed BI (€)	Net difference (€)	Accuracy	Symmetric Accuracy	Percentage Difference	Index Date	First BI record
Arzerra	ofatumumab	Primary	1,210,556	446,061	-764,495	0.37	2.71	171.39	23-May-2011	1-Dec-2010
Caprelsa	vandetanib	Primary	2,500,000	626,653	-1,873,347	0.25	3.99	298.95	1-Nov-2012	1-Jan-2013
Jevtana	cabazitaxel	Primary	5,900,000	8,214,398	2,314,398	1.39	1.39	-28.17	26-Sep-2011	1-Apr-2011
Nexavar	sorafenib	Primary	28,155,960	3,803,336	-24,352,624	0.14	7.40	640.30	27-Oct-2006	1-Jan-2007
Sprycel	dasatinib	Primary	14,470,548	6,896,229	-7,574,319	0.48	2.10	109.83	6-Mar-2007	1-Feb-2007
Sutent	sunitinib	Primary	27,751,023	11,860,601	-15,890,422	0.43	2.34	133.98	27-Oct-2006	1-Nov-2006
Tarceva	erlotinib	Primary	14,573,357	11,318,537	-3,254,821	0.78	1.29	28.76	9-Mar-2006	1-Jan-2006
Vectibix	panitumumab	Primary	4,977,360	5,355,496	378,136	1.08	1.08	-7.06	28-Apr-2008	1-Aug-2008
Vidaza	azacitidine	Primary	10,881,360	6,705,395	-4,175,965	0.62	1.62	62.28	25-May-2009	1-Mar-2009
Yervoy	ipilimumab	Primary	30,240,000	26,885,500	-3,354,500	0.89	1.12	12.48	23-Jan-2012	1-Mar-2012
Alimta	pemetrexed	Secondary	21,600,000	26,742,714	5,142,714	1.24	1.24	-19.23	22-Jun-2009	1-Nov-2004
Avastin	bevacizumab	Secondary	6,504,960	46,243,910	39,738,950	7.11	7.11	-85.93	22-Oct-2007	1-Feb-2005
Erbitux	cetuximab	Secondary	10,154,100	2,010,656	-8,143,444	0.20	5.05	405.01	29-Jan-2007	1-May-2004
Herceptin	trastuzumab	Secondary	18,000,000	68,700,329	50,700,329	3.82	3.82	-73.80	26-Apr-2010	1-Oct-2000
Mabcampath	alemtuzumab	Secondary	6,237,840	351,045	-5,886,795	0.06	17.77	1676.93	28-Aug-2006	1-Jan-2002
Torisel	temsirolimus	Secondary	2,506,837	606,085	-1,900,752	0.24	4.14	313.61	25-Aug-2008	1-Jan-2008
Tyverb	lapatinib	Secondary	10,498,488	234,418	-10,264,070	0.02	44.79	4378.54	26-Jan-2009	1-Sep-2011
Velcade	bortezomib	Secondary	9,921,500	12,289,775	2,368,275	1.24	1.24	-19.27	24-Sep-2007	1-Jun-2004
Zelboraf	vemurafenib	Secondary	26,459,460	4,561,173	-21,898,287	0.17	5.80	480.10	24-Feb-2014	1-Apr-2012

Five products (Erbitux, Tyverb, Zelboraf, Torisel, Mabcampath) display a very low ratio. Of these products, Mabcampath, Zelboraf and Erbitux had early market exposure relative to their index date. Despite this potential availability of market data, their BI estimates were still rather inaccurate.

Table 2. Aggregated budget impact accuracy. Primary analysis set indicates that a product's first BI record occurred within 6 months of the index date.

Analysis (n) (First BI record <6 months from index date)	Estimated BI (€)	Observed BI (€)	Mean Accuracy	Mean Symmetric Accuracy	Mean Percentage Difference (%)	Total net difference (€)
Primary (10)	140,660,163	82,112,205	0.64	2.50	142.2	-58,547,958
Secondary (9)	111,883,185	161,740,105	1.57	10.10	784.00	49,856,920
Combined (19)	252,543,348	243,852,310	1.08	6.10	446.25	-8,691,038

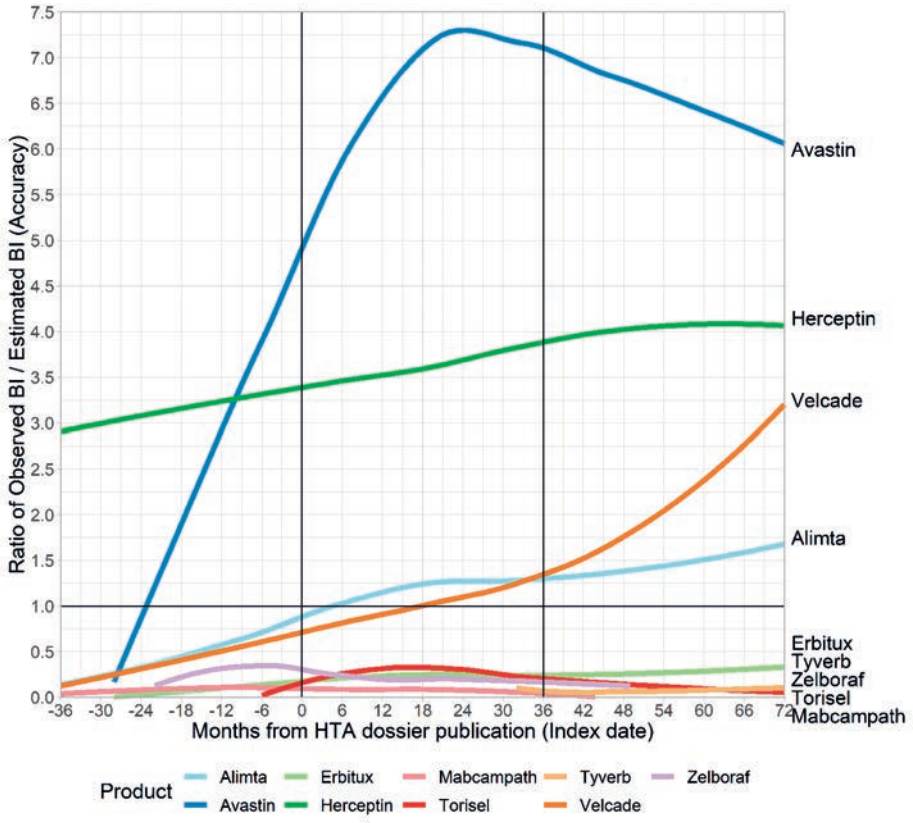


Figure 2. Estimation accuracy of budget impact in the secondary analysis. Secondary analysis indicates that a product's first BI record was separated from the index date by more than 6 months.

DISCUSSION

This study aimed to quantify the accuracy of BI estimates reported in reimbursement dossiers, using oncology products available in the Netherlands as an example. Of the 10 products included in the primary analysis, estimated BI was €140.7 million whilst only €82.1 million was observed. This difference would imply an aggregated accuracy of 0.58 whilst the mean accuracy of the individual products is 0.64. The mean symmetric accuracy and the mean percentage difference were 2.50 and 142%, respectively.

Keeping et al. assessed the accuracy of BIAs conducted by the AWMSG [13]. They reported that only 18% of estimations were within 40% of observed BI as defined by percentage difference. In the third year after introduction, Keeping et al., reported that 7 (14%) and 13 (27%) products of a total of 49 were within 40% and 100% of estimated BI. In our (primary) analysis, these figures were 4 (40%) and 5 (50%). The results are not directly comparable as Keeping and colleagues only included 2 oncology products.

Cha et al. and Broder et al. also investigated BIA accuracy but their relevance in this regard is limited as they did not target BIAs for informing reimbursement decisions [14,15]. They did however report considerable estimation inaccuracy (5.5-fold overestimation by Broder et al. and 60% of products deviated by > 40% by Cha), strengthening the evidence base that BI estimates are often inaccurate and thus pose a source of uncertainty to decision-makers that need to take BI into account.

The secondary analysis shows some interesting findings, with for example Mabcampath and Zelboraf, that show BI records preceding the index date by more than 6 months. For these products, observed BI remained well below the BI estimations reported in the eventual BIA. If the initial BI was generated by a different indication than covered in the BIA, the over-estimation relative to the indication in the BIA would be even greater. In other words, an indication that generated BI but is not covered in the BIA contributes to the observed BI, thereby increasing the value of the accuracy outcome (see equation 1). Therefore, for products with over-estimated BI (accuracy < 1; Mebcampath, Zelboraf, Tyverb, Torisel, Erbitux), it could be assumed that the reported accuracy is a best-case scenario. These five products incurred BI that deviated by >100% from estimated BI. We can therefore conclude that at least 10 of 19 (53%) products deviated by > 100%, and potentially even 14 of 19 (74%) products.

Our findings thus show that BI estimates are generally inaccurate. It is known that BI estimates are often used for confidential price negotiations between manufacturers and payers [31]. In that regard, a higher BI estimate could be an incentive and argument for a payer to attempt to negotiate a discount, although we have no evidence to support this suggestion. Our findings regarding BI estimation inaccuracy should however warrant careful consideration of

the role of BI in access decisions. Given the inaccuracy and large deviation in accuracy between products, uncertainty in BI is evident and should definitely be considered when BI informs decision-making.

Furthermore, as indicated by a review by Vooren et al., limitations in current BIA methodology and / or lack of adherence to the ISPOR budget impact guidelines might be a cause for relative lack of accuracy [32,33]. We believe that, given the importance of BI in decision making, efforts should be undertaken to improve BIA methodology and, for example, to properly address uncertainty in BI estimates.

We have some limitations to address. First, our sample size with regard to the number of products is quite small. We however chose to impose limitations on the deviation of the first BI record from the index date to assure that the comparison of the estimated BI is accurate. Our secondary analysis shows that for at least 5 more products, we can conclude that their BI assessment deviated by > 100%.

Second, we were unable to account for (off-label) indication extensions for which ZIN did not publish a BIA. As BIAs typically are conducted for expected BI over €2.5 million, we assume that indication extensions without BIA would have generated a relatively low BI and thus had little influence on our results [34,35].

CONCLUSION

The 10 products included in the primary analysis resulted in total BI estimates of €140.7 million whilst only €82.1 million was observed. For at least 53% of the included products, the observed BI differed from the estimated BI by more than 100%. These findings, combined with the large deviation in accuracy between different products, lead us to conclude that Dutch BI estimates for oncology drugs are often inaccurate. They are thus an inevitable source of uncertainty to the decision-maker impacting reimbursement decisions and therefore potentially leading to restricted access for patients. The role and use of BI in reimbursement decisions for these potentially life-saving drugs should therefore be considered carefully, as well as the need for improvements in BI estimation methodology.

ACKNOWLEDGEMENTS

We want to thank FarmInform for providing the observed BI data.

REFERENCES

1. Schumock GT, Vermeulen LC. The Rising Cost of Prescription Drugs: Causes and Solutions. Pharmacotherapy. 2017;37:9–11.
2. Prasad V, De Jesús K, Mailankody S. The high price of anticancer drugs: origins, implications, barriers, solutions. Nature Reviews Clinical Oncology. 2017;14:381–90.
3. Prasad V, De Jesús K, Mailankody S. The high price of anticancer drugs: origins, implications, barriers, solutions. Nature Reviews Clinical Oncology. 2017;14:381–90.
4. Green AK, Wood WA, Basch EM. Time to Reassess the Cancer Compendia for Off-label Drug Coverage in Oncology. JAMA. 2016;316:1541–2.
5. The Organisation for Economic Co-operation and Development (OECD). New Health Technologies: Managing Access, Value and Sustainability [Internet]. [cited 2018 Feb 7]. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/managing-new-technologies-in-health-care_9789264266438-en
6. Cohen JP, Stolk E, Niezen M. Role of Budget Impact in Drug Reimbursement Decisions. J Health Polit Policy Law. 2008;33:225–47.
7. Ghabri S, Mauskopf J. The use of budget impact analysis in the economic evaluation of new medicines in Australia, England, France and the United States: relationship to cost-effectiveness analysis and methodological challenges. Eur J Health Econ. 2018;19:173–5.
8. Niezen MGH, Bont A de, Busschbach JJV, Cohen JP, Stolk EA. Finding legitimacy for the role of budget impact in drug reimbursement decisions. International Journal of Technology Assessment in Health Care. 2009;25:49–55.
9. Barua S, Greenwald R, Grebely J, Dore GJ, Swan T, Taylor LE. Restrictions for Medicaid Reimbursement of Sofosbuvir for the Treatment of Hepatitis C Virus Infection in the United States. Annals of Internal Medicine. 2015;163:215.
10. Geenen JW, Boersma C, Klungel OH, Hövels AM. Accuracy of budget impact estimations and impact on patient access: a hepatitis C case study. Eur J Health Econ [Internet]. 2019 [cited 2019 May 2]; Available from: https://doi.org/10.1007/s10198-019-01048-z
11. Niezen MGH, Stolk EA, Steenhoek A, Uyl-De Groot CA. Inequalities in oncology care: Economic consequences of high cost drugs. European Journal of Cancer. 2006;42:2887–92.
12. American Society of Clinical Oncology Position Statement on Addressing the Affordability of Cancer Drugs. JOP. 2017;JOP.2017.027359.
13. Keeping S, Deslandes PN, Haines KE, Routledge PA. Estimated Versus Observed Expenditure Associated with Medicines Recommended by the All Wales Medicines Strategy Group. PharmacoEconomics Open [Internet]. 2019 [cited 2019 Jan 22]; Available from: https://doi.org/10.1007/s41669-019-0116-5
14. Cha M, Rifai B, Sarraf P. Pharmaceutical forecasting: throwing darts? Nature Reviews Drug Discovery. 2013;12:737–8.
15. Broder MS, Zambrano JM, Lee J, Marken RS. Systematic bias in predictions of new drugs' budget impact: analysis of a sample of recent US drug launches. Current Medical Research and Opinion. 2017;1–9.
16. Staten-Generaal TK der. Geneesmiddelenbeleid; Brief regering; Vergoeding Hepatitis C middel Sovaldi met financieel arrangement [Internet]. 2014 [cited 2018 Oct 23]. Available from: https://zoek.officielebekendmakingen.nl/dossier/29477/kst-29477-296?resultIndex=252&sorttype=1&sortorder=4
17. Ministerie van Volksgezondheid W en S. Grote budgetimpact effectief longkankermiddel Durvalumab - Nieuwsbericht - Zorginstituut Nederland [Internet]. 2019 [cited 2019 May 15]. Available from: https://www.zorginstituutnederland.nl/actueel/nieuws/2019/04/03/grote-budgetimpact-effectief-longkankermiddel-durvalumab

18. Aggarwal A, Fojo T, Chamberlain C, Davis C, Sullivan R. Do patient access schemes for high-cost cancer drugs deliver value to society?—lessons from the NHS Cancer Drugs Fund. Ann Oncol. 2017;28:1738–50.

19. Eichler H-G, Baird LG, Barker R, Bloechl-Daum B, Børlum-Kristensen F, Brown J, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. Clin Pharmacol Ther. 2015;97:234–46.

20. Pauwels K, Huys I, Vogler S, Casteels M, Simoens S. Managed Entry Agreements for Oncology Drugs: Lessons from the European Experience to Inform the Future. Front Pharmacol [Internet]. 2017 [cited 2019 Apr 4];8. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2017.00171/full

21. English - Farminform [Internet]. [cited 2018 Apr 26]. Available from: https://www.farminform.nl/English

22. G-Standaard — Z-Index [Internet]. [cited 2018 Aug 6]. Available from: https://www.z-index.nl/g-standaard

23. PHARMO Institute for Drug Outcomes Research [Internet]. Pharmo. [cited 2018 Apr 26]. Available from: https://www.pharmo.nl/

24. Ministerie van Volksgezondheid W en S. Home - Zorginstituut Nederland [Internet]. 2013 [cited 2018 May 1]. Available from: https://www.zorginstituutnederland.nl/

25. European Medicines Agency - Human medicines - European public assessment reports [Internet]. [cited 2018 May 3]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/epar_search.jsp&mid=WC0b01ac058001d125

26. Tofallis C. A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation [Internet]. Rochester, NY: Social Science Research Network; 2014 Jul. Report No.: ID 2635088. Available from: https://papers.ssrn.com/abstract=2635088

27. Tornqvist L, Vartia P, Vartia YO. How Should Relative Changes Be Measured? The American Statistician. 1985;39:43–6.

28. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: https://www.R-project.org/

29. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. Available from: http://ggplot2.org

30. Ministerie van Volksgezondheid W en S. Trastuzumab (Herceptin) bij borstkanker met tumoren die een overexpressie van HER2 vertonen - Rapport - Zorginstituut Nederland [Internet]. 2010 [cited 2018 Feb 20]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2010/04/26/trastuzumab-herceptin-bij-borstkanker-met-tumoren-die-een-overexpressie-van-her2-vertonen

31. Morgan SG, Vogler S, Wagner AK. Payers' experiences with confidential pharmaceutical price discounts: A survey of public and statutory health systems in North America, Europe, and Australasia. Health Policy. 2017;121:354–62.

32. Vooren K van de, Duranti S, Curto A, Garattini L. A Critical Systematic Review of Budget Impact Analyses on Drugs in the EU Countries. Appl Health Econ Health Policy. 2014;12:33–40.

33. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget Impact Analysis—Principles of Good Practice: Report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value in Health. 2014;17:5–14.

34. Zorginstituut Nederland. Kosteneffectiviteit in de praktijk [Internet]. 2015. Available from: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk/Kosteneffectiviteit+in+de+praktijk.pdf

35. Ministerie van Volksgezondheid W en S. Format Budget Impact Analyse GVS (voorheen kostenconsequentieraming) - Publicatie - Zorginstituut Nederland [Internet]. 2016 [cited 2018 Mar 23]. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/04/06/format-budget-impact-analyse-gvs-voorheen-kostenconsequentieraming

4

Methods for managing uncertainty

4.1

A novel method for predicting the budget impact of innovative medicines: validation study for oncolytics

Joost W. Geenen, Svetlana V. Belitser, Rick A. Vreman, Martijn van Bloois,
Olaf H. Klungel, Cornelis Boersma, Anke M. Hövels

Manuscript in preparation for submission.

ABSTRACT

Objectives

High budget impact (BI) estimations of new drugs have led to decision-making challenges potentially resulting in negative impact on patient access. However, current BI predictions are rather inaccurate and short-term. We therefore developed a new approach for BI prediction. Here, we describe the validation of our BI prediction approach using oncology drugs as a case study

Methods

We used Dutch population-level data to estimate BI, containing the products of list price multiplied by volume of all drugs from 2000 onwards. We included drugs in the antineoplastic agents ATC category which the European Medicines Agency (EMA) considered a New Active Substance and received EMA Marketing Authorisation (MA) between 2000 and 2017. A mixed-effects model was used for prediction and included tumour site, orphan-, first in class- or conditional approval designation as covariates. Data from 2000 – 2012 was the training set. BI was predicted monthly from 0-45 months after MA. Cross-validation was performed using a rolling forecasting origin with e^|Ln(observed BI/predicted BI)| as outcome.

Results

The training set and validation set included 25 and 44 products, respectively. Mean error, composed of all validation outcomes, was 2.94 (median 1.57). Errors are higher with less available data and at more future predictions. Highest errors occur without any prior data. From 10 months onwards, error remains constant.

Conclusions

We believe that, based on our validation, we have developed a valid method to predict BI. For payers or policymakers, our approach can yield a valuable addition to current BI predictions due to its ease of use, independence of indications and ability to update predictions to the most recent data.

INTRODUCTION

In recent years, the prices of new drugs, for example in oncology, increased considerably [1]. Combined with the increasing number of annual oncology approvals and expanding indications, drug treatment costs in this field have increased sharply. This has resulted in significant macro level budget impact (BI) discussions [1]. High BI estimations and the potential of a negative impact on affordability and patient access have, unsurprisingly, led to decision-making challenges and debate [2–5].

In many jurisdictions, patient access is governed by institutional payers or national reimbursement agencies [6]. When facing budgetary constraints, as is the case in for example England, (additional) spending on one drug must be covered by disinvesting in other interventions or services [7,8]. Budgetary limitations and budgeting policies cause payers or reimbursement agencies to limit access to high priced pharmaceuticals and/or products with a high BI and therefore burden to health care budgets and society [9].

It is a trend that more new drugs gain marketing approval with limited evidence packages [10]. The orphan designation and conditional approval legislation might have been successful in increasing the therapeutic options in some disease areas but it does have adverse effects on payers: much more uncertainty regarding clinical- and cost-effectiveness and BI [11–15]. The combination of high price and high uncertainty in (cost-) effectiveness as well as the potential population size and therefore budget impact, poses the greatest risk to payers or budget holders. In this study we will focus on the budget impact as source of uncertainty in reimbursement decision making. According to a review by Van de Vooren et al., many published budget impact analyses (BIA) still fail to reach an acceptable quality [16]. Many BIAs are short term (one year), quite subjective or based on expert opinion and determined by estimations of population size and eventual treatment regimen [17,18]. If the general methodological quality of BI analyses is indeed low, one would expect the predictive accuracy of these analyses to also be low.

Broder et al., who evaluated BI forecasts of US drug launches between 1 Sep 2010 and 1 Sep 2015, concluded that the average predicted BI was 5.5 times the observed BI. Cha et al. concluded that 60% of the drug forecasts were off by more than 40%. Keeping et al. recently wrote that BI estimates used by Welsh payers that were specifically produced to inform access decisions were off by more than 40% in 80% of the cases [19]. We believe that these findings illustrate that the methodological quality as well as the predictive accuracy of current BIAs can be considered as low.

Not only are these estimations insufficient in providing adequate clarity on the costs of a new drug, they also fail to quantify the uncertainty that is associated with these predictions. In other words, the current point estimates or ranges given are not based on an underlying probability

distribution and thus provide insufficient insight in the possible range of financial outcomes. Especially given the concerns regarding accuracy and methodological quality mentioned previously, insights into uncertainty surrounding BI estimates could prove to be a crucial step in increasing the use and validity of BIA. Consequently, noting that conducting (probabilistic) sensitivity analysis is now standard practice is cost-effectiveness analysis (CEA), allowing for proper sensitivity analysis in BIA could increase its validity.

Proper incorporation of (accurate) budget impact predictions in reimbursement decisions is essential for ensuring patient access and affordability [1–4]. Therefore, we developed a new approach for BI predictions using population-based drug volume data and a mixed-effects model aiming to improve prediction of future BI and quantification of uncertainty of the predicted BI. In this paper, we describe this BI prediction approach and the validation of this method using a Dutch perspective and using oncology drugs as a case study.

METHODS

We used population-level data provided by FarmInform to estimate and validate BI [20]. This data contains the monthly BI as list price multiplied by volume (generated in the in- and outpatient setting) of all prescription drugs in the Netherlands from 1 January 2000 to 1 October 2017. We denote these monthly products as BI data records. FarmInform crosschecks the data with Dutch patient-level PHARMO data to ensure generalisability [21,22].

We selected products based on being in the ATC antineoplastic agents (L01) category, having European Medicines Agency (EMA) Market Approval (MA) between 1 January 2000 and 1 October 2017 and having an initial oncology indication as specified in the EPAR [23]. We excluded biosimilars, generics and products that were not designated as a 'New Active Substance' by the EMA [23].

Data on EMA orphan designation at MA, EMA conditional approval or EMA MA under exceptional circumstances, EMA 'New Active Substance' classification and indication(s) at EMA MA were derived from the initial EMA European public assessment reports (EPAR) and European Commission Decision documents [23,24]. MA under exceptional circumstances and conditional approval were combined into one covariate denoted as 'CE'. Indications were subsequently categorised into cancer sites (e.g. breast, lung). Data on molecule type (e.g. small molecule, monoclonal antibody) were derived from EPARs. Food and Drug Administration (FDA) First in Class designation (FiC) was based on Eder et al. and FDA Novel Drug Approvals summaries [25,26]. As we chose the perspective of individual drug products and not drug-classes or patient populations, indication extensions or label changes of products were not included.

We used a mixed-effects model for prediction. Model building and validation were performed in R for Windows using the nlme package [27,28]. The dataset was split in a training set and a validation set based on the date of the monthly BI data record. The training set was used for constructing the mixed-effects model. The splitting point of the dataset for modelbuilding was set at 149 months, indicating that products with a BI data record prior to 1 May 2012 were selected as training set and products with a first BI data record after this date as validation set. Only the first 45 months of BI records per drug were included as this is the period we aim to predict, denoted as *t_max*. The duration of the period for the training set was based on a proper balance between number of products in the training set (n = 25) and an adequate number of months in the validation set to ensure capturing enough of the 45 months of data for a sufficient amount of validation set products (n = 44).

On the training set, model building was performed using forward stepwise selection (lowest Akaike Information Criterion (AIC), $p < 0.10$). Interactions with the square root of time and time to the power 1 - 6 were included as a possible step to model time dependence. Only a single time interaction per covariate was allowed.

As Shmueli stated, overfitting to training data is the biggest danger to generalisability of predictive models [29]. Moreover, it is explained that it is not required to explore the causal structure of variables as, in prediction models, predictor selection should be solely based on quality of the association between the predictor and response [29]. In order to limit risk of overfitting, we therefore did not force main effects of interactions to be included in the model. Due to right-skewed BI data, log transformed monthly BI (per lowest AIC) was selected as dependent variable. Random effects were composed of a random intercept and a random slope for time per product, based on lowest AIC. The correlation structure was defined as autoregressive with an order of 1 for time.

We then performed cross validation. Let *A* be the validation set products, *k* be a single product selected from *A* and *B* is the resulting list of products in the training set which does not include *k*. *A* is constructed by selecting all products with a first BI record after 1 May 2012. Figures 1a – 1c provide a schematic overview of the validation procedure.

We simulated the effect of the monthly addition of new data, thereby modelling the passing of time and the influence this has on prediction by using a rolling forecasting origin. Hence, we adapt the training set to include all BI records on *B* just one month prior to the date of the first BI record of *k*. *t_data* represents the number of months of data available to modelbuilding and prediction while *t_pred* indicates the month which is predicted. *t_split* governed the rolling forecast and indicates the date at which the dataset is split into training- and validation set.

The rolling forecasting origin, simulating passing time, used the following procedure: We set the initial cycle to start with zero BI records (*t_data = 0*) on *k; t_split* is set to the date of the first

BI record of *k*. The training set is constructed to include all data until *t_max* (45 months) and *t_split* on B. The first 45 months (or less if *k* has a shorter MA period) of BI are then predicted for *k* (*t_pred[1, t_max])* and compared with the observed BI data of *k*. In the next cycle, the first month of BI data will become available to the prediction model, so *t_data* = 1 and *t_split* is increased by one month. This implies that the first month of *k*'s records is added to the training set with *B,* governed by *t_split,* also advancing one month. Prediction and comparison with observed data is then performed for *k* for *t_pred [*2, *t_max]*. The sequence is repeated for *t_data[2, t_max - 1]* and all products in *A*. This yields a total of 45 + 44 + … = 1035 timepoints.

To improve robustness, a validation is also performed on a training set with *t_max = 42* which includes separate model building on this second training set. Subsequently, BI prediction validation is performed with 42 months of BI prediction. The mean of the absolute individual predictions are then calculated for all *k* with *t_data[0, t_max - 1]* and *t_pred[1, t_max]* by taking the average of these data points for the 42- and 45 *t_max* runs. This produces the final prediction results for each *k* with a specific *t_data* and *t_pred,* denoted as the prediction samples.

The previous paragraphs have outlined the role of the training set (selection of model structure) and the validation set (accuracy of predictions, given the chosen model structure). Coefficients are, unlike the traditional notion of a training set, not governed by the initial training set but are estimated for each prediction cycle, based on the available data (governed by *t_split*). We thus aim to validate a BI prediction approach which uses a fixed model structure and where the model is continuously retrained on future data. This validation approach is adopted as it represents the envisioned implementation that can adopt to patient and market dynamics.

We capped predicted BI to a minimum or maximum value in order to limit the effect of potential outliers. The maximum predicted monthly BI was determined as two times the maximum monthly BI in the total dataset. The minimum monthly BI was set at an arbitrary €5000. The influence of limiting these values is explored by means of scenario analysis.

Equation 1 describes the calculation of the prediction error.

$$1. \quad error = e^{\left|\ln \frac{Observed\ BI}{Predicted\ BI}\right|}$$

The resulting ratio is symmetric for over- and underprediction [30,31]. The purpose of this transformation is to yield ratios that have a positive sign for over- as well as underprediction so that overpredictions do not cancel out underpredictions. An error of 2 should therefore be interpreted as, in case of observed BI of €10,000, a predicted BI of €5000 or €20,000. This error was calculated for each prediction sample, yielding error samples.

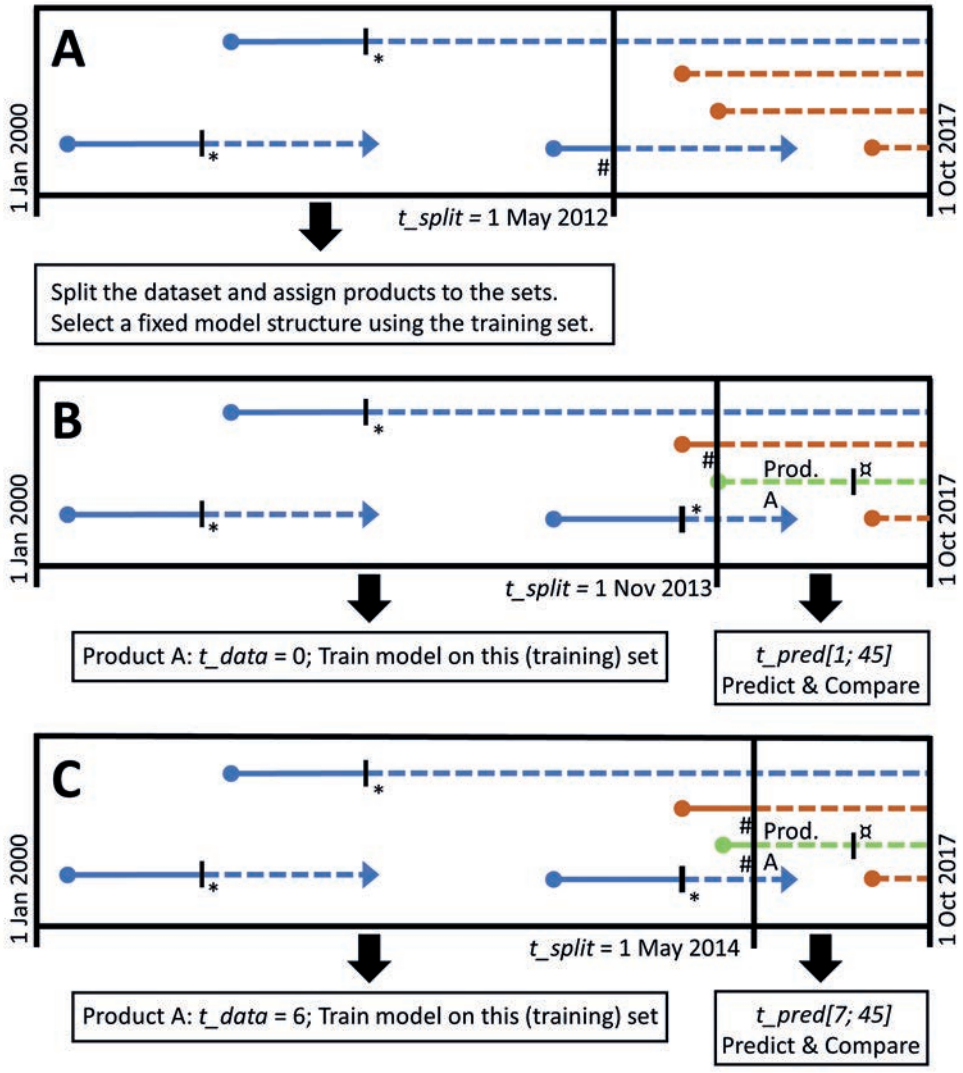


Figure 1. Schematic presentation of the role and construction of the training- and validation set, model development and the validation using a rolling forecast origin. 1a: Training set selection for model development and resulting selection of validation products. 1b: Validation of the product depicted in green with t_data = 0. The t_split = 1 Nov 2013, similar to the first date of recorded BI for this particular product. 1c: Validation of the product depicted in green with t_data = 6. The t_split = 1 May 2014. Arrows = BI record availability of a specific product; dashed line = trimmed data; solid line = data included in a training set; blue = training set products; orange = products that will be validated; green = product that, in this example, is validated; * = data cut-off based on t_max; # = data cut-off based on t_split; ¤ = the maximum value of t_pred which is identical to t_max.

Results were compiled in three ways:

1. Aggregated per *t_pred* and *t_data*: Error samples are aggregated for each point in (*t_pred, t_data*).

2. Aggregated per *t_data*: Error samples are aggregated for each *t_data*.

3. Not aggregated: Outcomes on all individual error samples.

In order to compare our results to previously published literature, we calculated the percentage of predictions that are between 40% to -40% and between 100% to - 100% of the observed BI. Per prediction, this percentage is calculated using equation 2:

$$2. \quad percentage\ difference = \frac{predicted\ BI - observed\ BI}{observed\ BI} * 100$$

For all products, we investigated whether a reimbursement dossier was published by the Dutch Healthcare Institute (ZIN), the authority that performs Health Technology Assessment (HTA) and advises the Dutch Minister of Health on reimbursement of new drugs. For the products with a reimbursement dossier, the amount of *t_data* on the date of publication of the report was recorded. Products can have *t_data* prior to publication of the dossier when it has been available for another indication or through an alternative reimbursement scheme. In our envisioned implementation, the available *t_data* just prior to publication of the dossier would be used to make an up-to date BI prediction for the product in question.

RESULTS

The training set used for modelbuilding contained 25 products with a mean of 33 months of data, 15 and 16 products had data until *t_max* of 45 and 42 months, respectively. The validation set included 44 products with an average of 27 months of data and 11 and 14 products having data until *t_max* of 45 and 42 months, respectively. This resulted in a total of 19,681 prediction- and error samples. The products included in the datasets are displayed in appendix 1.

Fixed effect selection for the 42- and 45 *t_max* models yielded the same fixed effects being time + time * CE, √(time) * Tumour site, molecule type, √(time) * FiC and √(time) * orphan designation. As random effects were not varied both model structures are identical. The final model syntax was:

[lme(fixed = log(observed BI) ~ Time + Time:CE + Molecule_type + sqrt(Time):(Orphan_status + FiC_status + Tumour site), random = ~ Time | Product, correlation = corARMA(p = 1, q = 0, form = ~ Time | Product)]

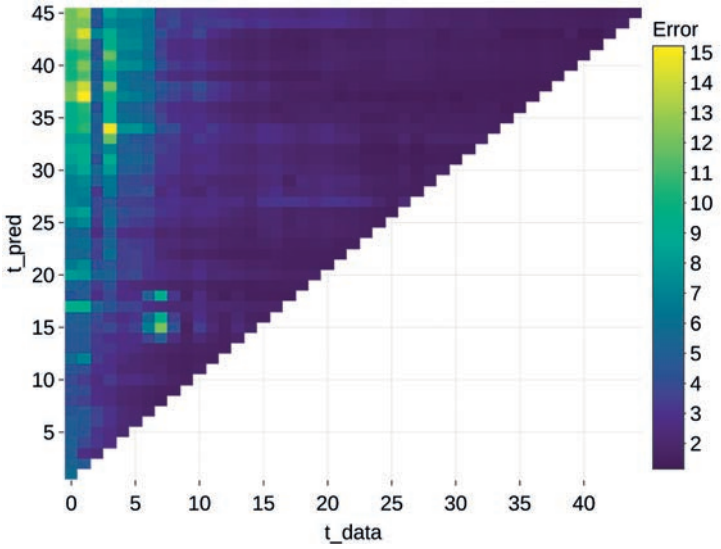


Figure 2. Mean error aggregated per t_pred and t_data. t_pred indicates the future month that is predicted and t_data indicates the amount of months of prior data which is available to the model.

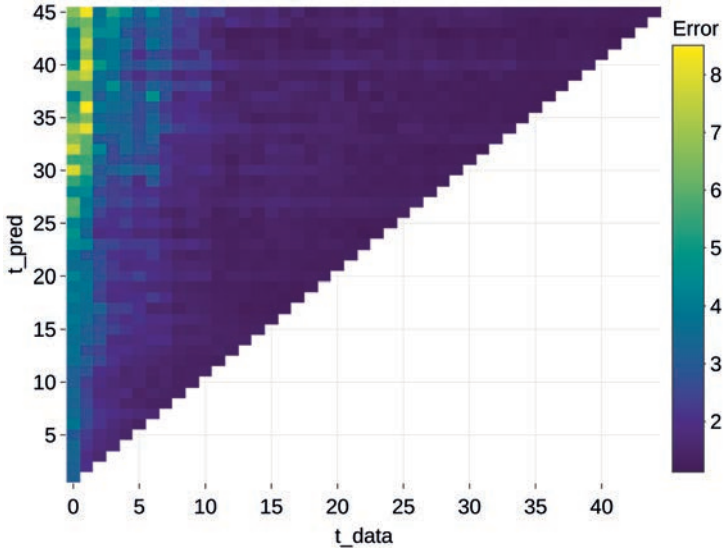


Figure 3. Median error aggregated per t_pred and t_data. t_pred indicates the future month that is predicted and t_data indicates the amount of months of prior data which is available to the model.

The results that are aggregated per t_pred and t_data are shown in figures 2 (mean) and 3 (median). These figures illustrate that the errors are higher in models with less available data (low *t_data*) and at predictions further in the future (a higher *t_pred*). The highest errors occur in the models without any prior data *(t_data* = 0) with a mean error ratio of 6.37. From *t_data* > 10, the error seems to remain constant.

In figure 4, we present the results that are aggregated per *t_data*. The errors are clearly left-skewed and significantly reduce with increasing *t_data* as established using a linear regression on the individual samples (coefficient = -0.117, se = 0.0039, $p < 0.0001$). The interquartile range (IQR) decreases with increasing t_data. Prediction performance increases substantially from increasing *t_data* from 0 to 5, from *t_data* > 10 model accuracy does not improve by adding more data.

The unaggregated mean and median error for all samples are 2.94 (Standard deviation (SD) = 5.64) and 1.57 (Interquartile range (IQR) = 1.42), respectively. The mean of Ln(observed BI/predicted BI), so without converting to absolute values, of all samples, depicted in figure 5, did significantly differ from 0 (t-test: mean = 0.057, n = 19,681, 95% CI = 0.043 ;0.070, $p < 0.0001$). In absolute terms, underprediction was significantly more likely than overprediction (exact binomial test, prob. underprediction = 0.515, 95% CI = 0.508 ; 0.522, $p < 0.0001$). Using equation 2, we calculated the percentage difference that can be compared with other literature. Of the 19,681 samples, 9,503 (48.3%) had a maximum percentage difference between 40%

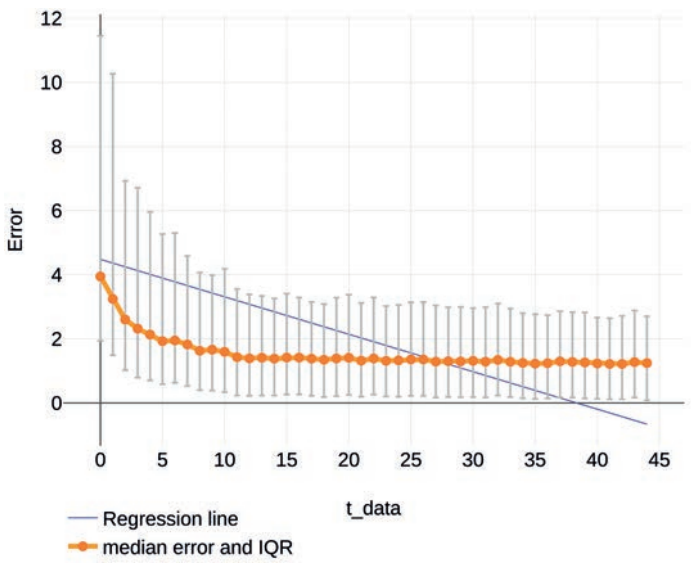


Figure 4. Median error aggregated per t_pred. Median error (orange) aggregated per t_pred, including error bars indicating the interquartile range and the regression line (blue). Coefficient = -0.096, se = 0.0035, $p<0.0001$

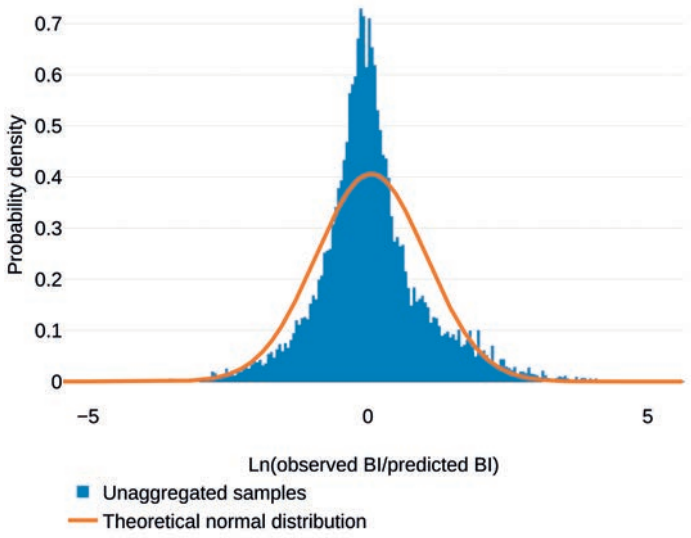


Figure 5. Histogram of the individual outcomes. Outcomes calculated as Ln(observed BI/predicted BI) (blue) and the theoretical normal distribution (orange).

Table 1. Main outcomes.

Outcome	Value
Mean error, aggregated per t_data and t_pred (SD)	3.01 (2.24)
Mean error, not aggregated (SD)	2.94 (5.63)
Median error, not aggregated (5^{th}, 25^{th},75^{th} and 95^{th} percentile)	1.57 (1.04, 1.21, 2.63, 8.59)

and -40%. For the 100% to -100% range, this number was 15,915 (80.9%). Table 1 summarises the main outcomes.

In the training set, 3 products did not have a reimbursement dossier and 9 products did not have *t_data* at the time of publication of the dossier. For the 13 training set products with *t_data*, the median and mean months of *t_data* were 22 and 29.8 (sd = 32.1), respectively. In the validation set, 28 products did not have a dossier and 10 products (with a dossier) had 0 *t_data*. The 6 products with a dossier and *t_data* > 0, had a median and mean *t_data* of 12.5 and 16.5 (sd = 11.8), respectively.

Scenario analyses

We explored the influence on limiting predictions to a minimum (< €5000) and maximum (> 2 times maximum recorded BI) on the outcomes by adopting scenarios where 1) only minimum values were adjusted, 2) only maximum values were adjusted and 3) where no predictions

were adjusted. In the base-case analysis, on a total of 19,681 samples, 930 minimum- and 44 maximum values were adjusted. The outcomes are presented in appendices 2 – 4.

These results indicate that not limiting minimum values has a profound negative influence on model performance as, presented in supplementary table 3, mean unaggregated error increases to 8.00 with a very large sd of 358.74. On the contrary, not limiting maximum only has a minor impact, as table 1 and supplementary table 2 yield nearly identical results. As only the more extreme values are concerned, it is logical that the median figures (appendices 2 - 4) are very similar to the median base case results.

DISCUSSION

Our prediction model was constructed using a mixed effects approach with a training set of 25 oncology products. The validation of this prediction model using a training set of 44 products yielded 19,681 samples and resulted in an overall mean error of 2.94. This error is higher when available data is more limited and when predicting further into the future. The decline in error with increasing available data seems to halt, at a median error of ± 1.5, around 5 – 10 months of data. This indicates that relatively accurate predictions can be generated with 5 – 10 months of data. There is a slight but significant higher probability of underprediction vs overprediction. The percentage of predictions that were within 40% to -40% and within 100% to -100% of observed BI were 48.3% and 80.9%, respectively.

We envision the following implementation: Initially, a model structure would be selected and validated using the procedures herein described. Then, with an estimate of the model performance, BI predictions, using up-to-date data of all other products based on which the model is trained, can then easily be generated for a new product (with or without prior BI data of that product). The validation results can yield insights into the expected accuracy for this new drug for a specific future month (*t_pred*) and a specific amount of available BI data (*t_data*). At some future moment, of which the specifics are beyond the scope of this paper, continuous model retraining on new data will probably not suffice as the validation set at that time will not be representative of the training set on which the model structure was developed. In that case, model selection and validation would have to be redone. This would also be applicable to using our methodology in other jurisdictions or geographic areas, for predicting different or entirely new drug classes or when adapting to changes in regulatory systems.

Our BI predictions are quite constant and rather accurate from 10 months of available data as the median aggregated error from 10 months onwards ranges from 1.20 to 1.42. As the error is highest with little available data, one could say that our approach is not useful for BI predictions when these predictions are part of a reimbursement dossier in a 'closed' reimbursement system, indicating reimbursement for new drugs is only available after an HTA decision. However, various countries have (partly) open systems (e.g. Germany, the Netherlands) wherein HTA

dossiers become available after the drug is already available and in use. In our dataset, we have shown that 50% of the products have a substantial amount of data available at the date of publication of the reimbursement dossier. It is therefore very probable that, at least for open reimbursement systems, a sufficient amount of BI data will in many cases be available to overcome the high errors associated with having less than 5 – 10 months of available data for prediction.

When extending the use of BIA beyond the initial reimbursement decision to a more dynamic drug life-cycle approach, for example as part of managed entry agreements, available BI data will keep increasing and will therefore rapidly be sufficient for achieving our reported maximum predictive accuracy [32].

Cha et al. analysed the accuracy of peak sales forecasts produced by so called sell-side analysts [17]. They categorised the forecasts in categories of percentage difference between forecasts and observed peak sales. Their highest deviation categories were < -80% (n=7/260) and >160% (n=57/260) and found a median error of 4%. They do state that most forecasts are poor and that the variance is high, but the 4% median error does not give clear insight in forecast error as overestimations can cancel out underestimations (i.e., their error is not symmetric) and as the maximum error is limited (-80% and 160%). We have partly applied the methodology used by Cha et al. to our dataset by also limiting the maximum error and by not making the error symmetric. Using this method, our median error is -3%; a major difference between our unaggregated and symmetric median error of 1.54 (154%). Cha and colleagues furthermore state that more than 60% of the forecasts were off by more than 40%, whereas in our analysis 51.7% of estimations had a higher deviation than 40%.

Broder et al. published a review of the bias in BI predictions of new drugs [18]. They used a US perspective and included formal, more scientific, BI predictions as well as informal predictions that are aimed at projecting share prices. All estimates were made less than 12 months before launch and nearly all estimates were for just the first year. Mean predicted BI in the sample was 5.5 times the observed BI. When excluding the informal predictions, the average overestimation is 5.6 times the observed BI. These values are asymmetric representations of under- and overprediction (i.e., the value is attenuated towards 1 due to underpredictions that have a value between 0 and 1) and are still higher than our (symmetric) mean error. Only 20% of the predictions were within 40% of the observed usage, compared to 48,3% for our model [18]. If we then relate to the differing lengths of forecast period (i.e., t_pred) of one year for Broder et al. and 45 months for our study, we could argue that our predictions seem to have better accuracy whilst providing more future predictions.

Keeping et al. investigated BI estimates that were part of pharmaceutical company submissions and compared them to the observed expenditure [19]. These company submissions were issued to the All Wales Medicines Strategy Group (AWMSG) for reimbursement decision making.

The AWMSG is the institution that appraises clinical and cost-effectiveness of new medicines being considered for National Health Service prescribing in Wales (United Kingdom). A total of 49 medicines were included and the percentage of predictions in the 40% and 100% range were 20.4% and 53.1%, respectively. Our model achieved 48.3% and 80.9% on these respective accuracy markers. Of the 49 products Keeping and colleagues included, only 3 - 6 (depending on the definition) had an oncological indication which is therefore quite different from our oncology cohort. Still, as the BI estimates investigated by Keeping et al. are those used by payers to inform decision making, the work of Keeping et al. is very relevant. Even though our results are not directly comparable, we still argue that our superior performance in the 40% and 100% range metric is a rather clear indicator that our method has the potential to be superior to current BI estimates used by payers and decision makers.

Our BI prediction approach potentially has several advantages over current BI estimation procedures. Firstly, our methodology is independent of indication extensions. Of course, additional indications do have an influence on BI and possibly on the accuracy of the predictions. We however chose to not include indication expansions as a predictor variable as this would be rather laborious to perform in practice for a large group of products. Unlike current Dutch Reimbursement authority ('National Health Care Institute') BI predictions, our model intrinsically adjusts for possible changes in indications as we apply a drug perspective irrespective of indication.

Another potential advantage of our BI prediction approach is the ease and speed with which BI predictions can be constructed. Updating the data, possibly performing a separate validation and then performing the prediction for a new drug would be a matter of hours whereas the current guidelines call for a much more time-consuming endeavour [33,34]. This advantage is especially profound if you include the option of semi-automatically updating the predictions as time passes and more data becomes available.

Finally, our model results yield predictions with a potentially quantifiable amount of uncertainty as the distribution of error is known and can be adjusted for the amount of data already available (*t_data*) and the number of future months (*t_pred*). This is hardly possible with current BI predictions that produce point estimates. Our approach could therefore serve as a basis for more profound modelling of uncertainty around BI predictions.

Our study has various limitations. First, we have only validated our model for a rather specific set of products and characteristics. Future products, for example novel advanced therapy medicinal products, are not validated and are therefore probably not accurately predictable by our current model. As is however described above, the dataset can be updated to future states and a new validation can then be done rather easily in order to accommodate new drug classes and/or characteristics.

Second, we have no direct comparison of our results to the current BI estimations used in practice. As our model is based on Dutch data it would be very insightful to compare our results with observed BI predictions published by the Dutch Reimbursement authority. In light of the Broder et al., Cha et al. and Keeping et al. findings, our prediction accuracy appears to be superior.

Third, we have capped minimum and maximum BI predictions which to an extent impacted the results. Our explicit assumption of a predicted maximum of two times the maximum monthly BI in the total dataset has no evidentiary basis. Potentially worse, there were records in the dataset with monthly BI below the minimum monthly amount of €5000. In other words, it is quite probable that we overestimate certain products with monthly BI below €5000. We however believe that these caps are justified as one of our main aims is to provide payers with better BI predictions. A difference between €5000 and €50 yields 4.6 log units of deviation but this difference, in absolute terms, is probably not very relevant to payers.

We have explored the influence of these value restrictions through scenario analyses. These have clearly indicated that only limiting the lower values to €5000 and not restricting maximum values delivers a nearly identical predictive performance. We thus believe that these limits improve the relevance of our outcomes as prediction errors that are irrelevant on a macro level, e.g. €5 vs €5000 per month, are omitted. The high-level caps are implemented as some modelling scenarios yielded predictions that were irrationally high (for example, higher than the entire Dutch healthcare budget) and can therefore be identified by potential users of this method. In order to limit these scenarios to realistic figures, the factor two limit was imposed.

Fourth, we understand that alternative potentially more advanced validation techniques have been developed. In order to construct a methodology that is suitable for informing decision making, the method has to be interpretable and transparent. We therefore abstain from adding more complexity to the current model in order to also keep it as practical as possible.

CONCLUSIONS

We believe that, based on our validation, we have developed a valid method to predict BI. We were able to compare our results with three independent studies using a metric that describes the number of predictions that are within 40% to -40% and 100% to -100% of the observed BI. Our model was superior to all these three studies and especially the study of Keeping et al. is important in this regard as they investigated the accuracy of BI predictions used by payers for reimbursement decision making. We can therefore conclude that our approach can be used to develop models that can provide improved predictive accuracy compared to the current practice of conducting BIA. Additionally, our data-driven approach would allow for a more

dynamic, life-cycle approach to predicting and managing BI of drugs. To conclude, we think that our approach can be a valuable addition to BI predictions due to its potential for increased accuracy, independence of indications and ability to keep updating the predictions to the most recent data.

ACKNOWLEDGEMENTS

We would like to thank FarmInform for providing the budget impact data.

REFERENCES

1. The Organisation for Economic Co-operation and Development (OECD). New Health Technologies: Managing Access, Value and Sustainability [Internet]. [cited 2018 Feb 7]. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/managing-new-technologies-in-health-care_9789264266438-en
2. American Society of Clinical Oncology Position Statement on Addressing the Affordability of Cancer Drugs. J Oncol Pract. 2017;JOP.2017.027359.
3. Barua S, Greenwald R, Grebely J, Dore GJ, Swan T, Taylor LE. Restrictions for Medicaid Reimbursement of Sofosbuvir for the Treatment of Hepatitis C Virus Infection in the United States. Ann Intern Med. 2015;163:215.
4. Re VL, Gowda C, Urick PN, Halladay JT, Binkley A, Carbonari DM, et al. Disparities in Absolute Denial of Modern Hepatitis C Therapy by Type of Insurance. Clin Gastroenterol Hepatol. 2016;14:1035–43.
5. Paulden M, O'Mahony J, McCabe C. Determinants of Change in the Cost-effectiveness Threshold. Med Decis Making. 2017;37:264–76.
6. Kleijnen S, Lipska I, Leonardo Alves T, Meijboom K, Elsada A, Vervölgyi V, et al. Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. Ann Oncol. 2016;27:1768–75.
7. Collins M, Latimer N. NICE's end of life decision making scheme: impact on population health. BMJ. 2013;346:f1363–f1363.
8. Maarse H, Jeurissen P, Ruwaard D. Results of the market-oriented reform in the Netherlands: a review. Health Econ Policy Law. 2016;11:161–78.
9. Flume M, Bardou M, Capri S, Sola-Morales O, Cunningham D, Levin L-A, et al. Approaches to manage 'affordability' of high budget impact medicines in key EU countries. J Mark Access Health Policy [Internet]. 2018 [cited 2018 Jul 10];6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5998770/
10. Eichler H-G, Pignatti F, Flamion B, Leufkens H, Breckenridge A. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma. Nat Rev Drug Discov. 2008;7:818–26.
11. Hughes-Wilson W, Palma A, Schuurman A, Simoens S. Paying for the Orphan Drug System: break or bend? Is it time for a new evaluation system for payers in Europe to take account of new rare disease treatments? Orphanet J Rare Dis. 2012;7:74.
12. Ermisch M, Bucsics A, Vella Bonanno P, Arickx F, Bybau A, Bochenek T, et al. Payers' Views of the Changes Arising through the Possible Adoption of Adaptive Pathways. Front Pharmacol [Internet]. 2016 [cited 2018 Mar 7];7. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2016.00305/full
13. Lipska I, Hoekman J, McAuslane N, Leufkens H, Hövels A. Does conditional approval for new oncology drugs in Europe lead to differences in health technology assessment decisions? Clin Pharmacol Ther. 2015;98:489–91.
14. Wiss J, Levin L-A, Andersson D, Tinghög G. Prioritizing Rare Diseases: Psychological Effects Influencing Medical Decision Making. Med Decis Making. 2017;37:567–76.
15. Coyle D, Cheung MC, Evans GA. Opportunity Cost of Funding Drugs for Rare Diseases: The Cost-Effectiveness of Eculizumab in Paroxysmal Nocturnal Hemoglobinuria. Med Decis Making. 2014;34:1016–29.
16. Vooren K van de, Duranti S, Curto A, Garattini L. A Critical Systematic Review of Budget Impact Analyses on Drugs in the EU Countries. Appl Health Econ Health Policy. 2014;12:33–40.
17. Cha M, Rifai B, Sarraf P. Pharmaceutical forecasting: throwing darts? [Internet]. Nat. Rev. Drug Discov. 2013 [cited 2018 Mar 7]. Available from: https://www.nature.com/articles/nrd4127

18. Broder MS, Zambrano JM, Lee J, Marken RS. Systematic bias in predictions of new drugs' budget impact: analysis of a sample of recent US drug launches. Curr Med Res Opin. 2018;34:765–73.

19. Keeping S, Deslandes PN, Haines KE, Routledge PA. Estimated Versus Observed Expenditure Associated with Medicines Recommended by the All Wales Medicines Strategy Group. PharmacoEconomics - Open [Internet]. 2019 [cited 2019 Jan 22]; Available from: https://doi.org/10.1007/s41669-019-0116-5

20. English - Farminform [Internet]. [cited 2018 Apr 26]. Available from: https://www.farminform.nl/English

21. PHARMO Institute for Drug Outcomes Research [Internet]. Pharmo. [cited 2018 Apr 26]. Available from: https://www.pharmo.nl/

22. Goettsch WG, Jong RB de, Kramarz P, Herings RMC. Developments of the incidence of osteoporosis in The Netherlands: a PHARMO study. Pharmacoepidemiol Drug Saf. 2007;16:166–72.

23. European Medicines Agency - European public assessment reports - European public assessment reports: background and context [Internet]. [cited 2018 Feb 9]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/general/general_content_000433.jsp

24. Community Register - European Commission [Internet]. [cited 2018 Feb 28]. Available from: http://ec.europa.eu/health/documents/community-register/

25. Eder J, Sedrani R, Wiesmann C. The discovery of first-in-class drugs: origins and evolution. Nat Rev Drug Discov. 2014;13:577–87.

26. Research C for DE and. New Drugs at FDA: CDER's New Molecular Entities and New Therapeutic Biological Products [Internet]. [cited 2018 Feb 28]. Available from: https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/default.htm

27. R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.

28. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2017). nlme: Linear and Nonlinear Mixed Effects Models [Internet]. Available from: https://CRAN.R-project.org/package=nlme

29. Shmueli G. To Explain or to Predict? Stat Sci. 2010;25:289–310.

30. Tofallis C. A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation [Internet]. Rochester, NY: Social Science Research Network; 2014 Jul. Report No.: ID 2635088. Available from: https://papers.ssrn.com/abstract=2635088

31. Tornqvist L, Vartia P, Vartia YO. How Should Relative Changes Be Measured? Am Stat. 1985;39:43–6.

32. Klemp M, Frønsdal KB, Facey K. What principles should govern the use of managed entry agreements? Int J Technol Assess Health Care. 2011;27:77–83.

33. Principles of Good Practice for Budget Impact Analysis [Internet]. [cited 2018 Apr 26]. Available from: https://www.ispor.org/workpaper/BudgetImpactAnalysis/BIA_TF0906.asp

34. Ministerie van Volksgezondheid W en S. Format Budget Impact Analyse GVS (voorheen kostenconsequentieraming) - Publicatie - Zorginstituut Nederland [Internet]. 2016 [cited 2018 Apr 26]. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/04/06/format-budget-impact-analyse-gvs-voorheen-kostenconsequentieraming

APPENDICES

Appendix 1. Included products

Supplemental table 1. list of included products in the training set and validation set.

Training set product name	Training set active substance name	Validation set product name	Validation set active substance name
Xeloda	capecitabine	Adcetris	Brentuximab
Avastin	bevacizumab	Zaltrap	aflibercept
Vectibix	panitumumab	Tagrisso	osimertinib
Erbitux	cetuximab	Inlyta	axitinib
Tyverb	lapatinib	Perjeta	pertuzumab
Glivec	imatinib	Venclyxto	venetoclax
Halaven	eribulin	Zykadia	ceritinib
Iressa	gefitinib	Erivedge	vismodegib
Votrient	pazopanib	Ibrance	palbociclib
MabCampath	alemtuzumab	Lartruvo	olaratumab
Nexavar	sorafenib	Pixuvri	pixantrone
Jevtana	cabazitaxel	Imbruvica	ibrutinib
Arzerra	ofatumumab	Teysuno	tegafur / gimeracil / oteracil
Tarceva	erlotinib	Jakavi	ruxolitinib
Sutent	sunitinib	Portrazza	necitumumab
Tasigna	nilotinib	Kadcyla	trastuzumab
Torisel	temsirolimus	Alecensa	alectinib
Alimta	pemetrexed	Lynparza	olaparib
Sprycel	dasatinib	Imlygic	talimogene laherparepvec
Velcade	bortezomib	Caprelsa	vandetanib
Vidaza	azacitidine	Mekinist	trametinib
Yervoy	ipilimumab	Keytruda	pembrolizumab
Zelboraf	vemurafenib	Yondelis	trabectedin
Herceptin	trastuzumab	Zydelig	idelalisib
Targretin	bexarotene	Opdivo	nivolumab
		Kyprolis	carfilzomib
		Cyramza	ramucirumab
		Darzalex	daratumumab
		Dacogen	decitabine
		Cotellic	cobimetinib
		Vargatef	nintedanib
		Ninlaro	ixazomib citrate
		Xalkori	crizotinib
		Atriance	nelarabine
		Gazyvaro	obinutuzumab
		Blincyto	blinatumomab
		Lonsurf	trifluridine / tipiracil

Supplemental table 1. (continued)

Training set product name	Training set active substance name	Validation set product name	Validation set active substance name
		Giotrif	afatinib
		Stivarga	regorafenib
		Tafinlar	dabrafenib
		Bosulif	bosutinib
		Lenvima	lenvatinib
		Farydak	panobinostat
		Evoltra	clofarabine

Appendix 2: Main outcomes, scenario where only minimum outliers were capped

Results of scenario analysis where only minimum outliers were capped.

Outcome	Value
Mean error, aggregated per t_data and t_pred (SD)	3.14 (2.87)
Mean error, not aggregated (SD)	3.03 (7.40)
Median error, not aggregated (5^{th}, 25^{th},75^{th} and 95^{th} percentile)	1.57 (1.04, 1.21, 2.63, 8.61)

Appendix 3: Main outcomes, scenario where only maximum outliers were capped

Results of scenario analysis where only maximum outliers were capped.

Outcome	Value
Mean error, aggregated per t_data and t_pred (SD)	6.02 (42.79)
Mean error, not aggregated (SD)	8.00 (358.74)
Median error, not aggregated (5^{th}, 25^{th},75^{th} and 95^{th} percentile)	1.58 (1.04, 1.22, 2.67, 9.09)

Appendix 4: Main outcomes, scenario where no outliers were capped.

Results of scenario analysis where no outliers were capped.

Outcome	Value
Mean error, aggregated per t_data and t_pred (SD)	6.15 (42.82)
Mean error, not aggregated (SD)	8.09 (358.77)
Median error, not aggregated (5^{th}, 25^{th},75^{th} and 95^{th} percentile)	1.58 (1.04, 1.22, 2.67, 9.13)

4.2

Increasing the information provided by probabilistic sensitivity analysis: the relative density plot

Joost W. Geenen*, Rick A. Vreman*, Cornelis Boersma, Olaf H. Klungel, Anke M. Hövels, Renske M.T. ten Ham

*These authors contributed equally.

Manuscript in preparation for submission.

ABSTRACT

Background

Results of probabilistic sensitivity analyses (PSA) are typically visualised as a scatter plot. Although useful, such scatter plots have two major limitations. First, high density areas cannot be correctly interpreted due to overlap of individual estimates (overdrawing). Second, relative density cannot be interpreted which may cause decision-makers to give too much weight to relatively infrequent scenarios. To overcome these limitations, we developed a novel visualisation of PSA results: The Relative Density plot (PSA-ReD). Here, we demonstrate PSA-ReD using one theoretical and two real-world case studies.

Methods

The PSA-ReD combines a density plot and a contour plot to visualise PSA results. Density is calculated using kernel density estimation. Relative density, depicted using a colour gradient, is transformed to cumulative probability. Contours are then plotted over regions with a specific cumulative probability. We use one theoretical case study (normal distribution) and two real-world case studies (published health-economic models) to demonstrate the PSA-ReD plot. The PSA-ReD plot was created using R. The R-script and manual are publicly available on GitHub.

Results

In the case studies, the PSA-ReD provided additional visual information that could not be understood from the traditional scatter plot. High density areas were identified by colour-coding and the contour plot allowed for quantification of PSA iterations within areas of the cost-effectiveness plane, diminishing overdrawing and putting infrequent iterations in perspective.

Conclusions

The PSA-ReD plot is easy to implement, presents more of the information enclosed in PSA data, and prevents inappropriate interpretation of PSA results. Thus, this new PSA presentation provides modellers with additional information about model behaviour and can help decision-makers to more appropriately interpret probabilistic model results.

INTRODUCTION

Health economic models have become an integral part of healthcare decision making [1]. These models rely on input parameters associated with uncertainty which must be taken into account when calculating and presenting model results [2]. Deterministic and probabilistic sensitivity analyses (DSA & PSA) are systematic approaches that quantify the impact of uncertainties related to model inputs on the outcomes of the model [3]. Providing results of sensitivity analyses is advised by modelling and reporting guidelines and is often mandatory for the submission of health technology assessment dossiers [4–6]. The PSA has been the most prominent method to quantify the impact of combined uncertainty of all model input parameters [7]. In a PSA all input parameters are simultaneously varied along predefined ranges according to their specific distribution, with the outputs presented as a scatter plot in the cost-effectiveness plane (CE-plane) [8].

The traditional PSA scatter plot is useful to quickly visualise the distribution of PSA results as well as the correlation between the cost and the effect measure of interest [9]. A critical aspect of the scatter plot is its ability to illustrate the distribution of PSA samples over the quadrants of the cost-effectiveness plane (i.e., increased Quality Adjusted Life Years [QALY] and increased costs or decreased QALYs and increased costs). The scatter plot itself is not the sole measure to quantify and interpret parameter uncertainty as, for example, the likelihood of cost-effectiveness is typically illustrated with a cost-effectiveness acceptability curve (CEAC). Still, the aforementioned properties make the PSA scatter plot an intuitive, useful and usually mandatory figure in communication towards stakeholders, who might be less familiar with uncertainty analyses. Despite these advantages, the traditional scatter plot has two major limitations.

The first limitation is that in the traditional scatter plot, individual point estimates are overlapping in high density areas. This so-called overdrawing makes it hard to assess the relative density of point estimates in various areas of the plot [10]. Second, due to difficulty in estimating this relative density, infrequent scenarios appear very prominent in the traditional figure. This may cause overestimation of the occurrence of these scenarios and may incorrectly inform decision making.

To overcome these two limitations a novel presentation of the PSA scatter plot is desired. Increased computational power combined with increased popularity of open source software, such as R, provide the tools to improve the traditional PSA presentation [11]. Two R-packages have been developed to display PSA scatter plot results. The heemod package uses coloured hexagons to display relative density which gives some information on overdrawing [12]. The BCEA package by Baio et al. provides the tools to draw a contour plot using ellipses in discrete intervals. However, the BCEA package requires purchase of a costly manual [13]. Both

packages have the same major drawback; they require the use of package specific syntax to be able to use and apply the package features. This requires extensive R-skills which can put-off users less familiar with programming language. Additionally, the features available within these packages provide either a plot showing relative density (heemod) or a contour plot (BCEA). Neither provides a combination of both these plot elements.

We therefore developed a novel open source graphical presentation of PSA results, incorporating relative density and probability contours, overcoming both overdrawing and outlier overestimation. The method is independent from modelling software and relies only on an import of PSA results in .csv format. We call this new PSA presentation the Relative Density plot (PSA-ReD).

The aim of this paper is to illustrate the concept and functionalities of the PSA-ReD plot. We do so using one theoretical and two real-world case studies. We also provide the R code designed for direct application to any user's own research outputs together with a user manual on GitHub [14].

METHODS

Relation to traditional cost-effectiveness plane

A traditional PSA output is a two-dimensional black and white scatter plot presented on a CE-plane (Figure 1a). The PSA-ReD plot (Figure 1b) combines a multi-coloured density plot (appendix 1, figure a) and a contour plot (appendix 1, figure b). The combination of these two plots allows the reader to identify and distinguish high density areas using a colour scale, as well as a quantification of the point estimate density within the CE-plane, thus visualising the information that remains hidden in the traditional scatter plot. This increases the information that can be understood from the scatter plot and improves understanding of the parameter uncertainty which a PSA is aimed to address.

The two features of the PSA-ReD plot can be constructed separately. The density plot (Appendix 1, figure b) could be interpreted as a two-dimensional histogram. Like a traditional one-dimensional histogram, the two axes are divided in sections. These sections on both axes divide the two-dimensional space in distinct rectangular regions. Then, as in a one-dimensional histogram, the number of data points per region is counted and transformed to present the relative frequency using a colour scale. Low density is presented by a green to blue scale and high density is presented by a yellow to red scale.

When using a histogram, the choice of the anchor point of the plot area (i.e., the range and starting points of the axes) has influence on the graphical outcome [15]. This effect would be most pronounced in a one-dimensional histogram with relatively few datapoints: Shifting

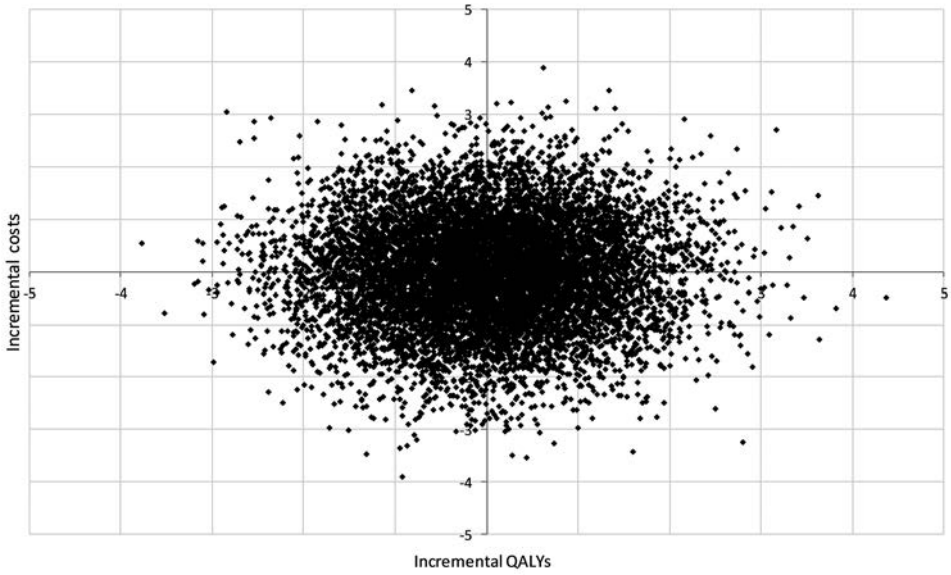


Figure 1a. Traditional scatterplot displaying PSA results. Displaying a bivariate normal distribution with mean = 0, sd = 1, 10,000 iterations

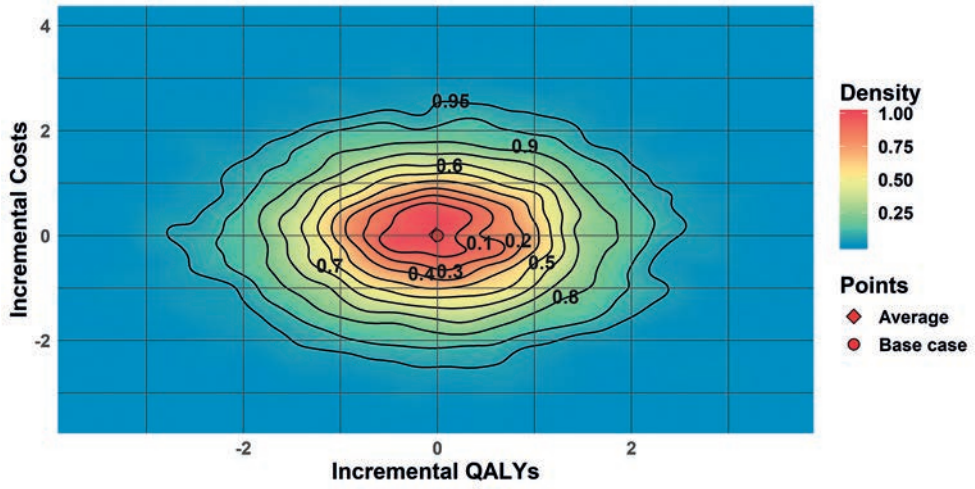


Figure 1b. New graphical presentation of PSA using the Relative Density plot (PSA-ReD). Displaying a bivariate normal distribution with mean = 0, sd = 1, 10,000 iterations and 1000 bins. PSA-Probabilistic Sensitivity Analysis.

the x-axis would change the histogram as, by chance, the number of datapoints falling within each bin would differ with each x-axis shift. Crucially, the underlying data remains the same and this bias would also be present in two-dimensional histograms [15]. To overcome this, bivariate kernel density estimation (kde) is used as it provides a more accurate representation

of the probability density [15]. Instead of counting the number of data points per rectangular section, each data point is surrounded by a kernel which are summed to yield the kernel density estimate. Each data point is thereby smoothened over a small surrounding area (data kernel) instead of being a single data point [15]. The size of this area is determined by the data as explained in the 'technical aspects' paragraph.

The PSA-ReD plot combines the density plot with a contour plot (appendix 1, figure B). The contours indicate the boundaries of regions with a specific cumulative probability. This cumulative probability is calculated by converting the density estimates to cumulative densities per area. These cumulative area densities are then mapped to represent the cumulative probability range of 0 to 1. A contour line is then drawn joining areas with specific pre-specified values of cumulative probability (e.g. 0.1, 0.5, 0.95).

Hardware and software

The script to realise the PSA-ReD plot was developed and tested using R version 3.5.1 and Rstudio version 1.1.453 [11,16]. For our analyses, we used a standard consumer grade personal computer (Dell Optiplex 9020). In appendix 2, we provide detailed information on the hardware and software used.

The R script that we used is available in a GitHub repository as well as in appendix 6 [14]. We adhered to Google's R Style guide and provide step-by-step guidance using comments embedded in the script [17]. The R script is licensed under the GNU General Public License v3.0. In short, this means that users are free to run, study, share and modify the software. The license dictates, among other things, that the software (or derivative work) must be open source and that derivative work must be published using the same license [18]. This guarantees that our project can be used and optimised by anyone whilst ensuring that it remains open to all.

Technical aspects of plot generation in R

In R, we use the kde2d function from the MASS package to perform the aforementioned kernel density estimation [19]. In essence, the outcome of kde is a density value per area of a prespecified size, comparable to the number of data points within each bin in histograms. Detailed information is provided in the work by Silverman and in the documentation of the MASS package [19,20]. As these density values are very small and hard to interpret, we normalise these values by taking the reciprocal of the maximum density value to yield values ranging from 0.0 to 1.0. With these, we generate an easy to interpret plot with a scale from 1.0 (highest density) to 0.0 (lowest density).

The kde2d function has, besides the x and y values, two arguments that influence the kde. These are n (the number of bins in each dimension) and h (the bandwidth that determines the level of smoothing). The number of bins defines the number of sections on each axis. The total number

of areas within the resulting plots is therefore horizontal bins * vertical bins (e.g. 100*100 = 10000). An easy analogy of these areas would be to regard them as pixels, the bins then determine the resolution in both directions. This pixel-analogy only reflects to the number of underlying bins. As we outline in the appendix 2 regarding the saving of plots, the actual resolution of the figures can be specified and is irrespective of the number of bins used. As the number of bins can be interpreted as the resolution of the figure, a larger number of bins produces a more precise figure. However, increasing the number of bins also increases computation time which means a balance must be struck.

In appendix 3, we present the influence of different bin sizes. Using 50-500 bins (appendix 3, figures a and b), yields a density gradient that is not smooth and may appear like the image is pixelated. With 1000 bins (appendix 3, figure c), the image is smooth, no pixilation can be identified, and all the computation is performed within one minute on the aforementioned consumer grade computer. With more bins (2000, appendix 3, figure d), the image does not get smoother but it does lead to increased RAM usage and computation time. We therefore recommend using 1000 bins and have used this number of bins in all figures throughout the manuscript, unless otherwise stated.

The h argument of the kde2d function determines the bandwidth of the kernel areas. It can be interpreted as the size of the kernels that is applied when converting each data point to a data kernel. We have chosen to leave this at the default setting where the bandwidth is automatically selected based on the data by the well-established MASS package (specifically, the bandwidth.nrd function) [19]. This guarantees generalisability of results.

Number of PSA iterations

As in any PSA, it is preferred to run as many iterations as necessary to reach model convergence [3]. We explored the influence of the number of iterations used by varying this between 1000 – 100,000 iterations, as presented in appendix 4. As RAM usage and computation time increases when more iterations are used, we recommend using a maximum of 10,000 iterations. Running the script with 10,000 iterations takes a maximum of 1 minute. In all figures throughout this manuscript, we have used 10,000 iterations unless otherwise stated.

User modifications

Other parameters that can be altered by the user are contour levels, axis-, legend- and plot titles, font sizes and font types. In the supplied script, it is explained how and where this can be done. Apart from these cosmetic changes, we provide means to zoom on a particular area of the plot and generate a new plot from that specific area, as well as two rendering options to avoid clipping of contour labels in these zoomed images. Appendix 5 displays this zooming capability. We also provide a feature that allows users to plot willingness to pay (WTP) thresholds in the PSA-ReD plot, as well as plotting the base case scenario and the average of the PSA. Appendix 2 provides in-depth explanations on the use of the various features described above.

Case study demonstration

To demonstrate the novel graphical presentation, the concept was applied to three exemplary case studies. The first case study is a theoretical example which assumes a model with only two standard normally distributed parameters (mean = 0 and standard deviation = 1) that define the incremental costs and incremental QALYs.

The two real-world case studies were selected as a convenience sample as we needed access to the raw PSA results and because the case studies should have been published. The two selected cases each show a different pattern within the PSA results. Both patterns are commonly seen in economic evaluations. The first real-world case study assesses the influence of three characteristics (cost, specificity and sensitivity) on cost-effectiveness of a hypothetical pharmacogenomic test for prevention of angiotensin-converting enzyme inhibitor induced angioedema (denoted as 'eHTA study') [21]. The second real-world case study used a three-state partitioned survival model to investigate cost-effectiveness of periodic therapeutic drug monitoring of endoxifen levels in breast cancer patients (denoted as 'TDM study') [22].

RESULTS

Normal distribution case study

The traditional CE-plane of the bivariate normal model would look like figure 1a. The base case would be at zero incremental costs and zero incremental QALYs. Though we can see that the borders of the area are less densely populated, it is unclear how the density of iterations is spread over the populated area. If instead we look at the PSA-ReD plot in figure 1b, it becomes clear that the density is evenly spread around the base case, as would be expected for this normally distributed data. Additionally, the contours give insight into the spread of the iterations. In this case study, the area containing 95% of the iterations will approximate that of a 95% confidence interval because we used normal distributions. A bivariate normal distribution is distributed according to the χ^2-distribution with two degrees of freedom [23]. Taking the square root of the critical value for the 95% confidence interval (5.99), results in the area borders of the confidence interval (2.45). This is clearly shown by the contours in the PSA-ReD plot. In general, the probability that the values for two variables within a joint distribution together fall in any area of their two dimensions is given by the volume (or cumulative probability) under the density function above that area. This is exactly what the PSA-ReD method calculates when providing the contours.

eHTA case study

The results of the PSA of the eHTA study are presented in figure 2. This figure shows the PSA results both in traditional presentation (2a) as well as via the PSA-ReD plot (2b). The figures are both based on 5000 PSA iterations, as this reflects the number of iterations in the published paper

[21]. The classic CE-plane implies more spread due to a small number of iterations that generate relatively high incremental QALYs. However, the PSA-ReD plot shows these are extremely infrequent and fall outside the contour area that includes 95% of the iterations. Additionally, apparently 10% of all iterations appear within an area of approximately 0.05 incremental QALYs (0.0 – 0.05) and 1000 incremental euros (4000 – 5000). Particularly interesting is that the base case falls well outside this most dense area. This is contrary to what would be expected in a PSA as generally, the most likely outcome for the incremental cost-effectiveness ratio (ICER) based on the individual distributions of parameters is close to the base case. Thus, one would expect the highest density area to be surrounding the base case.

However, when one of the model parameter distributions is skewed (i.e. a beta or gamma distribution), the resulting average of all PSA samples will, by definition, not lie on the point of highest density as the average will lean towards the tail of the specific distribution. In certain parameterisations of the beta and gamma distributions (e.g. when $\alpha < 1$ and $\beta > \alpha$), the base case value will not be the value with the highest probability density of that specific distribution. Instead, the value of 0 will have the highest probability density. In the eventual PSA-ReD plot, this effect attenuates the area of highest density away from the base case towards 0 as is especially apparent in this case study. This information cannot be interpreted from the traditional CE-plane. Therefore the PSA-ReD plot can provide modellers with information regarding model behaviour.

A .csv datafile with the incremental QALYs (x-values) and incremental costs (y-values) of the eHTA PSA results is provided in the GitHub repository to allow the reader to recreate the PSA-ReD [14].

TDM case study

Figure 3 shows the PSA results from the TDM case study both in traditional presentation (3A) as well as via the PSA-ReD plot (3B). Both figures are based on 10,000 PSA iterations. Density in the classic CE-plane is not interpretable but suggests a relatively high density around the base case and in the upper left corner of the plane. Additionally, there seems to be accumulation of iterations along the Y-axis. The PSA-ReD plot more precisely clarifies the high density that is found within the small area near the origin. Additionally, the relatively high density suggested by the CE-plane around the Y-axis is put into perspective by the PSA-ReD plot, clarifying that these scenarios are relatively infrequent.

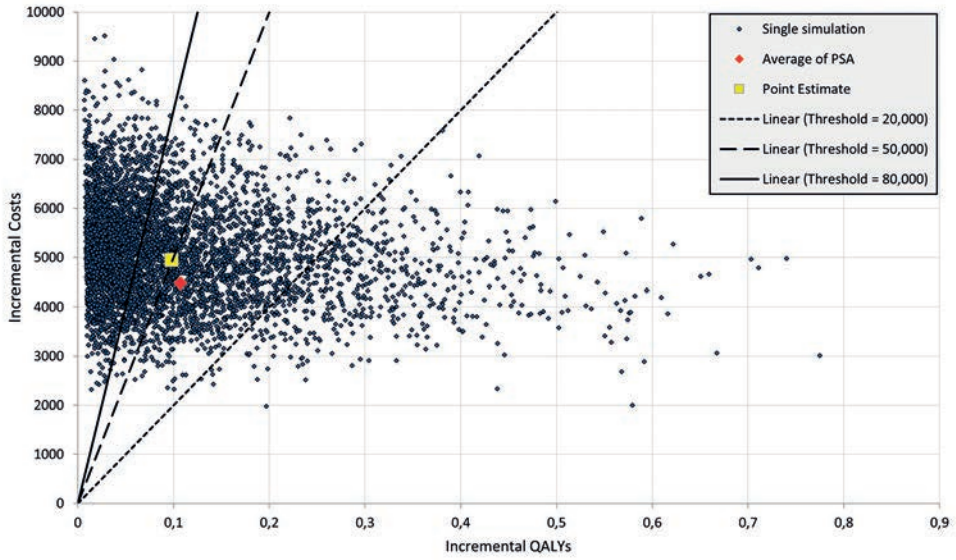


Figure 2a. Probabilistic sensitivity analysis results of eHTA-study using the traditional scatter plot presentation. Generated using 5000 PSA iterations.

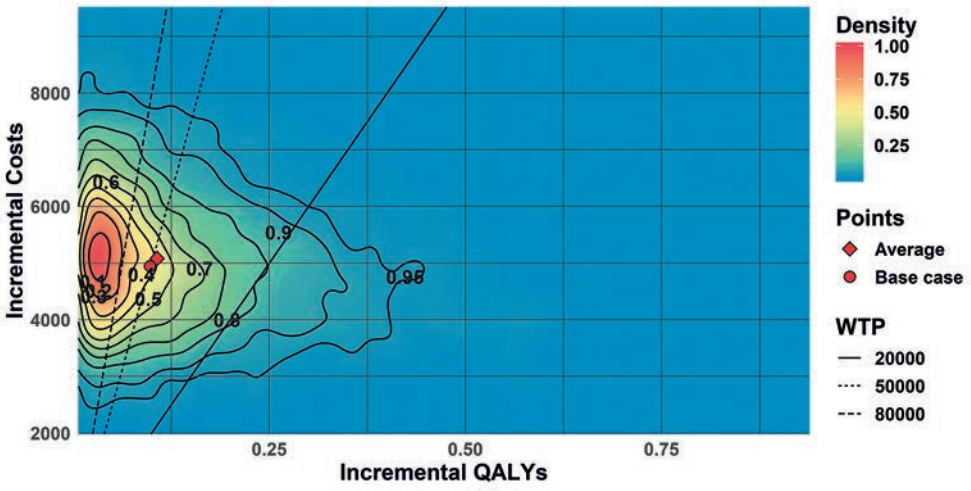


Figure 2b. Probabilistic sensitivity analysis output of eHTA-study using the PSA-ReD presentation. Generated using 5000 PSA iterations and 1000 bins.

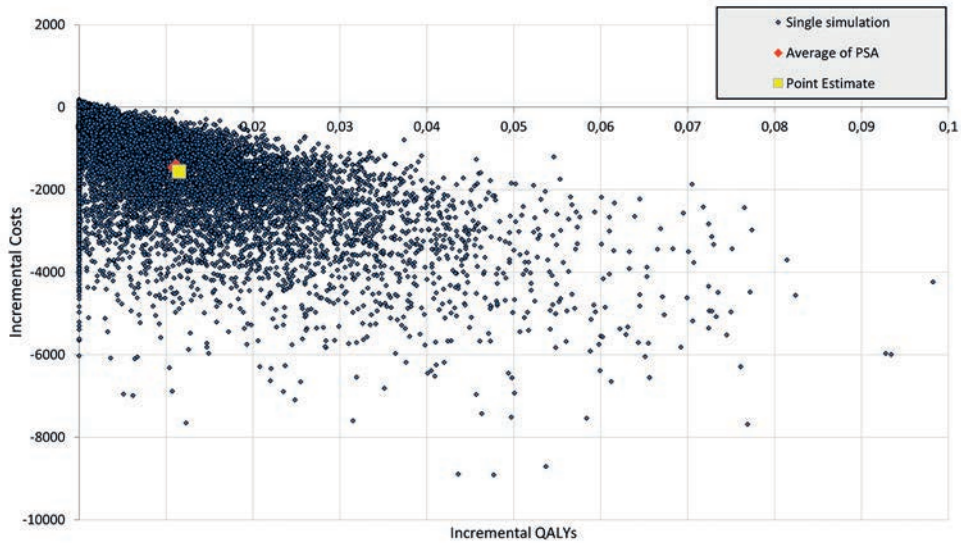


Figure 3a. Probabilistic sensitivity analysis results of TDM study using the traditional scatter plot presentation. Generated using 10,000 PSA iterations.

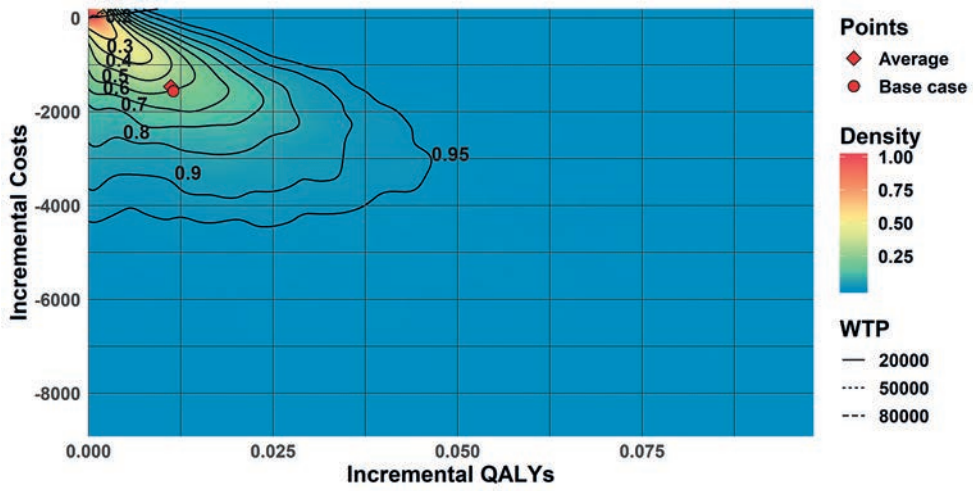


Figure 3b. Probabilistic sensitivity analysis results of TDM study using the PSA-ReD presentation. Generated using 10,000 PSA iterations and 1000 bins.

DISCUSSION

In the theoretical case study we demonstrated the application and interpretation of the relative density estimation function and the applied normalisation of the values on which the PSA-ReD plot is based. In the real-world case studies we demonstrated how the PSA-ReD plot provides more insight into the relative density and cumulative area probabilities. Thus, the PSA-ReD plot provides visual information that is not provided by the traditional PSA scatter plot within the CE-plane nor by solely a density plot or a contour plot.

The benefits of the PSA-ReD plot over the traditional scatter plot are evident. The accumulation of PSA results within certain areas of the cost-effectiveness plane cannot be interpreted by the traditional scatter plot. The PSA-ReD plot not only clearly visualises the location of these high-density areas, it also provides a quantification of the proportion of PSA iterations within these areas. Additionally, inappropriate significance could be attributed to relatively infrequent PSA iterations in the traditional scatter plot. The PSA-ReD plot diminishes this effect.

The PSA-ReD plot provides modellers with increased insight into the relation between all input parameter distributions and the subsequent distribution of model outcomes. This can serve as an additional validation to confirm the model works as intended. Besides additional information for modellers, the PSA-ReD plot provides additional insights for decision-makers. Decision-makers are often the end user of models but are generally not as familiar with health economic modelling practices. To assist decision-making by end users it is crucial to provide intuitive and informative presentations of the outcomes of health economic models. The PSA-ReD plot informs decision-makers about the relative and cumulative likelihood of areas of incremental costs and incremental outcomes in an intuitive figure.

Currently, R packages exist that provide the option for plotting density figures. However, the corresponding documentation is typically hard to decipher and interpret for inexperienced users, the packages lack abilities for user adjustments and the packages typically require the user to perform all model syntax according to the construct of these packages. The heemod package for example, is a package specifically designed for cost-effectiveness analysis [12]. Though it does provide the option of generating a density plot, this does not generate contours nor does it provide user options such as the plotting of WTP-thresholds. To display results in a density plot using the heemod package, users need to understand and use the heemod package syntax. Another example is the BCEA package which has a variety of graphical capabilities but also requires users to use the specific syntax [13]. An alternative previously described approach to illustrate areas with a specific cumulative probability is the ellipse, for example implemented as a 95% confidence ellipse by Pradelli et al. and as an option in the software suite TreeAge [24,25]. This approach has several weaknesses. First, it assumes the underlying distribution is circular. This would be correct for our normally distributed example but is clearly not suitable for the two

real-world case studies. Our non-parametric density estimation does not rely on this assumption. Second, there is no single or clear method on how to generate the ellipse which potentially limits generalisability. Indeed, the example of Pradelli et al. does not describe the methodology used to generate their ellipse. Third, there is no readily available and generic implementation of the ellipse methodology in for example Excel or R, so this functionality could only be available if the health-economic model is built within a specific proprietary software package.

Our approach to the PSA-ReD plot is specifically designed to combine a density plot with a contour plot in one figure and to be used with any model and any software, as long as the user is able to extract the PSA x- and y-values and save them as an .RData, Microsoft Excel or .csv file which thereafter can be imported into R using our script.

To facilitate the use of our method we provided a step-by-step tutorial on GitHub to generate the PSA-ReD plot based on PSA results from any user's own research This tutorial is designed to also accommodate users with very basic R knowledge. Additionally, generating a variety of PSA-ReD plots is easier and quicker than generating multiple attractive plots in Excel.

For modelers who do not wish to use or explore R, It is possible to generate a 2D histogram with colours within Excel. This approximates the density part of the PSA-ReD plot but lacks the kernel density estimation and contours. It also does not provide the option for adding WTP thresholds nor any of the user options to adjust the figure to make it more visually attractive. An Excel file including the Visual Basics Application syntax can be requested from the authors by any interested reader. We however highly recommend to use R for PSA-ReD generation.

The PSA-ReD script bases the size of the plot exactly on the minimum and maximum values of the PSA iterations in the dataset. This means that four PSA points (or less if they define a corner) lie exactly on the borders of the PSA-ReD figure. As plots usually have some space around the minimum and maximum values, this may make the initial interpretation of the total range slightly harder, but we believe that this yields the best insight into the (distribution of) high density areas as the plot size is kept as small as possible.

The provided R script provides a selection of user options to modify the generated plot. These options and settings are aimed at providing all the functionalities that users of the current scatter plots require. Though experienced R users may be able to further customise the script, novel users are encouraged to apply the options provided in this paper to ensure generalisability of the PSA-ReD plot generated by different users.

CONCLUSION

The proposed PSA-ReD plot facilitates intuitive visual interpretation of information included in PSA results that cannot be interpreted from the traditional scatter plot. Specifically, the PSA-ReD plot provides quantitative information on the relative density of PSA iterations throughout the cost-effectiveness plane and the cumulative probability of PSA iterations within predefined areas of the cost-effectiveness plane. The PSA-ReD plot is particularly useful to identify the location of the highest-density areas, quantify their cumulative density, and to reduce over-emphasis of infrequent PSA iterations. We suggest using the PSA-ReD plot to visualise PSA results in order to benefit interpretation of PSA results of health-economic models.

REFERENCES

1. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM modeling good research practices task force working group-6. Medical Decision Making. 2012;32:722–32.
2. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. Health Economics. 1994;3:95–104.
3. Hatswell AJ, Bullement A, Briggs A, Paulden M, Stevenson MD. Probabilistic Sensitivity Analysis in Cost-Effectiveness Models: Determining Model Convergence in Cohort Models. PharmacoEconomics. Springer International Publishing; 2018;36:1421–6.
4. Zorginstituut Nederland (ZIN), Staal P, Heymans J, Ligtenberg G, Derksen J, Couwenbergh B. Beoordeling stand van de wetenschap en praktijk. 2014;90.
5. NICE. Guide to the processes of technology appraisal Process and methods [PMG19].
6. Adalsteinsson E, Toumi M. Benefits of probabilistic sensitivity analysis – a review of NICE decisions. Journal of Market Access & Health Policy. 2013;1:21240.
7. Baio G, Dawid AP. Probabilistic sensitivity analysis in health economics. Statistical Methods in Medical Research. 2015;24:615–34.
8. Drummond MF, Sculpher MJ, Torrance GW, O'Brien, Stoddart BJ and GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press. Oxford University Press; 2005.
9. William S. Cleveland RM. The Many Faces of a Scatterplot. Theory and Method. 2012;79:807–22.
10. Gleicher AM and M. Splatterplots: Overcoming Overdraw in Scatter Plots. NIH Public Access. 2013;19:1526–38.
11. R Core team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
12. Filipovic-Pierucci A, Zarca K, Durand-Zaleski I. Markov Models For Health Economic Evaluation Modelling In R With The Heemod Package. Value in Health. 2016. p. A369.
13. Baio G, Heath A, Berardi A. Bayesian Cost-Effectiveness Analysis with the R package BCEA. Switzerland: Springer Nature; 2017.
14. Geenen JW. GitHub repository for the PSA-ReD scripts and manuals. [Internet]. 2019 [cited 2019 Jul 19]. Available from: https://github.com/joostgeenen/PSA-ReD
15. April S. Density Estimation for Statistics and Data Analysis Chapter 1 and 2. Most. 2003. p. 8.
16. RStudio. RStudio: Integrated development environment for R (Version 0.97.311). J. Wildl. Manage. Boston, MA: RStudio, Inc.; 2011. p. 1753–66.
17. Google. Googles R Style Guide. 2013. p. 1–6.
18. Free Software Foundation Inc. GNU General Public License v3.0. 2007.
19. Venables WN, Ripley BD. Modern Applied Statistics with S. New York: Springer; 2002.
20. Silverman BW. Density Estimation for Statistics and Data Analysis. Chapman and Hall/CRC; 1986.
21. Geenen J, Baranova E, Asselbergs F, de Boer A, Maitland-van der Zee A, Hovels A. Early HTA in Pharmacogenomics: A Case Example in Cardiovascular Drugs. Value in Health. 2016;19:A636.
22. van Nuland M, Vreman RA, Ten Ham RMT, de Vries Schultink AHM, Rosing H, Schellens JHM, et al. Cost-effectiveness of monitoring endoxifen levels in breast cancer patients adjuvantly treated with tamoxifen. Breast Cancer Research and Treatment. 2018;
23. Davies HM, Hope K. Methods of Multivariate Analysis. 3rd ed. The Mathematical Gazette. Wiley; 2007.
24. Pradelli L, Povero M, Muscaritoli M, Eandi M. Updated cost-effectiveness analysis of supplemental glutamine for parenteral nutrition of intensive-care patients. European Journal of Clinical Nutrition. 2015;69:546–51.

25. TreeAge Software, Inc. TreeAge Pro 2017 R2 User's Manual. 2017;795.
26. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016.
27. Hocking T. directlabels: Direct Labels for Multicolor Plots [Internet]. 2018.
28. Auguie B. gridExtra: Miscellaneous Functions for "Grid" Graphics [Internet]. 2017.
29. Wickham H. Reshaping Data with the reshape Package. Journal of Statistical Software. 2007;21:1–20.

APPENDICES

Appendix 1: Figures of Components of PSA-ReD

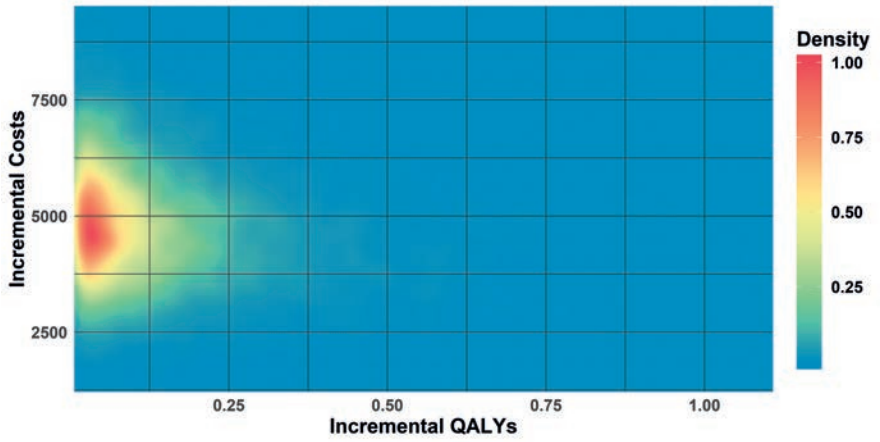


Figure A. Figure with only the density rendered, generated using 1000 bins and 10,000 PSA iterations.

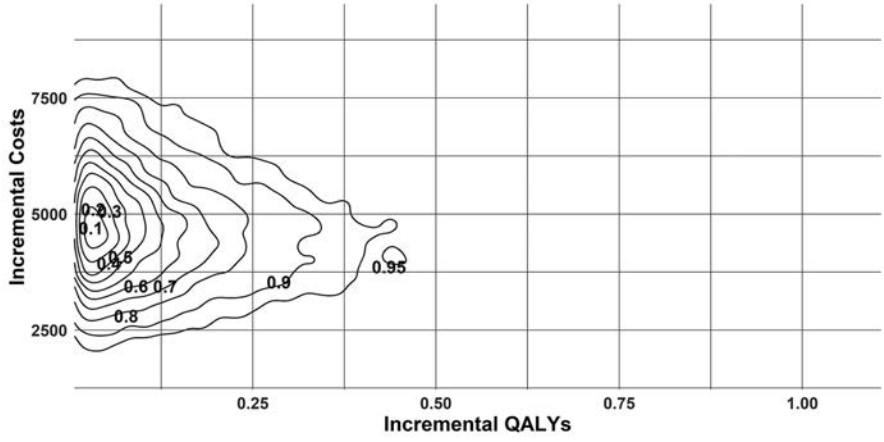


Figure B. Figure with only the contours rendered, generated using 1000 bins and 10,000 PSA iterations.

Appendix 2: Detailed information on user modifications and the hard- and software used for PSA-ReD development

Hardware and software used

Various packages exist for rendering plots and figures. We have selected ggplot2 as renderer as it is free, open source, very versatile, easy to interpret and very well supported [26]. Besides ggplot2 version 3.0.0, the following R packages were used: MASS version 7.3-50 to perform kde, directlabels version 2018.05.22 to add contour labels, grid version 3.5.1 and gridExtra version 2.3 for conveniently displaying and zooming the figures and reshape2 version 1.4.3 for data preparation [19,27-29]. The supplied script handles installing and/or loading of these packages. For our analyses, we used a standard consumer grade personal computer (Dell Optiplex 9020, Intel® Core™ i5-4590 CPU @ 3.30 GHz, 8.00 GB Random Access Memory (RAM), 500GB 7200RPM Hard Disk Drive, Windows 10 Enterprise © 2017 Microsoft Corporation).

Zooming

As the initial plot range is chosen on the maximum and minimum x and y values, this implies that some data points are removed when zooming in. We do however think that adding this option is justified as, when using a very large PSA sample (i.e. 100,000). This is because a few extreme values will increase the plot size and may limit they interpretability of the area of interest. The implementation of zooming is chosen so that the plot itself is trimmed instead of trimming the underlying data. This conserves the underlying data structure and is analogous to simply selecting a small part of an existing larger figure. Users can use the zoom functionality by specifying a specific x and y range.

When zooming, users have to pay attention to whether the contour labels are clipped from the plot area. We have supplied a setting that can deal with this issue but it does require the user to evaluate their zoomed figure and choose the most appropriate figure. In the scenario analyses, we have shown the influence of this setting.

Saving

Figures can be saved using functions that are supplied in the script. The path and filename can be set by the user. Besides these settings, one can choose the resolution (as dpi) with which the figure is saved. We specifically chose the dpi as setting to alter the eventual figure size, as it does not influence the relative sizes of the various figure elements.

When saving the figure, one can choose the figure to be saved exactly like it appears in the Rstudio "Plots" panel. As different users and computers could have differently sized "Plots" panels, we supply a means to specify the required width and height of the figures. This ensures that different users can easily generate the same PDP when using the same data and settings.

Plotting Willingness To Pay threshold

Willingness to pay (WTP) thresholds are regularly plotted in the traditional PSA scatter plot. We have built a feature that allows users to specify one or more WTP thresholds. If one wishes to plot these thresholds, the WTP thresholds should be stored in the vector "WTP.thresholds" which is present in the script.

Appendix 3: Influence of bin size on PSA-ReD plots

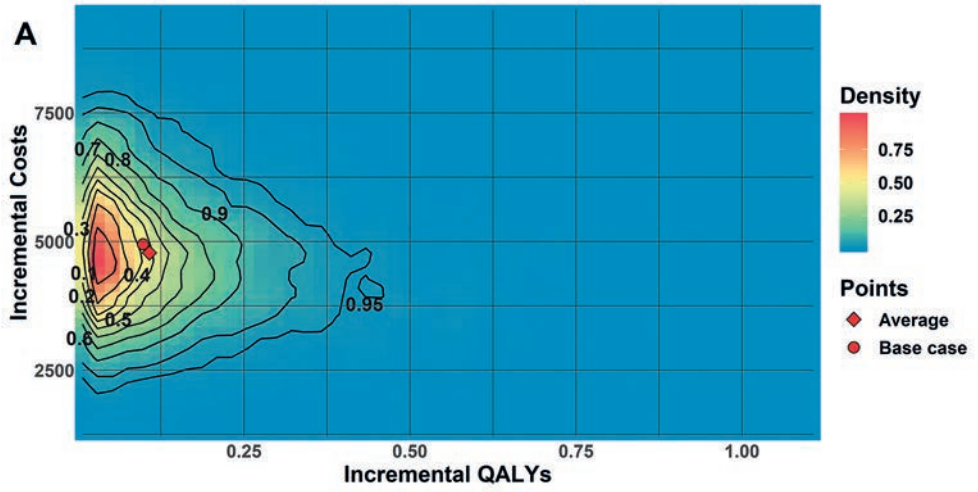


Figure a. Bin size = 50, 10,000 PSA iterations

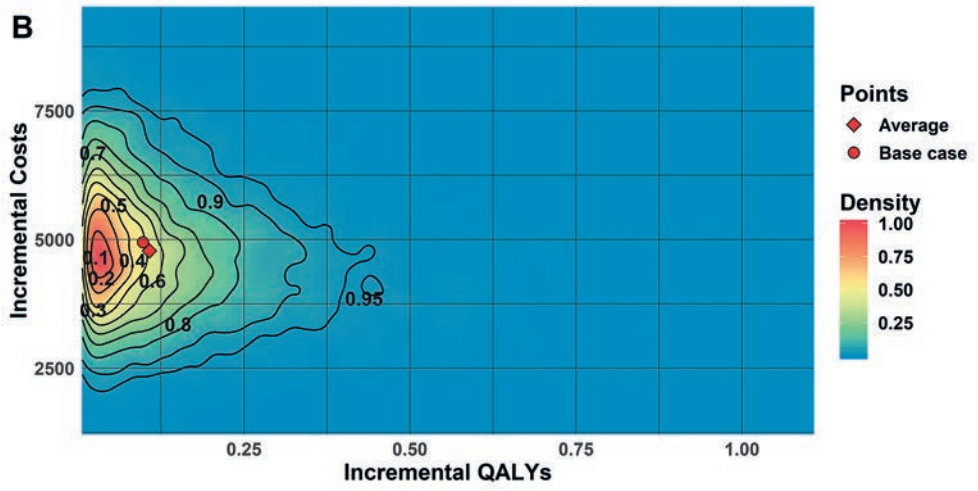


Figure b. Bin size = 500, 10,000 PSA iterations

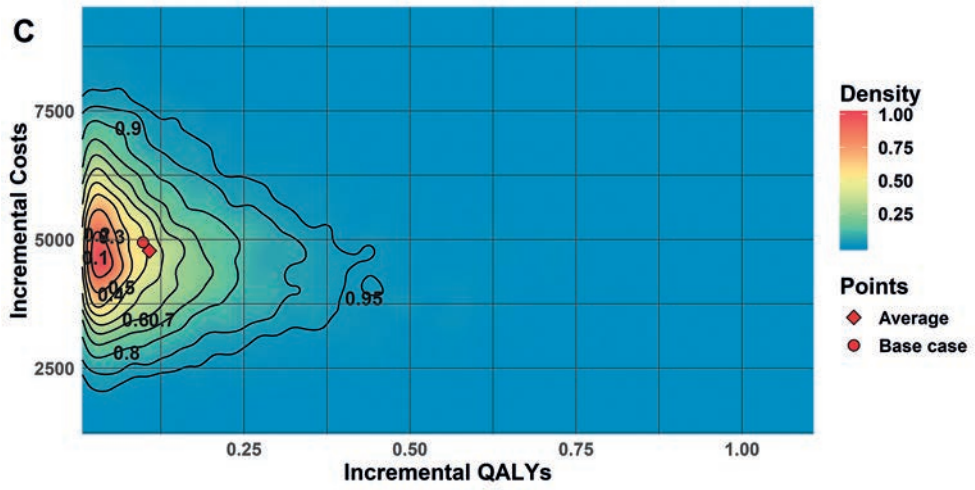


Figure c. Bin size = 1000, 10,000 PSA iterations

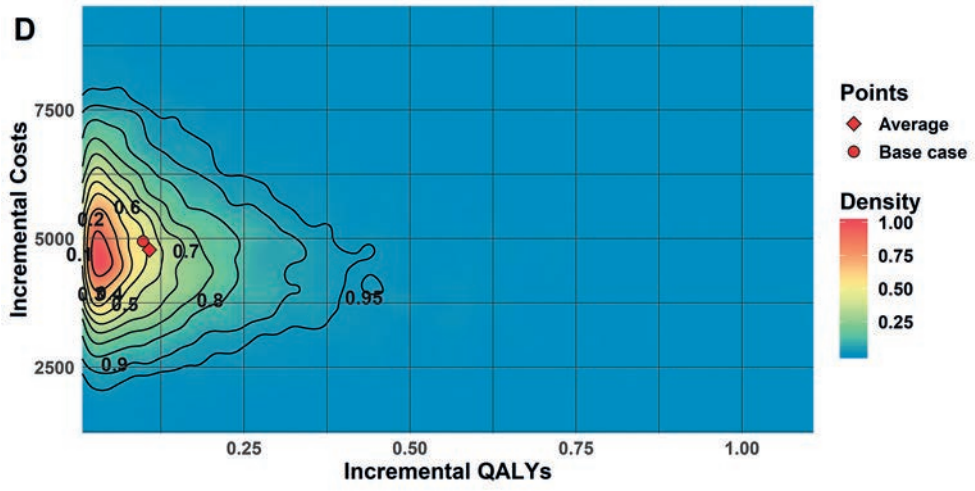


Figure d. Bin size = 2000, 10,000 PSA iterations

Appendix 4: Influence of number of PSA iterations on PSA-ReD plots

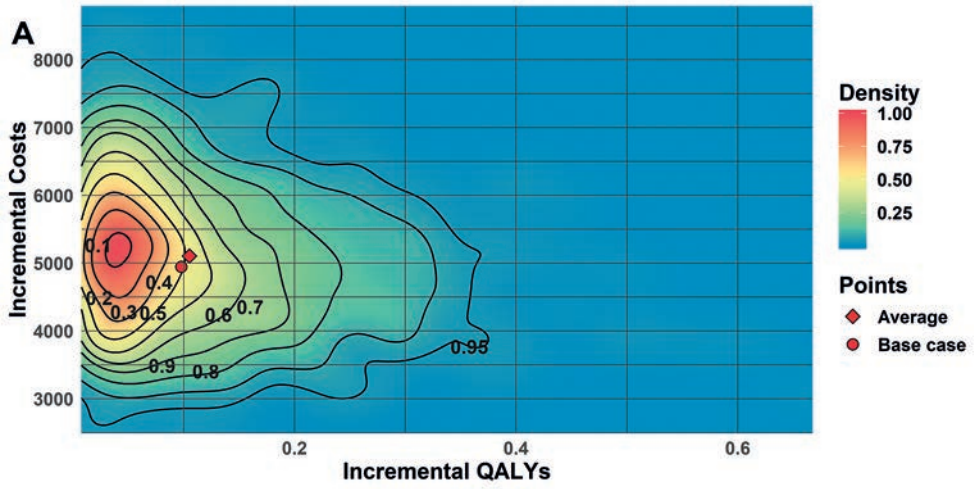


Figure a. Bin size = 1000, 1000 PSA iterations

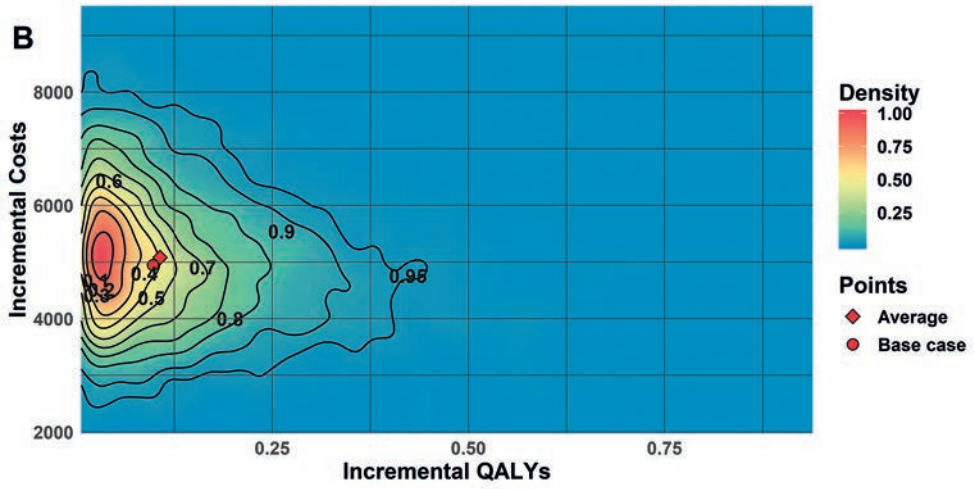


Figure b. Bin size = 1000, 5000 PSA iterations

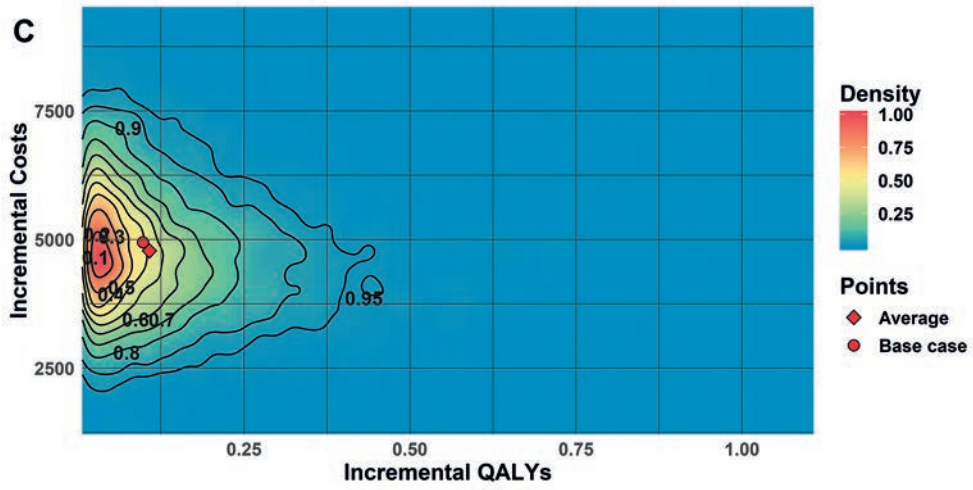


Figure c. Bin size = 1000, 10,000 PSA iterations

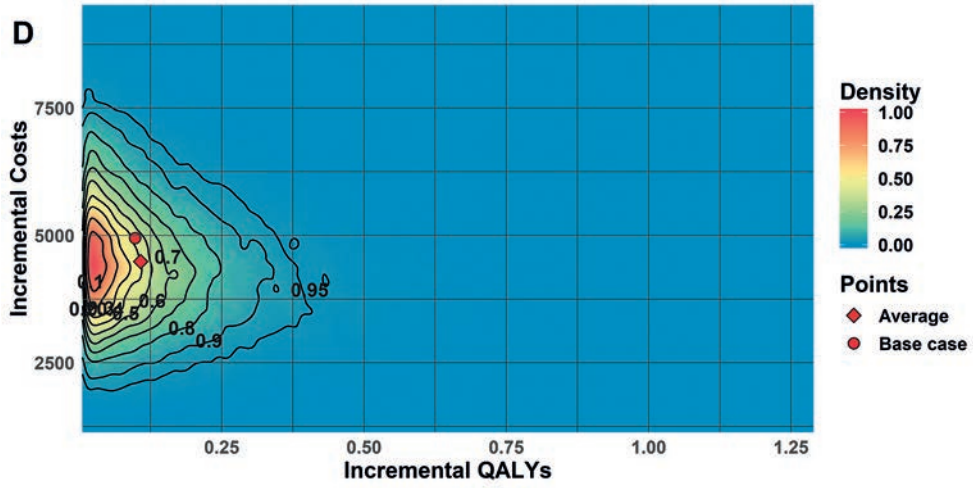


Figure d. Bin size = 1000, 100,000 PSA iterations

Appendix 5: Zoom functionality and influence of clip argument

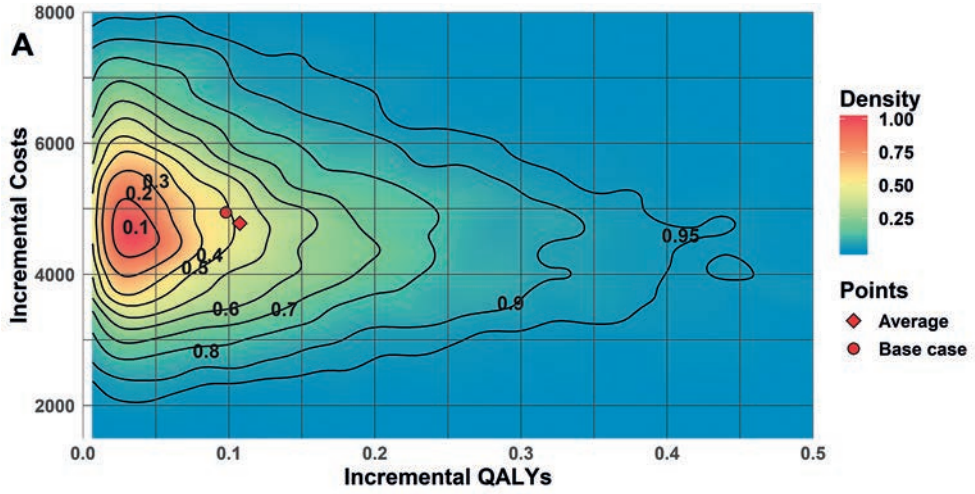


Figure a. zoomed PSA-ReD plot, clip = FALSE. Bin size = 1000, 10,000 PSA iterations

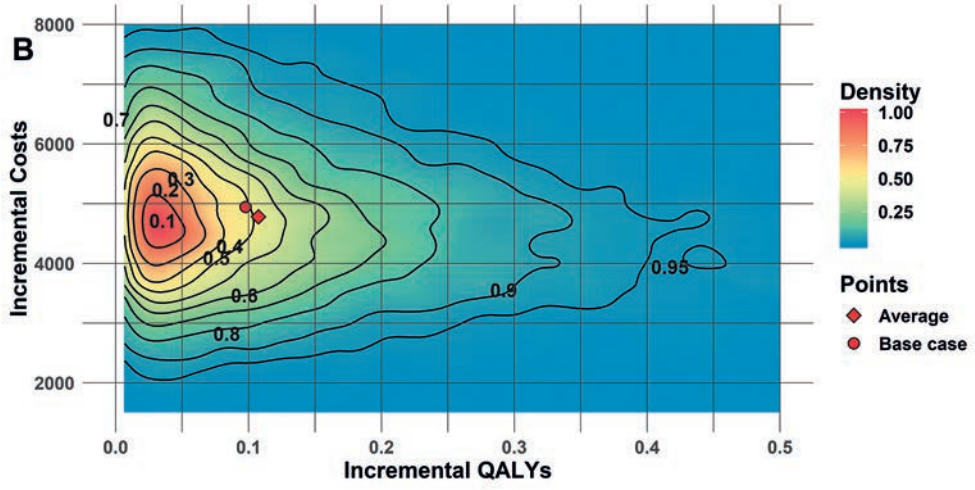


Figure b. zoomed PSA-ReD plot, clip = TRUE. Bin size = 1000, 10,000 PSA iterations

Appendix 6: Script for generated PSA-ReD plots

```
# PSA-ReD Plot Generator v1.0.1. Use this script to make your own PSA-ReD plots.
# Copyright (C) 2019, Joost Geenen
# This program is free software: you can redistribute it and/or modify it under
# the terms of the GNU General Public License as published by the
# Free Software Foundation, either version 3 of the License,
# or (at your option) any later version.

# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
# See the GNU General Public License for more details.

# You should have received a copy of the GNU General Public License along with
# this program. If not, see <https://www.gnu.org/licenses/>.

### -------------- Do not change the part in between / below -------------- ###
### ----------------------------------------------------------------------- ###

# The code below installs and/or loads the required packages.

if (!require('ggplot2')) {
  install.packages("ggplot2", dependencies = TRUE)
}
library('ggplot2')

if (!require('MASS')) {
  install.packages("MASS", dependencies = TRUE)
}
library('MASS')

if (!require('directlabels')) {
  install.packages("directlabels", dependencies = TRUE)
}
library('directlabels')

if (!require('grid')) {
  install.packages("grid", dependencies = TRUE)
}
library('grid')
```

```
if (!require('gridExtra')) {
  install.packages("gridExtra", dependencies = TRUE)
}
library('gridExtra')

if (!require('reshape2')) {
  install.packages("reshape2", dependencies = TRUE)
}
library('reshape2')

# The code below loades custom functions

ProcessContourData <- function(kde.data) {
  # Processess kde data for plotting the contours.
  #
  # Args:
  #  kde.data: a list containing the kernel density data,
  #  which is generated using the 'data' which the user loaded.
  #
  # Returns:
  #  A dataframe containing the cumulative density per x and y.

  kde.dx <- diff(kde.data$x[1:2]) # Width of 1 bin (x-axis)
  kde.dy <- diff(kde.data$y[1:2]) # Height of 1 bin (y-axis)
  kde.sz <- sort(kde.data$z)      # Sorted density per bin
  kde_c1 <- cumsum(kde.sz) * kde.dx * kde.dy # Sorted density per bin area

  dimnames(kde.data$z) <- list(kde.data$x, kde.data$y) # Density per x and y
  kde_dc <- melt(kde.data$z)      # Melt list

  # Interpolate the density values to a range of [1,0]
  # This yields cumulative probability as density is sorted from high to low
  kde_dc$contour.levels <- approx(kde.sz, 1 - kde_c1, kde_dc$value)$y

  # Convert to data.frame
  plot.data.contour <- data.frame(x = kde_dc[, 1],
                                  y = kde_dc[, 2],
                                  contour.levels = kde_dc[, 4])
  return(plot.data.contour)
}
```

```
GenerateNormalisedDensity <- function(kde.data) {
  # Normalises the density to 1 and generates a matrix for plotting the density
  #
  # Args:
  #  kde.data: a list containing the kernel density data,
  #  which is generated using the 'data' which the user loaded.
  #
  # Returns:
  #  A dataframe containing the relative density per x and y.
  x <- kde.data$x
  y <- kde.data$y

  df.density   <- expand.grid(X = x, Y = y)
  df.density$Z <- as.numeric(unlist(kde.data$z))

  df.density.normalised   <- df.density
  df.density.normalised$Z <- df.density.normalised$Z *
    (1 / max(df.density.normalised$Z))

  return(df.density.normalised)
}

GeneratePlot <- function(df.density.normalised,
                         legend.title,
                         plot.data.contour,
                         contour.levels,
                         font.size,
                         font.face,
                         font.family,
                         x.axis.title,
                         y.axis.title,
                         x.range,
                         y.range,
                         clipping,
                         extend.panel,
                         WTP.thresholds,
                         basecase,
                         average.PSA) {
  # Generates the ggplot object that can be plotted.
  #
```

```r
# Args:
#   df.density.normalised: A dataframe consisting of the normalised relative
#                          density per x and y.
#   legend.title:          The title of the legend.
#   plot.data.contour:     A dataframe containing the cumulative density
#                          per x and y.
#   contour.levels:        A vector containing the specified contour levels.
#   font.size:             The font size of characters in the plot
#   font.face:             The font face (ie, bold, italic) of the characters
#                          in the plot
#   font.family:           The font family (ie, sans) of the characters
#                          in the plot
#   x.axis.title:          The x-axis title.
#   y.axis.title:          The y-axis title.
#   x.range:               The range of values on the x-axis,
#                          as a vector (min, max)
#   y.range:               The range of values on the y-axis,
#                          as a vector (min, max)
#   clipping:              A Boolean specifying whether contour labels
#                          may be clipped from the plot area.
#   extend.panel:          A Boolean specifying whether the grid panel
#                          may be extended.
#   WTP.thresholds:        A vector containing WTP thresholds to draw.
#   basecase:              A vector as: (incremental QALYs, incremental costs)
#   average.PSA:           A vector as: (incremental QALYs, incremental costs)
#
# Returns:
#   A gtable object with the plot data.

if (missing(x.range)) {
  x.range <- c(min(df.density.normalised$X), max(df.density.normalised$X))
}

if (missing(y.range)) {
  y.range <- c(min(df.density.normalised$Y), max(df.density.normalised$Y))
}

if (!is.null(average.PSA) | !is.null(basecase) | !is.null(WTP.thresholds)) {
  density.barheight <- NULL
} else {
  density.barheight <- 15
```

```
}

element.list <- list()
label.list1  <- list("far.from.others.borders",
                     "calc.boxes",
                     "enlarge.box",
                     rot = 0,
                     hjust = 0,
                     vjust = 0,
                     box.color = NA,
                     fill = "transparent",
                     "draw.rects")

if (!is.null(WTP.thresholds)) {
  # Initialise segment dataframe
  segment.df <- as.data.frame(matrix(0, ncol = 6,
                                     nrow = length(WTP.thresholds)))
  min.x <- min(df.density.normalised$X)
  min.y <- min(df.density.normalised$Y)
  max.x <- max(df.density.normalised$X)
  max.y <- max(df.density.normalised$Y)
  colnames(segment.df) <- c("segment,start.x", "segment.end.x",
                            "min.y", "max.y", "WTP Threshold", "i")
  # Calculate the coordinates of each WTP segment
  for(i in 1:length(WTP.thresholds)) {
    segment.start.x <- min.y / WTP.thresholds[i]
    segment.end.x <- max.y / WTP.thresholds[i]

    if (segment.end.x > max.x){
      segment.end.x <- max.x
      max.y <- WTP.thresholds[i] * max.x
    }
    if(segment.start.x < min.x) {
      segment.start.x <- min.x
      min.y <- WTP.thresholds[i] * min.x
    }

    # For when a segment is entirely out of the plot window
    if( segment.start.x > max.x){
      segment.start.x <- min.x
      segment.end.x <- min.x
```

```
      min.y <- min(df.density.normalised$Y)
      max.y <- min(df.density.normalised$Y)
    }
    segment.df[i,] <- c(segment.start.x, segment.end.x, min.y, max.y,
                        WTP.thresholds[i], i)
  }
  WTP <- factor(segment.df[, 6], labels = as.character(segment.df[, 5]))
}

# Generate plot using ggplot() call
plot.contour <- ggplot(data = df.density.normalised,
                       aes(x = X, y = Y, z = Z)) +
  geom_tile(aes(fill = Z), alpha = 1) +
  scale_fill_distiller(name      = legend.title,
                       palette   = "Spectral",
                       direction = -1,
                       guide     = "colourbar") +
  theme_minimal()  +
  geom_contour(aes(z  = plot.data.contour$contour.levels),
               breaks = rev(contour.levels),
               size   = 0.5,
               colour = "black") +
  theme(panel.grid.major = element_line(colour = "gray30", size = 0.25),
        panel.grid.minor = element_line(colour = "gray30", size = 0.25),
        panel.ontop      = TRUE,
        text             = element_text(size = font.size,
                                        family = font.family,
                                        face   = font.face),
        legend.spacing.y = unit(0.15, "cm")) +
  labs(x = x.axis.title,
       y = y.axis.title) +
  guides(fill = guide_colourbar(barheight = density.barheight))

if (!is.null(WTP.thresholds)) {
  element.list <- append(element.list,
                         geom_segment(data = segment.df,
                                      aes(x    = segment.df[, 1],
                                          xend = segment.df[, 2],
                                          y    = segment.df[, 3],
                                          yend = segment.df[, 4],
                                          linetype = WTP),
```

```
                                    color = "black",
                                    size  = 0.5,
                                    inherit.aes = F))
}

if (!is.null(average.PSA) & !is.null(basecase)) {
  point.df <- rbind.data.frame(average.PSA, basecase)
  point.df <- cbind(point.df, c("Average", "Base case"))
  colnames(point.df) <- c("x", "y", "Type")
  points.data <- factor(c(1, 2), labels = as.character(point.df$Type))

  element.list <- append(element.list, geom_point(inherit.aes = F,
                                                  data = point.df,
                                                  aes(x = x,
                                                      y = y,
                                                      group = "Type",
                                                      shape = points.data),
                                                  color = "black",
                                                  fill  = "red",
                                                  size  = 3))
  element.list <- append(element.list,
                         scale_shape_manual(name = "Points",
                                            values = c(23, 21),
                                            labels = c("Average",
                                                       "Base case")))
} else if (!is.null(average.PSA) & is.null(basecase)) {
  point.df <- rbind.data.frame(average.PSA)
  point.df <- cbind(point.df, c("Average"))
  colnames(point.df) <- c("x", "y", "Type")
  points.data <- factor(c(1), labels = as.character(point.df$Type))

  element.list <- append(element.list, geom_point(inherit.aes = F,
                                                  data = point.df,
                                                  aes(x = x,
                                                      y = y,
                                                      group = "Type",
                                                      shape = points.data),
                                                  color = "black",
                                                  fill  = "red",
                                                  size  = 3))
  element.list <- append(element.list,
```

```
                        scale_shape_manual(name = "Points",
                                           values = c(23),
                                           labels = c("Average")))
} else if (is.null(average.PSA) & !is.null(basecase)) {
  point.df <- rbind.data.frame(basecase)
  point.df <- cbind(point.df, c("Base case"))
  colnames(point.df) <- c("x", "y", "Type")
  points.data <- factor(c(1), labels = as.character(point.df$Type))

  element.list <- append(element.list, geom_point(inherit.aes = F,
                                                  data = point.df,
                                                  aes(x = x,
                                                      y = y,
                                                      group = "Type",
                                                      shape = points.data),
                                                  color = "black",
                                                  fill  = "red",
                                                  size  = 3))
  element.list <- append(element.list,
                         scale_shape_manual(name = "Points",
                                           values = c(21),
                                           labels = c("Base case")))
}

element.list <- append(element.list,
                       geom_dl(aes(label = ..level..,
                                   x = plot.data.contour$x,
                                   y = plot.data.contour$y,
                                   z = plot.data.contour$contour.levels,
                                   fontface = "bold"),
                               inherit.aes = F,
                               color = "gray15",
                               cex = 0.75,
                               method = label.list1,
                               stat = "contour",
                               breaks = rev(contour.levels)))

if (clipping == FALSE) {
  # Limits rendering to given coordinates,
  # does not exclude (clip) data during plot generation
  element.list <- append(element.list, coord_cartesian(xlim = x.range,
```

```
                                                    ylim = y.range,
                                                    expand = FALSE))
  } else {
    # Clip datapoints to fall within a range
    element.list <- append(element.list, xlim(x.range))
    element.list <- append(element.list, ylim(y.range))
  }

  # Add elements stored in element.list
  plot.contour <- plot.contour + element.list

  if (extend.panel == TRUE) {
    # Extend the plot panel outside of the density area
    # so that contour labels are not partially cropped.
    plot.contour <- ggplot_gtable(ggplot_build(plot.contour))
    plot.contour$layout$clip[plot.contour$layout$name == "panel"] <- "off"
  }
  return(plot.contour)
}

message("Copyright (C) 2019, Joost Geenen\n",
        "This program comes with ABSOLUTELY NO WARRANTY.\n",
        "This is free software, and you are welcome to redistribute it",
        " under certain conditions.\n",
        "You should have received a copy of the GNU",
        " General Public License along with this program.\n",
        "If not, see <https://www.gnu.org/licenses/>.")
### ------------------------------------------------------------------------ ###
### -------------- Do not change the part in between / above -------------- ###

### ------------------------------------------------------------------------ ###
### ------------------ Setup your data and plot settings ------------------ ###

# Your raw data should have the following characteristics:

# - Saved as a .csv file
# - The first column should be incremental effects (QALYs / LYs / etc)
# - The second column should be incremental costs

# Then, prepare to load your data:
```

```
# - set the filename of your raw datafile, do not forget the .csv
# - set whether your data has a header (ie, column names)
# - set the type of decimal seperator (, or . within " ")
# - set the column separator used for your .csv file (within " ").
# You can see the seperator when opening your csv file with, for example, Excel.
filename   <- "YourData.csv"
has.header <- FALSE
decimal.separator <- ","
column.separator  <- ";"

# We recommend 10.000 rows, although somewhere between
# 1.000 and 100.000 will produce proper figures.
# More than 10.000 will slow various computations whilst
# it does not add information to your plots.
# We therefore recommend a maximum number of rows of 10.000
# If you would like to limit your rows to this number,
# set the 'limit_rows' variable to TRUE and specify
# the number of rows in number_rows:
# If your data has less than 10.000 rows, leave this setting at FALSE.
limit.rows <- TRUE
number.rows <- 10000

# The number of bins determines the granularity of the plot.
# Warning: Larger bin numbers require more RAM.
# 1000 bins typically produces images without pixelation.
# More bins does not provide better images whilst it increases computation time.
# Less bins (eg, 100, 500), provide decent figure with limited RAM usage and
# reduced computation time.
# We therefore recommend using 1000 bins.

# set bin.number here
bin.number <- 1000

# You can specify the following plot characteristics:
# - specify desired contour levels.
# - set as, for example: contour.levels <- c(0.9, 0.5, 0.1)
contour.levels <- c(0.9, 0.5, 0.1)

# - Specify WTP thresholds to add to the plot.
#   Example: WTP.thresholds <- c(30000, 80000)
#   If you don't want to add these, run "WTP.thresholds <- c()"
```

```
WTP.thresholds <- c(20000, 50000, 80000)

# - If you want to add the base-case results to your plot,
#   set as: "basecase <- c(<basecase incr. QALYs>, <basecase incr. costs>)
#   Example: basecase <- c(0.5, 2000)
#   If you don't want to add this, set: "basecase <- c()"
basecase <- c()

# - If you want to add a marker with the average of the PSA iterations,
#   set add.average.PSA <- TRUE
add.average.PSA <- TRUE

# - Set Font Title for the Plot
x.axis.title <- "Incremental QALYs"
y.axis.title <- "Incremental Costs"
font.size    <- 14

# - Set font characteristics for the plot
font.face    <- "bold"         ## choose bold or plain or italix
font.family  <- "sans"         ##
legend.title <- "Density"      ## Set your desired legend title

# Now, proceed running the following parts,
# parts within a 'do not change the part in between / below (or above)
# should just be run but do not require input from the user.

### -------------------- End of data and plot settings -------------------- ###
### ----------------------------------------------------------------------- ###

### -------------- Do not change the part in between / below -------------- ###
### ----------------------------------------------------------------------- ###

# Load data
data <- read.csv(file   = filename,
                 header = has.header,
                 dec    = decimal.separator,
                 sep    = column.separator)

print(paste("your data has", as.character(nrow(data)), "rows"))
if (limit.rows == TRUE & nrow(data) > number.rows){
  data <- data[1:number.rows, ]
```

```
}

# Perform kernel density estimation
kde.data <- kde2d(data[, 1], data[, 2], n = bin.number)
plot.data.contour     <- ProcessContourData(kde.data)
df.density.normalised <- GenerateNormalisedDensity(kde.data)

# Calculate average PSA results
if (add.average.PSA == TRUE) {
  average.PSA <- c(mean(data[, 1]), mean(data[, 2]))
} else {
  average.PSA <- NULL
}

# Generate the ggplot plot object
contour.plot <- GeneratePlot(df.density.normalised,
                             legend.title,
                             plot.data.contour,
                             contour.levels,
                             font.size,
                             font.face,
                             font.family,
                             x.axis.title,
                             y.axis.title,
                             clipping = FALSE,
                             extend.panel = TRUE,
                             WTP.thresholds = WTP.thresholds,
                             basecase = basecase,
                             average.PSA = average.PSA)

# Display the plot
grid.newpage()
grid.draw(contour.plot)
### ------------------------------------------------------------------------ ###
### -------------- Do not change the part in between / above -------------- ###

### -------------------- Options for Saving your plot ---------------------- ###
### ------------------------------------------------------------------------ ###

# If you wish to save this plot, enter the required filename and settings here
```

```
# Set the file name of the new plot.
# WARNING: It will overwrite files / plots with the same name!
filename <- "Figurename.PNG"

# Set plot dpi (resolution)
dpi <- 600

# Use size of the "Plots" Panel in Rstudio?
# "TRUE" will guarantee that your figure is saved exactly like you see it now.
# Another user, with a differently sized "Plots" panel,
# will then however get a different plot using the same data
# Selecting "FALSE" allows you to specifiy your own, fixed figure size.
# This figure will be different from the one you see in your "Plots" panel
# But you will then reproduce this exact figure using the size values values.
use.my.panel.size <- FALSE

# If you have set use.my.panel.size to FALSE, set the plot size in inches:
plot.width  <- 8.47
plot.height <- 4.25

# Saving is now set-up, run the part below.

### -------------- Do not change the part in between / below -------------- ###
### ----------------------------------------------------------------------- ###
if (use.my.panel.size == TRUE) {
  ggsave(contour.plot, filename = filename, dpi = dpi)
} else {
  ggsave(contour.plot, filename = filename, dpi = dpi,
         width = plot.width, height = plot.height, units = "in")
}
### ----------------------------------------------------------------------- ###
### -------------- Do not change the part in between / above -------------- ###

### -------------- Options for zooming in on a specific area -------------- ###
### ----------------------------------------------------------------------- ###

# If you wish, you can zoom in on a particular area of the plot.
# Set x- and y range as c(min, max)
x.range <- c(0, 0.5)
y.range <- c(1500, 8000)
```

```
# To prevent the contour labels from clipping of the sides,
# the drawing panel is extended.
# This however, has some consequences when zooming.

# Select ONE the following and then go to “Now, run the part below”:

# 1 If your contour labels all lie within the plot, set:
clip <- FALSE

# If your contour labels (partly) fall outside the plot, you have 2 choices:
# 1: Clip them off, which may yield a pretier plot. set:
clip <- FALSE

# 2: Extend panel and clip data to fit the panel, this may yield a pretier plot.
# Set:
clip <- TRUE

# Now, run the part below.

### -------------- Do not change the part in between / below -------------- ###
### ---------------------------------------------------------------------- ###
if (clip == FALSE) {
  extend.panel <- FALSE
  clipping     <- FALSE
} else{
  extend.panel <- TRUE
  clipping     <- TRUE
}

zoomed.plot <- GeneratePlot(df.density.normalised,
                            legend.title,
                            plot.data.contour,
                            contour.levels,
                            font.size,
                            font.face,
                            font.family,
                            x.axis.title,
                            y.axis.title,
                            x.range,
                            y.range,
                            clipping,
```

```
                              extend.panel,
                              WTP.thresholds,
                              basecase,
                              average.PSA)
grid.newpage()
grid.draw(zoomed.plot)
### ------------------------------------------------------------------------ ###
### ---------- Do not change the part in between / above ------------------ ###

### -------------- Options for zooming in on a specific area -------------- ###
### ------------------------------------------------------------------------ ###

# If you wish to save this plot, enter the required filename and settings here

# Set the file name of the new plot.
# WARNING: It will overwrite files / plots with the same name!
filename <- "Figurename.PNG"

# Set plot dpi (resolution)
dpi <- 600

# Use size of the "Plots" Panel in Rstudio?
# "TRUE" will guarantee that your figure is saved exactly like you see it now.
# Another user, with a differently sized "Plots" panel, will then however
# get a different plot using the same data
# Selecting "FALSE" allows you to specifiy your own, fixed figure size.
# This figure will be different from the one you see in your "Plots" panel
# But you will then reproduce this exact figure using the size values values.
use.my.panel.size <- FALSE

# If you have set use.my.panel.size to FALSE, set the plot size in inches:
plot.width  <- 8.47
plot.height <- 4.25

# Saving is now set-up, run the part below.

### -------------- Do not change the part in between / below -------------- ###
### ------------------------------------------------------------------------ ###

if (use.my.panel.size == TRUE) {
  ggsave(zoomed.plot, filename = filename, dpi = dpi)
```

```
} else {
  ggsave(zoomed.plot, filename = filename, dpi = dpi,
         width = plot.width, height = plot.height, units = "in")
}
### ------------------------------------------------------------------------ ###
### -------------- Do not change the part in between / above -------------- ###
```

5

Integrating cost-effectiveness, affordability and uncertainty: a new conceptual framework

5.1

Joint appraisal of cost-effectiveness, affordability and associated uncertainty in reimbursement decision-making: an integrated conceptual framework combining net monetary benefit and a dynamic willingness-to-pay threshold

Joost W. Geenen, Cornelis Boersma, Rick A. Vreman,
Saskia Knies, Olaf H. Klungel, Anke M. Hövels

Manuscript in preparation for submission.

ABSTRACT

Objectives

Cost-effectiveness and affordability, typically quantified as Incremental Cost-Effectiveness Ratio (ICER) and Budget Impact (BI), are usually appraised separately to inform reimbursement decisions. Generally, uncertainty in ICER and BI are also assessed separately. Furthermore, evidence suggests that Willingness to Pay thresholds, which currently are usually static, should be more dynamic to properly include opportunity costs. We aim to provide a conceptual framework for united appraisal of BI, WTP and ICERs and their associated uncertainty where WTP is dynamic and influenced by BI.

Methods

We selected the lung cancer drug nivolumab as a case study. We use three methods to quantify a potential relationship between WTP and BI, 1) a method based on a historical real-world reimbursement decision, an arbitrary method and a method based on a paper describing healthcare displacement in the Netherlands. We adapt Net monetary benefit (NMB) to a societal NMB (pNMB) and use this as outcome. pNMB can be calculated from WTP, ICER, BI and average treatment cost per patient. ICER, BI and treatment cost per patient, including uncertainty distribution, are adopted from the Dutch nivolumab reimbursement dossier.

Results

A fixed WTP and the dynamic WTP method based on a study on displacement in Dutch healthcare yielded results where only the ICER determines whether pNMB is positive or negative. In this case, only the ICER determines the reimbursement decision and yielded a reimbursement likelihood of 63% and 54%, respectively. When using a method with a stronger relationship between BI and WTP, the sign of pNMB and therefore the reimbursement decision was simultaneously influenced by BI and the ICER.

Conclusions

We have shown that pNMB can combine affordability and cost-effectiveness into a single metric and thus a single decision. The existence of an explicit relationship between BI and WTP is a prerequisite. This pNMB approach enables decision-makers to identify (combinations of) threshold values for the ICER and BI that are required, thereby potentially leading to improved decision-making.

INTRODUCTION

Cost-effectiveness and affordability, typically quantified as Incremental Cost-Effectiveness Ratio (ICER) and Budget Impact (BI), are two distinct aspects that are typically appraised separately to inform a reimbursement decision [1,2]. Crucially, uncertainty exists in the quantification of the ICER as well as BI and many different types of Managed Entry Agreements (MEA) have been designed to limit uncertainty (and related risk to a payer) in either affordability, cost-effectiveness or both [3–5]. Work has been undertaken to integrate the two concepts of affordability and cost-effectiveness but so far, none of these approaches have explicitly included a joint appraisal of uncertainty in both ICER and BI [1,6].

Budget Impact Analysis (BIA) is required for reimbursement applications in many jurisdictions [7–9]. Whilst submitting a BIA is often mandatory, the role of BI estimates in decision-making is less clear or less formal than, for example, the role of cost-effectiveness [10–13]. Many recent cases, (e.g. new drug introductions for hepatitis C), have however shown that BI can be a crucial and even a decisive factor in reimbursement decisions [7,11,14–19].

In a healthcare system with fixed budgets, new innovations can only be funded by savings or disinvesting in other care and thus cause displacement [1,8,20]. In healthcare systems with less restricted budgets, resources are still not unlimited so at least some opportunity costs will exist and decision-makers are still likely to prefer lower BI over higher BI [8,11]. Research shows that a high and / or highly uncertain BI is a potential risk to decision makers and that they are then more likely to limit reimbursement or to issue a type of MEA [9,14,19,21].

These displacement effects and opportunity costs should, at least to some degree, be reflected in The Willingness to Pay (WTP) threshold [1,8,9,17,20]. The WTP is typically used to reflect the maximum amount a decision-maker is willing pay for one additional Quality Adjusted Life Year (QALY). WTP could however also be regarded as the marginal cost per additional QALY [20]. If we assume that decision-makers have a preference for displacing high ICER care vs low ICER care and are able to implement this preference, we could state the following: A high BI would displace more care than a low BI would. So, when care is displaced from high to low ICER, the higher the BI, the lower the total ICERs that are displaced. As WTP should include displacement, a higher BI yields a lower WTP [20].

Currently, the WTP is primarily used to assess whether an intervention is cost-effective (i.e., WTP > ICER) and various jurisdictions employ a formal WTP threshold, which has explicit influence on decision-making [10,11]. The height of this WTP threshold is widely debated with recent evidence suggesting values of, for example, £12,946, €24,226, €74,000, and other more extreme values per QALY [20,22–25]. For example England and the Netherlands apply respective WTP ranges of £20,000 - £30,000 and €20,000 – €80,000 in practice [22,26]. WTP

ranges are for example based on a jurisdictions' gross domestic product or on disease severity but they do not include BI as a factor [22,27].

Literature suggests that, given the implicit relation between BI, available budgets, displacement and WTP, achieving a single threshold is impossible and should never be used in practice but that the threshold should instead be related to BI and available budgets [1,17,28,29].

Recent work on this topic that aims to link affordability with cost-effectiveness using a more dynamic, BI based WTP still assesses affordability and cost-effectiveness as separate components [1]. Pearson describes that stakeholders struggle with assessing these separate components and their relation to a dynamic WTP [1].

Net Monetary Benefit (NMB) is a metric that has the potential to unify cost-effectiveness and affordability to provide healthcare gains and losses in monetary terms. It is typically defined as [30]:

$$1. \quad NMB = WTP * \Delta E - \Delta C$$

Where NMB is calculated per person with ΔE being the difference in effectiveness (as QALYs) and ΔC being the difference in costs [30,31]. The ICER is calculated as ΔC / ΔE. It is possible to include the ICER in the per patient NMB calculation through dividing equation 1 by ΔE, resulting in:

$$2. \quad NMB = (WTP - ICER) * \Delta E$$

In equation 2, WTP - ICER indicates the incremental monetary benefits or losses procured per gained QALY and ΔE indicates the number of QALYs gained per individual. The NMB on a population level could be derived by multiplying this per patient NMB by the number of patients treated with the new intervention. The number of patients treated can be derived from the BIAs as this is one of the underlying parameters on which BI is based [8,27,32,33]. Population NMB (pNMB) is thus calculated as:

$$3. \quad pNMB = \big((WTP - ICER) * \Delta E\big) * \frac{BI}{Treatment\ cost\ per\ patient}$$

Crucially, this formula retains the nonlinear characteristics of the initial NMB implementation as defined by equation 1.

The first part of equation 3 (WTP - ICER) specifies whether the intervention is cost-effective per unit of the intervention whilst the second part (BI / treatment cost per patient) can be interpreted as the number of units of the intervention that will be acquired.

With a fixed WTP, BI has no influence on the sign of pNMB. As a decision rule, a positive pNMB (or Net Present Value, its general economic counterpart), would warrant a decision to invest whilst a negative pNMB would reject an investment [34]. Thus, based on equation 3, BI would never have a role in deciding whether to invest or not. When WTP would however be more dynamic, as literature has suggested it should be, pNMB could be a tool to assess reimbursement decisions that incorporates both cost-effectiveness and BI.

We aim to provide a conceptual framework for uniting BI, WTP and ICERs and their associated uncertainty where WTP is dynamic and influenced by BI. We use nivolumab as a case study and employ three methods to describe and to quantify the influence of BI and WTP and provide insight into the impact of such assumptions on decision-making. We furthermore aim to show that pNMB, a metric based on the proven NMB, combined with a dynamic WTP could be the missing link in appraising and deciding on affordability vs cost-effectiveness with full incorporation of their individual uncertainty.

METHODS

Case study selection

We selected nivolumab as a case study example from the Netherlands as its introduction was met with debate regarding its relatively high price and its high base-case BI combined with a large off-label potential posing a substantial risk for even greater BI [35,36]. Additionally, the reimbursement dossier included a scenario (the scenario presented by the manufacturer) with an ICER (€62,277) below the Dutch non-binding WTP of €80,000 for diseases with highest disease severity [37]. With our aim of presenting a new conceptual framework for decision-making, using a cost-effective scenario is more informative than a non-cost-effective scenario as the latter might be a reason to reject reimbursement irrespective of BI. Furthermore, the Probabilistic Sensitivity Analysis (PSA) yielded results that resemble a normal distribution, improving interpretability.

The average treatment cost per patient (stratified per gender) was stated in the reimbursement dossier. We converted treatment cost in cost per patient (irrespective of gender) of €41,201, thereby having subtracted the €3222 of docetaxel substitution that was noted in the reimbursement dossier [36].

Willingness to Pay

As stated in equation 3, BI has no role in determining the sign of pNMB and should therefore have no role in the investment decision although recent literature shows that BI has definitely had influence on reimbursement decisions and that it should have influence on WTP [14,16,17,19,29].

We use three methods to illustrate and quantify a potential relationship between WTP and BI (henceforth referred to as *WTP = f(BI)*), being a historical real-world Dutch reimbursement decision, a completely arbitrary method and a recent paper describing displacement in Dutch healthcare (POINT tool). These methods are solely used for *WTP = f(BI)* and have no further role in our nivolumab case study.

Method 1: Reimbursement decision in Hepatitis C

We aim to derive a possible relationship between BI and WTP using the introduction of sofosbuvir, a hepatitis C Direct-Acting Antiviral drug, as example. We acknowledge that this is a simplification of the real world and is associated with some assumptions, however the apparent simplicity makes for a rather clear and illustrative example of a WTP that is influenced by BI. The underlying assumptions are presented in box 1.

The derived coefficient from the assumptions in box 1, calculated as (€80,000 – €47,481)) / ((€2.5 million – €38.79 million), equals -0.000896. This means that for each €1 and €1116 of BI above €2.5 million, WTP is lowered by €0.000896 and €1, respectively.

Method 2: Arbitrary relationship

As method 1 is based on hepatitis C and therefore potentially not representative for our nivolumab case study, we also include an arbitrary coefficient of -0.0004. This value was chosen so that an annual BI of 100 million, considered very high in the Netherlands, would lead to a €40,000 lower WTP.

Method 3: POINT tool

The relationship between BI and WTP has been explored and modelled by Adang et al [20]. The goal of their study, performed from a Dutch hospital perspective, was to investigate the influence of introducing innovations on budget allocation and the resulting change in wellbeing of the Dutch population. As part of this research, a tool (POINT 1.0) was developed which needs BI as input and yields a marginal WTP. The POINT tool was available as a Microsoft Excel workbook. As the function describing WTP was not linear and could not easily be replicated, we developed a macro that inputs all simulated BI values and records the resulting WTP.

In 2014, the Dutch reimbursement dossier for sofosbuvir was published [38]. Sofosbuvir was deemed cost-effective but carried a risk of very high BI. Therefore, the Dutch Minister of Health, Welfare and Sport (MoH) decided that reimbursement was limited to the most severe patients [39]. This case is described by Geenen et al. in detail [14].

The hepatitis C population consists of a variety of subpopulations with different ICERs [38]. The highest reported ICER for any of the HIV-negative subpopulations was €47,481. Based on BI, reimbursement was limited to patients with most severe disease, denoted as F4 – F3 on the METAVIR scale [38,39]. We therefore conclude that BI for F4 – F3 was still acceptable but unacceptable for F4 – F2. The ICER was not differentiated per METAVIR score.

The reimbursement dossier assumed that 49.1% of the patients belonged to F4 – F2 so we assume that F4 – F2 accounted for 49.1% of total BI. The highest BI estimate in the reimbursement dossier was €79 million per year, resulting in €38.79 million for F4 – F2. We assume that, as the MoH did not reimburse an intervention with a BI of €38.79 and an ICER of €47,481, BI had resulted in a WTP below the ICER.

At the time, the estimated BI below €2.5 million did not warrant conducting a cost-effectiveness analysis so we assume that a BI < €2.5 million had no influence on WTP [27]. We thus derive that a change in BI from €2.5 million to €38.79 million changed the WTP from €80,000 to €47,481.

For simplicity, we assume that the relationship between BI and WTP is linear. We also assume that the WTP will never be lower than €20,000, regardless of BI. This value is arbitrary, but it is equal to the current Dutch WTP for interventions with a low burden of disease [27].
To conclude, we assume:

- BI = €38.79 million: WTP = ICER
- BI > €38.79 million: €20,000 < WTP < ICER
- BI < €38.79 million: €80,000 > WTP > ICER
- BI < €2.5 million: WTP = €80,000

Box 1. Assumptions regarding Hepatitis C reimbursement decision and its use in describing a potential relationship between WTP and BI.

ICER & effectiveness

As ICER data source, we use the nivolumab PSA results from the manufacturer's base-case as published in the reimbursement dossier [36]. The Dutch Healthcare Institute (ZIN) provided us with the raw data. All ICER samples are located in the upper-right quadrant of the cost-effectiveness plane. We sorted these ICERs, distributed them amongst 100 groups of the same size and then calculated the mean ICER per group. Each ICER group (hereafter scenarios) thus represents a scenario with a given ICER and a 1% probability of being the true ICER. The difference in effectiveness (ΔE) within an ICER scenario was aggregated to yield a mean ΔE per scenario.

BI

Quantification of BI uncertainty remains limited to scenario analyses where BI is typically presented as a point estimate accompanied by a range [33]. BI estimates are however inaccurate [14,40–42]. Keeping et al. recently described that BI estimates used by payers deviated from actual BI with more than 40% in 80% of the cases [40].

The nivolumab reimbursement dossier presents an estimated BI of €46 – 74 million per year [36]. To simulate the BI estimation uncertainty, we assume that the point estimate is the mean of this range at €60 million per year. The standard deviation (sd) could then be interpreted as €14 million. To better reflect the reported BI uncertainty reported by, for example, Keeping et al. we multiply the sd by 1.5 to yield 21 million. Assuming BI estimation error is normally distributed, these parameters still only yield a 25% chance on a BI deviation > 40%. We generate 50,000 samples and, like the ICER, sort these samples, distribute them among 100 groups of equal size and then calculate the mean BI estimate.

RESULTS

Distribution of WTP for different *WTP = f(BI)* methods

The distribution of WTP per BI with different *WTP = f(BI)* methods is shown in figure 1. The data of the main figure are the grouped BI estimates (100) and the corresponding WTPs (100 per method). The density plot for the BI estimates, shown above the main plot, is identical for all *WTP = f(BI)* methods. At the right of the main plot, density WTP plots for each method are presented.

The POINT method and Fixed WTP yield very narrow density plots, indicating that these WTP are insensitive to BI [20,43]. The base WTP of these two methods are however different. The -0.000896 coefficient generates a normally distributed spread of WTP that nearly spans the entire WTP range. The mean WTP is ± €58,000. There is also some condensing of the tails of the distribution in the first and last BI groups. The -0.000896 coefficient, based on method 1, leads to a WTP that is more sensitive to BI. This results in lower WTP and significant accumulation of WTP at the minimum of €20,000.

Presentation of pNMB results

The pNMB results are displayed graphically in figures 2 – 4. To improve interpretability, we provide a numerical example in the form of table 1. In the example of table 1, the results of a fixed WTP are shown. The only difference between the results of table 1 and figure 2 is that 5 BI and ICER scenarios were used for the table compared to 100 for the figure.

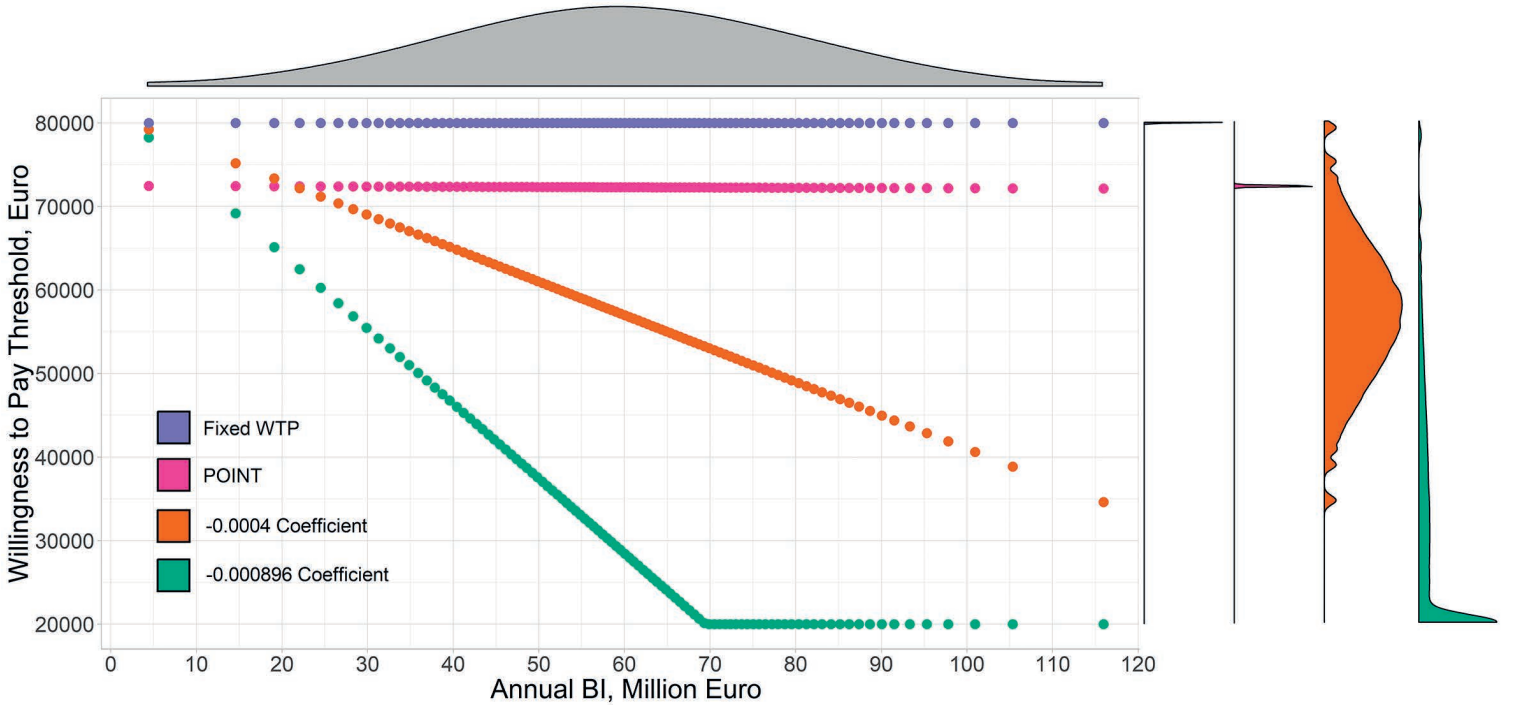


Figure 1. Distribution of WTP using different *WTP = f(BI)* methods. Including density plots for both axes. 100 BI scenarios per assumption are used as data. The BI density plot is the same for the 4 methods

BI and ICER scenario 1 are the lowest BI and ICER estimates, scenario 5 corresponds to the highest estimates. The area defined by the dotted line presents the pNMB (in million €) per BI and ICER scenario and is presented as a heatmap (panel B) in figures 2 – 4. This shows that higher ICER estimates lower the pNMB, a higher BI estimate increases the minimum as well as maximum pNMB.

The row "% Positive ICER Scenarios (A)" describes the percentage of ICER scenarios with a positive pNMB, per single BI scenario. In the figures, this is displayed in panel A. The column "% Positive BI scenarios (D)" describes the percentage of BI scenarios with a positive pNMB, per single ICER scenario and is presented in panel D. As the sign of the pNMB determines the investment decision, the % positive per ICER or BI scenario provides information on the influence of ICER or BI on the investment decision. For example, if the percentage of positive ICER scenarios does not change per BI scenario (as is the case in table 1), reducing uncertainty in BI does not help to determine the right decision as the probability of a wrong decision remains 40%.

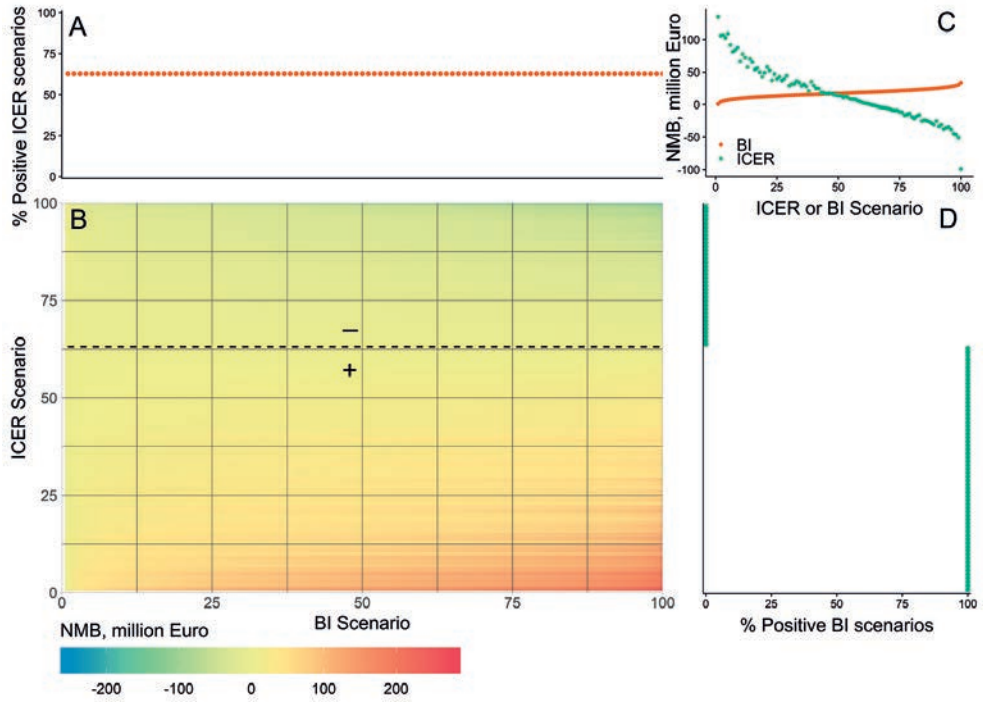


Figure 2. pNMB results generated with a Fixed WTP. 2a: The percentage of positive ICER scenario given a single BI scenario. 2b: heatmap of pNMB per BI and ICER scenario. The dashed line defines the boundary between positive and negative pNMB. The dashed contours define the boundary of specific pNMB values. 2c: mean pNMB per BI and ICER scenario. 2d: Ratio of positive BI scenarios per ICER quantile.

Table 1. Tabular example of the data presented in figures 2 – 4. Shows the results of a fixed WTP (€80,000) and 5 BI and ICER scenarios. k = €1000, m = €1,000,000.

	Average pNMB of each BI scenario (C)	**-3.3m**	**-5.2m**	**-6.4m**	**-7.6m**	**-9.5m**	**% Positive BI scenarios (D)**	**Average of each ICER scenario (C)**
	% Positive ICER scenarios (A)	**60%**	**60%**	**60%**	**60%**	**60%**		
ICER Scenario (B)	**5** (ICER = 269k)	*-97.5m*	*-156.2m*	*-191.5m*	*-227.2m*	*-285.3m*	**0%**	**-191.5m**
	4 (ICER = 89k)	*-4.9m*	*-7.9m*	*-9.7m*	*-11.5m*	*-14.4m*	**0%**	**-9.7m**
	3 (ICER = 70k)	*7.2m*	*11.5m*	*14.1m*	*16.7m*	*21.0m*	**100%**	**14.1m**
	2 (ICER = 57k)	*23.5m*	*37.6m*	*46.1m*	*54.7m*	*68.6m*	**100%**	**46.1m**
	1 (ICER = 38k)	*55.5m*	*88.9m*	*109.0m*	*129.4m*	*162.4m*	**100%**	**109.0m**
		1 (BI = 31m)	**2** (BI = 59m)	**3** (BI = 60m)	**4** (BI = 71m)	**5** (BI = 89m)		
				BI Scenario (B)				

The top row and righthand column denote the average pNMB (in million €) per BI and ICER scenario, respectively. These average results are combined in panel C. This provides information on the influence of uncertainty in BI and ICER on average pNMB. Relating this to decision-making, it represents the spread in absolute pNMB given the uncertainty in the ICER and BI. If, for example, the average pNMB per ICER scenario ranges from deeply negative to deeply positive, it informs the decision-maker that reducing ICER uncertainty could prevent significant losses.

Panels A and D thus describe which parameters determine the probability whether the decision results in potential profits. Figure C then informs on the potential losses or profits of a decision. The heatmap (panel C) is a visualisation of the bivariate pNMB distribution.

pNMB results of fixed WTP (€80,000)

In figure 2, the results for a selected fixed WTP of €80,000 are presented. A lower quantile reflects a lower ICER and BI estimate. Figure 2a indicates that for each BI scenario, the percentage of positive BI scenarios stays the same. BI does however influence the magnitude of the pNMB as the orange line in figure 2b, indicating mean pNMB per scenario, is not constant and as there is a colour gradient in the direction of the x-axis in figure 2c, indicating difference in pNMB.

Figure 2d shows that the ICER does influence the sign of pNMB and that the 63 scenarios with the lowest ICER yield a positive pNMB and the resulting 37 higher ICER scenarios result in a negative pNMB. Figure 2b shows that, logically, ICER has a great influence on the mean pNMB.

pNMB Results of Method 1 (real-world reimbursement decision)

Figure 3 presents the results using the *WTP = f(BI)* based on method 1. Figure 3a shows that BI now influences the sign of the pNMB and could thus influence the investment decision. Furthermore, 3c shows that the influence of BI on mean pNMB increased compared to 2c. The overall pNMB of method 1, as shown in 3b and 3c, is lower than in figure 2 due to lower mean WTP of this assumption.

Interestingly, mean pNMB as depicted in 3c is negative for all ICER and BI scenarios whilst the heatmap (3b) does show a positive region. There are thus BI and ICER combinations that yield positive pNMB but no single BI or ICER delivers an pNMB that is, on average, positive. For a decision-maker, this would imply that lowering uncertainty in only BI or ICER is probably insufficient but that a joint reduction in uncertainty would be needed. Furthermore, a decision-maker could use the information that a BI beyond scenario 51 never yields a positive pNMB, as a basis for a price / volume arrangement or volume cap so that BI is guaranteed to stay within scenarios 0 – 51. Similarly, the ICER should in any case be within scenario 0 – 62, potentially inspiring the use of pay-for performance schemes.

The ICER, shown in 3d, still has influence on the sign of pNMB although now in a less binary and more gradual manner than in the fixed WTP setting. In 3b, the dashed contour defining an pNMB of 0 shows the shared or bivariate influence on pNMB.

pNMB Results of Method 2 (arbitrary coefficient)

The implications of the arbitrary coefficient of -0.0004 are shown in figure 4. As in figure 3, BI clearly has influence on pNMB and the theoretical investment decision by means of the sign of the pNMB per ICER scenario. Compared to figure 3, the lower coefficient of -0.0004 results in higher NMB caused by less reduction in WTP. All BI scenarios in figure 4 have a potential for positive pNMB whereas in figure 3 only a low BI combined with a low ICER would yield a positive pNMB.

From a decision-making perspective, ICER scenarios 100 – 62 never yield a positive pNMB, regardless of BI. So, if efforts were to be undertaken to manage the ICER risk (i.e., prevent

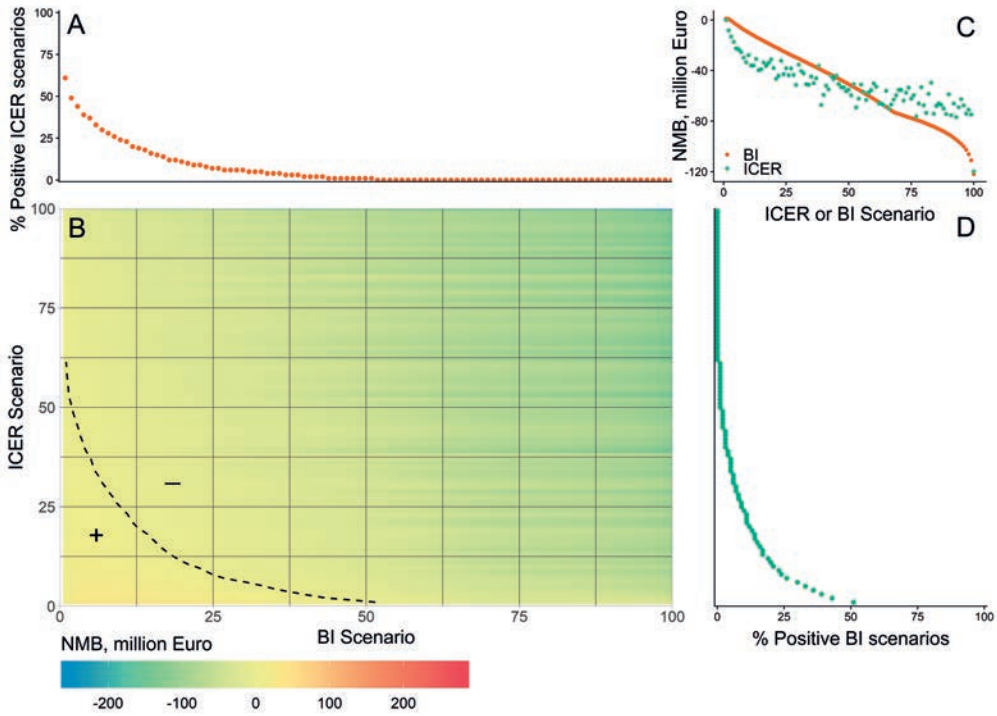


Figure 3. pNMB results of method 2 (*WTP = f(BI)* coefficient of -0.000896). 3a: The percentage of positive ICER scenario given a single BI scenario. 3b: heatmap of pNMB per BI and ICER scenario. The dashed line defines the boundary between positive and negative pNMB. The dashed contours define the boundary of specific pNMB values. 2c: mean pNMB per BI and ICER scenario. 2d: Ratio of positive BI scenarios per ICER quantile.

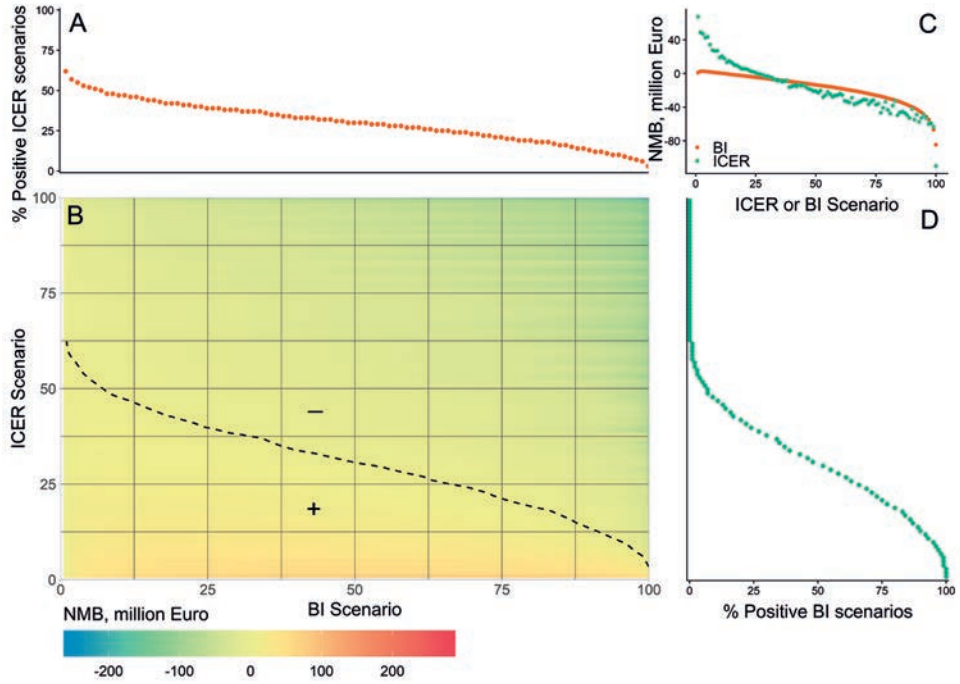


Figure 4. pNMB results of method 2 (*WTP = f(BI)* coefficient of -0.0004). 4a: The percentage of positive ICER scenario given a single BI scenario. 4b: heatmap of pNMB per BI and ICER scenario. The dashed line defines the boundary between positive and negative pNMB. The dashed contours define the boundary of specific pNMB values. 4c: mean pNMB per BI and ICER scenario. 4d: Ratio of positive BI scenarios per ICER quantile.

scenarios 62 – 100 from happening) like pay-for performance schemes, management of BI would not be necessary.

pNMB results of Method 3 (POINT)

As can be seen in figure 1, the POINT method is rather insensitive to BI in the simulated range and yields a WTP of ± €72,500 for all our BI values. The results are therefore nearly identical to those of a fixed WTP of €72,500. The results are shown in Appendix 1 and show that the influence of ICER and BI is (nearly) identical to those shown in figure 2.

DISCUSSION

We demonstrated that pNMB, a monetary value of health gained or lost to a system, has the potential to integrate ICER, WTP, and BI. Furthermore, we showed that without a role for displacement effects, defined by a fixed WTP, BI has no role in influencing an pNMB-based reimbursement decision.

As our results show, a WTP that is dependent on BI does enable BI to influence an pNMB guided investment decision [1,17,28,29]. Using nivolumab as case study and using three different methods for a dynamic WTP, we have shown that this approach could indeed lead to an pNMB that is influenced by the ICER as well as BI. This allows for a decision-making framework where affordability and cost-effectiveness are integrated into a single metric and enables a joint appraisal of these two entities.

Claxton et al. have evaluated the opportunity costs, reflected as WTP, of marginal (i.e., 'small') expenses in a healthcare system and determined this to be £12,936 per QALY in the UK [24]. Lomas et al. then used a similar approach to determine the opportunity cost (thresholds) for non-marginal (i.e., 'large') BI [44]. The definition of nonmarginal BI is discussed by Paulden et al [45]. Lomas and colleagues specifically addressed the influence of different BI values on the threshold and used a hepatitis C case-study to illustrate the implications of their research [44]. They derive an approximated linear relationship between expenditure (BI) and marginal productivity (WTP), resulting in a WTP of £12,542 and £12,166 for (UK) BI of £250 million and £2500 million, respectively. As UK thresholds as well as UK BI differs greatly from the herein adopted Dutch perspective, we decided not to explicitly incorporate the *WTP = f(BI)* influence as described by Lomas et al.

Their approach is however included more implicitly by means of the POINT *WTP = f(BI)* method. This method, described by Adang et al. used a technique similar to the work of Claxton et al. (and in that respect, also comparable to Lomas et al.) as they aimed to assess the Dutch opportunity costs of marginal expenditure using claims data on expenditure and mortality and quality of life data to assess QALYs, as well various demographic characteristics [20]. The results of Adang and colleagues, made accessible by means of an Excel workbook, however, show a very low sensitivity of WTP for BI values below approximately €2 billion. The relationship being non-linear, arbitrary annual BI values of €1 million, €100 million, €1 billion, €2 billion and €5 billion yield respective WTP values of €72,473, €72,189, €69,536, €66,341 and €20,066.

In 2017, total Dutch expenditure on 'specialist pharmaceutical care', reflecting mainly expensive specialty drugs for inpatient use, was €2 billion [46]. The previously mentioned Dutch hepatitis C case study noted a maximum annual BI of €79 million and, based on this estimate, drastic patient access restrictions were advised and implemented [14,38,39]. These two observations highlight a potential mismatch between solid empirical work by, for example, Adang et al., Claxton et al. and Lomas et al. and actual decision-making practice where BI appears to play a much more prominent role [20,24,44].

Some argue whether BI should actually play a role in reimbursement decision or if it is merely a budgetary practicality [8]. This theoretical discussion is however superseded by the fact that current decision-making practice is definitely being influenced by BI and that there is

no indication that this is likely to change [7,14–19]. To the contrary, a potential treatment for Alzheimer's disease (or any other severe and highly prevalent disease) would present healthcare systems with even greater financial pressure and consequent BI guided access decisions. In light of this reality, we believe that BI should then at least be properly integrated in decision-making instead of the current plethora of rather inconsistent, one-off decisions.

This proper or better integration of BI could be extended to the domain of uncertainty. Methodology on quantifying and managing ICER uncertainty is widely adopted and ICER sensitivity analysis is mandatory in many reimbursement files [27,47]. For BI, the exploration of uncertainty is limited to scenario analysis and lacks the advanced characterisation of its ICER counterpart [33]. Although we do not provide tools or methods for improved management of BI uncertainty, the combined influence of BI and ICER on pNMB and its graphical representation that we presented could provide means to give more insight in the combined uncertainty of ICER and BI on an intervention's potential value.

Using our approach, decision-makers could identify specific ranges or thresholds for the ICER and / or BI that are required to yield a positive pNMB. If, for example, a specific BI threshold may not be exceeded, it could provide the decision maker with an incentive to opt for price-volume arrangements. With our pNMB approach, uncertainty in the ICER can be combined with these BI thresholds: a certain ICER threshold could warrant a certain BI threshold and vice-versa. In practice, this could lead to access-schemes where aspects of pay-for-performance or coverage with evidence development (ICER-related risk and uncertainty) and price/volume arrangements (BI-related risk / uncertainty) could be combined. This will be work for follow-up research. Next to that, we believe that the true novelty of this paper, being the combination of BI, WTP and ICER and their uncertainty into one metric (pNMB), could serve as a tool to aid decision-makers with (combined) appraisal of cost-effectiveness and affordability.

Our study has a number of limitations. First, a WTP that is directly influenced by BI might be unrealistic. It is described in Claxton et al. that decision-making should be driven by a WTP that is based by the marginal opportunity cost of new investment within the system [30]. This new investment, as Adang has for example described, has influence on the marginal opportunity cost [20]. Besides the Adang, Claxton and Lomas studies, more evidence in favour of a potential of BI influencing WTP is present. This evidence is not only theoretical but also empirical [8,9,14,19,21,44].

Second, the methods used to quantify the relationship between BI and WTP are derived from practice (method 1 + 3) or even arbitrary (method 2) and the validity of these quantifications can be questioned. We do however believe that the methods used are based on realistic examples. Besides this, the coefficients are specific for the Netherlands, but the methods used to calculate these coefficients are not likely to be completely different between countries. Our method could thus be used in different settings or jurisdictions, as long as there are decision-making examples

on which to base a *WTP = f(BI)* quantification. We furthermore believe that our assumptions and methods, crude as they may be, perfectly illustrate the idea and potential merits of a WTP that is driven by BI.

Third, by dividing treatment cost per patient by annual BI to calculate pNMB, we assume that a full treatment (including its potentially life-time horizon ICER) happens in exactly one year. A treatment can however last for multiple years and would thus incur BI in multiple years. For example, for a 2-year treatment with a total cost of €100,000 and annual BI of €1,000,000 in the first year, we would assume that 10 patients would be treated and would have started to incur their ICER which would be translated to pNMB. In reality however, 20 patients would have started treatment. In the second year, if we for simplicity assume that no more patients started treatment and annual BI of €1,000,000 is again incurred, we would again find that 10 patients have completed treatment which brings the total number of treated patients to the true value of 20. We thus conclude that if treatments span multiple years, we underestimate pNMB in the first year(s). In the reimbursement dossier of our nivolumab case study, nivolumab was assumed to be used until disease progression, resulting in a median treatment duration of 3.5 months [36]. We do not believe that the annualisation of the ICER (composing a life-time horizon of costs and effects) into an annual pNMB is an issue as these ICER components are typically discounted to reflect their present-day value [48,49].

CONCLUSION

A WTP threshold that reflects displacement and opportunity costs is widely cited to be needed in reimbursement decision-making. Besides, cost-effectiveness and affordability (and their associated uncertainty) are currently appraised as separate entities and decision-makers are known to struggle with this separate appraisal. We have shown that a decision-making framework using societal NMB can combine affordability and cost-effectiveness into a single metric and thus a single decision. A prerequisite of this framework is the existence of an explicit relationship between BI and WTP. Using various methods, we have provided examples and implications of such relationships on pNMB. BI has an explicit influence on WTP and when integrated into pNMB, it could lead to better assessment of the impact of uncertainty of BI and ICER on an innovation's value. This pNMB approach enables decision-makers to identify (combinations of) threshold values for the ICER and BI that are required for an intervention to add value to the health care system, thereby informing on suitable managed entry agreements or pricing arrangements. To conclude, we believe that the decision-making concept presented here could lead to a truly combined and united consideration of cost-effectiveness and affordability.

REFERENCES

1. Pearson SD. The ICER Value Framework: Integrating Cost Effectiveness and Affordability in the Assessment of Health Care Value. Value Health. 2018;21:258–65.
2. Birch S, Gafni A. The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold. J Health Serv Res Policy. 2006;11:46–51.
3. Klemp M, Frønsdal KB, Facey K. What principles should govern the use of managed entry agreements? Int J Technol Assess Health Care. 2011;27:77–83.
4. Makady A, van Veelen A, de Boer A, Hillege H, Klungel OH, Goettsch W. Implementing managed entry agreements in practice: The Dutch reality check. Health Policy Amst Neth. 2019;123:267–74.
5. Pauwels K, Huys I, Vogler S, Casteels M, Simoens S. Managed Entry Agreements for Oncology Drugs: Lessons from the European Experience to Inform the Future. Front Pharmacol [Internet]. 2017 [cited 2019 Apr 4];8. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2017.00171/full
6. Chhatwal J, Kanwal F, Roberts MS, Dunn MA. Cost-effectiveness and budget impact of hepatitis C virus treatment with sofosbuvir and ledipasvir in the United States. Ann Intern Med. 2015;162:397–406.
7. Ghabri S, Mauskopf J. The use of budget impact analysis in the economic evaluation of new medicines in Australia, England, France and the United States: relationship to cost-effectiveness analysis and methodological challenges. Eur J Health Econ. 2018;19:173–5.
8. Cohen JP, Stolk E, Niezen M. Role of Budget Impact in Drug Reimbursement Decisions. J Health Polit Policy Law. 2008;33:225–47.
9. Niezen MGH, Bont A de, Busschbach JJV, Cohen JP, Stolk EA. Finding legitimacy for the role of budget impact in drug reimbursement decisions. Int J Technol Assess Health Care. 2009;25:49–55.
10. Simoens S. Use of Economic Evaluation in Decision Making: Evidence and Recommendations for Improvement. Drugs. 2010;70:1917–26.
11. Ghijben P, Gu Y, Lancsar E, Zavarsek S. Revealed and Stated Preferences of Decision Makers for Priority Setting in Health Technology Assessment: A Systematic Review. PharmacoEconomics. 2018;36:323–40.
12. Franken M, Nilsson F, Sandmann F, Boer A de, Koopmanschap M. Unravelling Drug Reimbursement Outcomes: A Comparative Study of the Role of Pharmacoeconomic Evidence in Dutch and Swedish Reimbursement Decision Making. PharmacoEconomics. 2013;31:781–97.
13. Vreman RA, Bouvy JC, Bloem LT, Hövels AM, Mantel-Teeuwisse AK, Leufkens HGM, et al. Weighing of Evidence by Health Technology Assessment Bodies: Retrospective Study of Reimbursement Recommendations for Conditionally Approved Drugs. Clin Pharmacol Ther. 2019;105:684–91.
14. Geenen JW, Boersma C, Klungel OH, Hövels AM. Accuracy of budget impact estimations and impact on patient access: a hepatitis C case study. Eur J Health Econ [Internet]. 2019 [cited 2019 May 2]; Available from: https://doi.org/10.1007/s10198-019-01048-z
15. Flume M, Bardou M, Capri S, Sola-Morales O, Cunningham D, Levin L-A, et al. Approaches to manage 'affordability' of high budget impact medicines in key EU countries. J Mark Access Health Policy. 2018;6:1478539.
16. American Society of Clinical Oncology Position Statement on Addressing the Affordability of Cancer Drugs. J Oncol Pract. 2017;JOP.2017.027359.
17. Neumann PJ, Cohen JT, Weinstein MC. Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold. N Engl J Med. 2014;371:796–7.
18. Mauskopf J, Chirila C, Masaquel C, Boye KS, Bowman L, Birt J, et al. RELATIONSHIP

BETWEEN FINANCIAL IMPACT AND COVERAGE OF DRUGS IN AUSTRALIA. Int J Technol Assess Health Care. 2013;29:92–100.

19. Barua S, Greenwald R, Grebely J, Dore GJ, Swan T, Taylor LE. Restrictions for Medicaid Reimbursement of Sofosbuvir for the Treatment of Hepatitis C Virus Infection in the United States. Ann Intern Med. 2015;163:215.
20. Adang E, Stadhouders N, Parsons C, Wammes J, Govaert P, Determann D, et al. Verdringingseffecten binnen het Nederlandse zorgstelsel [Internet]. 2018. Available from: https://www.iqhealthcare.nl/media/124601/eindrapport-verdringing-2018.pdf
21. Niezen MGH, Stolk EA, Steenhoek A, Uyl-De Groot CA. Inequalities in oncology care: Economic consequences of high cost drugs. Eur J Cancer. 2006;42:2887–92.
22. Shiroiwa T, Sung Y-K, Fukuda T, Lang H-C, Bae S-C, Tsutani K. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ. 2010;19:422–37.
23. 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;24:1–7.
24. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess Winch Engl. 2015;19:1–503, v–vi.
25. Ryen L, Svensson M. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. Health Econ. 2015;24:1289–301.
26. Carrera P, IJzerman MJ. Are current ICER thresholds outdated? Valuing medicines in the era of personalized healthcare. Expert Rev Pharmacoecon Outcomes Res. 2016;16:435–7.
27. Zorginstituut Nederland. Kosteneffectiviteit in de praktijk - Rapport [Internet]. 2015 [cited 2019 Apr 24]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk
28. Bertram MY, Lauer JA, De Joncheere K, Edejer T, Hutubessy R, Kieny M-P, et al. Cost–effectiveness thresholds: pros and cons. Bull World Health Organ. 2016;94:925–30.
29. Cleemput I, Neyt M, Thiry N, De Laet C, Leys M. Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. Int J Technol Assess Health Care. 2011;27:71–6.
30. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015.
31. Messori A, Trippoli S. The results of a pharmacoeconomic study: incremental cost-effectiveness ratio versus net monetary benefit. Heart. 2017;103:1746–1746.
32. Zorginstituut Nederland. Format Budget Impact Analyse GVS (voorheen kostenconsequentieraming) - Publicatie [Internet]. 2016 [cited 2018 Mar 23]. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/04/06/format-budget-impact-analyse-gvs-voorheen-kostenconsequentieraming
33. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2014;17:5–14.
34. Ross SA. Uses, Abuses, and Alternatives to the Net-Present-Value Rule. Financ Manag. 1995;24:96–102.
35. Zorginstituut Nederland. Advies Zorginstituut: longkankermiddel nivolumab tegen huidige prijs niet in basispakket - Nieuwsbericht [Internet]. 2015 [cited 2018 May 1]. Available from: https://www.zorginstituutnederland.nl/actueel/nieuws/2015/12/08/advies-zorginstituut-longkankermiddel-nivolumab-tegen-huidige-prijs-niet-in-basispakket

36. Zorginstituut Nederland. Pakketadvies nivolumab (Opdivo®) bij gevorderde (plaveiselcel) niet-kleincellige longkanker - Advies [Internet]. 2015 [cited 2019 Jun 19]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2015/12/08/pakketadvies-nivolumab-opdivo

37. Zorginstituut Nederland. Pakketadvies in de praktijk: wikken en wegen voor een rechtvaardig pakket - Rapport [Internet]. 2017 [cited 2019 May 29]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2017/09/06/rapport-pakketadvies-in-de-praktijk-wikken-en-wegen-voor-een-rechtvaardig-pakket

38. Zorginstituut Nederland. GVS-advies sofosbuvir (Sovaldi®) bij chronische hepatitis C - Advies [Internet]. 2014 [cited 2018 May 3]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2014/05/20/gvs-advies-sofosbuvir-sovaldi-bij-chronische-hepatitis-c

39. Staten-Generaal TK der. Geneesmiddelenbeleid; Brief regering; Vergoeding Hepatitis C middel Sovaldi met financieel arrangement [Internet]. 2014 [cited 2018 Oct 23]. Available from: https://zoek.officielebekendmakingen.nl/dossier/29477/kst-29477-296?resultIndex=252&sorttype=1&sortorder=4

40. Keeping S, Deslandes PN, Haines KE, Routledge PA. Estimated Versus Observed Expenditure Associated with Medicines Recommended by the All Wales Medicines Strategy Group. PharmacoEconomics - Open [Internet]. 2019 [cited 2019 Jan 22]; Available from: https://doi.org/10.1007/s41669-019-0116-5

41. Cha M, Rifai B, Sarraf P. Pharmaceutical forecasting: throwing darts? Nat Rev Drug Discov. 2013;12:737–8.

42. Broder MS, Zambrano JM, Lee J, Marken RS. Systematic bias in predictions of new drugs' budget impact: analysis of a sample of recent US drug launches. Curr Med Res Opin. 2017;1–9.

43. IQ healthcare - Nederlands. Eindrapport Verdringing [Internet]. [cited 2019 Jul 3]. Available from: https://www.iqhealthcare.nl/nl/kennisbank/rapporten/e/eindrapport-verdringing/

44. Lomas J, Claxton K, Martin S, Soares M. Resolving the "Cost-Effective but Unaffordable" Paradox: Estimating the Health Opportunity Costs of Nonmarginal Budget Impacts. Value Health. 2018;21:266–75.

45. Paulden M, O'Mahony J, McCabe C. Determinants of Change in the Cost-effectiveness Threshold. Med Decis Making. 2017;37:264–76.

46. Monitor geneesmiddelen in de medisch specialistische zorg januari 2019 - Nederlandse Zorgautoriteit [Internet]. [cited 2019 Jun 18]. Available from: https://puc.overheid.nl/nza/doc/PUC_264248_22/1/

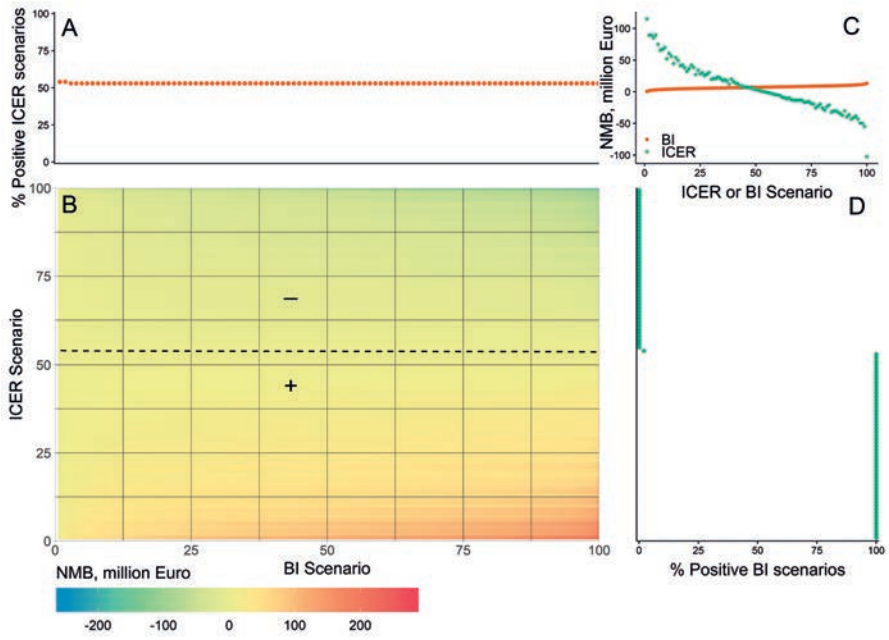
47. Foreword | Guide to the methods of technology appraisal 2013 | Guidance | NICE [Internet]. [cited 2019 May 13]. Available from: https://www.nice.org.uk/process/pmg9/chapter/foreword

48. Attema AE, Brouwer WBF, Claxton K. Discounting in Economic Evaluations. PharmacoEconomics. 2018;36:745–58.

49. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg - Publicatie [Internet]. 2016 [cited 2018 Oct 23]. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg

APPENDICES

Appendix 1



Appendix 1. Data generated with the POINT *WTP = f(BI)* assumption. 1a: The percentage of positive ICER scenario given a single BI scenario. 1b: heatmap of NMB per BI and ICER scenario. The dashed line defines the boundary between positive and negative NMB. The dashed contours define the boundary of specific NMB values. 1c: mean NMB per BI and ICER scenario. 1d: Ratio of positive BI scenarios per ICER quantile.

5.2

Real options analysis to inform reimbursement decisions

Joost W. Geenen, Cornelis Boersma, Joost J. Enzing, Olaf H. Klungel, Anke M. Hövels

Manuscript in preparation for submission.

ABSTRACT

Objectives

Informing reimbursement decisions based on economic evaluations is associated with two important challenges related to: i. the value of separate versus united assessment of cost-effectiveness and budget impact (BI); and ii. the importance of incorporation of timing of decisions based on the level of uncertainty. We aim to develop a real options analysis (ROA) based method that addresses these challenges in order to allow for more optimal reimbursement decision-making. To achieve this, we use an oncology case study from a Dutch perspective.

Methods

Net Monetary Benefit (NMB) is the main outcome and is calculated as: NMB = ((WTP – ICER) * incremental effectiveness) * (BI/treatment cost). Opdivo (nivolumab) was selected as case-study. Data on the ICER was derived from the reimbursement dossier whilst BI data was generated using a validated BI prediction model. For WTP, three methods for the influence of BI on WTP were used. For ROA implementation, we assumed that the true BI could be observed after one month.

Results

We compared traditional 'now or never' decisions to the option of waiting for more data. For some scenarios, waiting for 10 months of data was the optimal decision as risk due to uncertainty in the first 10 months outweighed immediate benefit (NMB). The different methods describing the relationship between WTP and BI had great influence on NMB (- €42 million for a fixed WTP vs €69 million for a WTP method based on a real-world Dutch reimbursement decision).

Conclusion

Based on a unified assessment of cost-effectiveness and BI by means of NMB and then incorporated timing using ROA and demonstrated that our ROA based method can be used to inform on the timing of reimbursement decisions. ROA could therefore be a suitable methodological tool for providing early guidance on flexible and adaptive reimbursement decisions, deemed essential in the current landscape of ever higher uncertainty at market access of new costly drugs.

INTRODUCTION

Current reimbursement decision-making is driven by the clinical or therapeutic value and at least to some extent by the Incremental Cost-Effectiveness Ratio (ICER) and Budget Impact (BI) [1–3]. The influence of the outcomes of the economic analysis on decision-making is at least partly driven by whether a formal Willingness to Pay Threshold (WTP) exists [1–3]. The relative importance of BI and cost-effectiveness in decision making differs and is not clearly defined [1,3]. Also, the exact role and relative importance of economic outcomes varies between various jurisdictions [1,3]. Inevitably, quantification of these outcomes involves uncertainty and this uncertainty is currently a crucial aspect in decision making [1]. The methods for quantification, visualising and accommodating for the uncertainty in ICER and BI are also different.

For the ICER, probabilistic sensitivity analysis (PSA) and the resulting Value of Information (VOI) analysis have been developed and are now widely used [4]. It has been shown that decision makers are less likely to reimburse drugs with a highly uncertain ICER [1].

The WTP, or ICER threshold, has different fixed values or fixed ranges in various jurisdictions [5,6]. In England, an upper limit of £20,000 - £30,000 per QALY is deemed cost-effective whilst an informal threshold varies between €20,000 - €50,000 - €80,000 in the Netherlands. The threshold value has however been a topic of scientific debate as for example Shiroiwa et al. and Claxton et al. have described [6,7]. WTP values of, for example, £12,936 and €74,000 been proposed as more accurate ICER thresholds [8,9]. A recent literature review reported an even wider range of WTP estimates and found a mean of €24,226 per QALY [10].

For BI, the availability- and type of guidance or legislation on the role of BI in reimbursement making varies per jurisdiction, although in general it could be stated that role of BI on access decisions is less clear than the role of the ICER [2,11–13]. Besides this, the scientific rigor of Budget Impact Analysis (BIA) is less developed than is the case for the ICER [14]. Recent examples have however clearly shown that in cases of a very high (expected) BI, rigorous access restrictions have been imposed [15–17].

As BIA are generally constructed using point estimates of various uncertain parameters and time-horizons, uncertainty in BI estimations is inevitable [18,19]. Mainly due to limited data, quantification of BI uncertainty remains limited to scenario analyses [18]. Therefore, BI is typically presented (for one or more scenarios) as a point estimate accompanied by a minimum and maximum value. A highly uncertain BI could lead to less or deferred access by means of, for example, managed entry agreements (MEA) [13,20,21].

Interestingly, ICER and BI (and their uncertainty) are typically appraised separately whilst they both inform on the same decision: does the investment deliver a health gain? This is

especially relevant as recent studies have shown that a relationship between WTP and BI exists, and that this relationship can be quantified [8,22–24]. This would mean that BI could have influence on the question whether an innovation is deemed cost-effective (i.e., ICER < WTP). A joint assessment of these three components could thus be crucial for solid reimbursement decision making.

Another limitation in current decision-making practice can be found in the lack of incorporating timing of decisions, evolution in dynamic health care practices and development of evidence and uncertainty over time. MEAs, Conditional Reimbursement (CR) or Coverage with Evidence Development (CED) are examples of tools designed to allow for granting rapid access whilst evidence on outcomes or BI, is still developing [25–30]. Besides, active management of innovations or label changes (e.g., changes in indication) do happen and influence cost-effectiveness and should thus be included in economic analyses [31,32]. Currently tools as MEA, CR and CED are quite widely used and they should be accompanied by reimbursement decision making methodology that is, at its core, suitable for implementing the factor of time [28].

Current cost-effectiveness analyses (CEA) and BIA however employ the Net Present Value (NPV) paradigm; future benefits and costs are discounted towards a present-day value and the investment or reimbursement decision is to be taken now or never. Flexibility can of course be added by means of scenario analyses or reperforming a CEA or BIA after some time, but in the essence of these analyses, delaying the decision is not an available option. Therefore, current CEA or BIA methodology is unable to fully incorporate the role of active management of healthcare related projects or the development of uncertainty over time.

Real Options Analysis (ROA) is a technique used in economics for valuing investment decisions [33]. The NPV, a classic approach for assessing investment opportunities, only allows for the decision to invest or to not invest [33]. ROA on the other hand specifically recognises the postponement of the investment (decision) as an option [33]. ROA is thus inherently designed for coping with timing of an investment and could, as for example Grutters et al., Attema et al. and Favato et al. have shown, be used to inform healthcare related decisions [31,34,35].

To summarise, current reimbursement decision making methodology has two major issues: It lacks a united assessment of CE and BI and it lacks incorporation of timing of decisions and development of evidence and uncertainty over time. In this study, we aim to develop and demonstrate a ROA-based method that amends these issues and therefore allows for better reimbursement decision-making. To achieve this, we use an Opdivo (nivolumab, oncology) case study from a Dutch perspective.

METHODS

Net Monetary Benefit

The main outcome measure used in our analyses is based on Net Monetary Benefit (NMB). NMB is typically defined as [36]:

$$1. \quad NMB = WTP * \Delta E - \Delta C$$

Where WTP is the amount the decision-maker is willing to pay for one unit of increased effectiveness, ΔE is the difference in effectiveness and ΔC the difference in costs [36]. When ΔE and ΔC are per patient, the resulting NMB is also per patient. It is possible to rewrite equation 1 so that the ICER is included, yielding equation 2:

$$2. \quad NMB = (WTP - ICER) * \Delta E$$

Again, if ΔE is per patient, the resulting NMB is per patient. In equation 2, WTP – ICER determines whether the innovation is cost-effective per (theoretical) unit, ΔE can then be interpreted as the number of units that are procured. The NMB to the entire healthcare system or society can be calculated by multiplying the pNMB by the number of patients receiving the new intervention. This population NMB (pNMB) is calculated as:

$$3. \quad pNMB = \big((WTP - ICER) * \Delta E\big) * \frac{BI}{Treatment\ \text{cost}\ per\ patient}$$

This equation retains the characteristics of the traditional NMB, like the linearity regarding change in ΔC and ΔE and insensitivity to different CE quadrants.

Equation 3 states that, if ΔE, BI and treatment cost per patient are positive numbers, only WTP and the ICER determine whether pNMB is positive. As a decision rule, a positive pNMB (or NPV, its general economic counterpart) warrants investment whilst a negative pNMB does not [37]. Thus, with equation 3, only WTP and ICER can determine the investment decision and, crucially, BI would never have a role in deciding whether to invest.

Data on input parameters

ICER and ΔE

The ICER and the associated uncertainty are quantified using the results of PSA as this usually is the best estimate of the influence of combined parameter uncertainty on the ICER. The Dutch

Health Care Institute (ZIN) provided us with the Opdivo PSA results. As the ICER is not necessarily normally distributed, we sort the ICERs (as they are in the same CE quadrant), distribute them amongst groups of the same size and then calculate the mean ICER per group. Each ICER group (hereafter scenarios) thus represents a scenario with a given ICER and a certain probability of being the true ICER. We used 100 groups for the ICERs so that each resulting scenario has a 1% probability of being the true ICER. The difference in effectiveness (ΔE) within an ICER scenario was aggregated to yield a mean ΔE per scenario.

BI

Current BI estimates provide very little insight into the actual uncertainty or probability distribution of BI, aspects that are crucial for ROA. We therefore used a validated BI prediction model that is described in detail elsewhere [38].

In short, this data-driven regression-based prediction model is trained and validated using monthly Dutch BI data, where BI is defined as volume * list price per drug. This population-level BI data source covers inpatient as well as outpatient prescriptions and is validated to be representative of the Netherlands [38]. The prediction model was validated using a of set oncology products, included various product characteristics (i.e., orphan status, cancer site, First in Class designation) and was limited to predicting the first 45 months of BI. Crucially, the model was validated using a rolling forecasting origin. This approach allows for the monthly addition of new BI data and thus retrains the model with each addition of monthly data and therefore mimics the envisioned real-world use of such a model.

We demonstrate this rolling forecasting origin and its implementation using the nivolumab case study. We denote the months of data that is available for predicting the BI of a specific product as t_data. In our dataset, the first BI record of nivolumab was 1 August 2015. The reimbursement dossier was published on 8 Dec 2015. So, at this time, 4 months of BI data (t_data) were available to the prediction model. As the observed BI data cut-off is 1 March 2018, a total of 31 months of predictions can be compared to the observed data. As the prediction model is validated for a maximum of 45 months, the last 14 predicted months cannot be compared to the observed data and we are unable to extend t_data beyond 31 for Opdivo (as this data is unavailable). The validation indicated a reduction in prediction error with increasing availability of data (t_data).

We aggregate all prediction errors per t_data, sort them and distribute them amongst groups of the same size and then calculate the mean prediction error per group. Each BI error group (hereafter scenarios) thus represents a scenario with a given prediction error and a certain probability of being the true error. As with the ICER, we used 100 groups, so each resulting scenario has a 1% probability of being the true scenario. For each predicted month, we multiply the predicted BI (for that month) with all error scenarios (for that t_data) and thus yield 100 possible BI estimates that are distributed based on the prediction error of a specific t_data.

Average treatment cost

The Opdivo reimbursement dossier specified average treatment cost per gender (€46,200 for males, €42,646 for females) and €3222 of docetaxel substitution [39]. For simplicity, we assume 50% of Opdivo users is female and thus average the treatment cost per gender. After subtracting the noted substitution cost, we yield an average treatment cost of €41,201 per patient.

Willingness to Pay

As base-case, we use the Dutch €80,000 threshold that is designated for indications with a high burden of disease [40]. As stated in equation 3, BI has no role in determining the sign of pNMB and should therefore have no role in the investment decision. Various studies, for example by Lomas et al., Claxton et al, Adang et al., and Geenen et al., have however described that a relationship exists between BI and WTP and have therefore quantified this relationship [8,22–24].

In this study, we consider three potential relationships between WTP and BI (henceforth referred to as $WTP = f(BI)$) that are described in detail elsewhere [24]. The first method, denoted as POINT, is based on empirical research by Adang et al. who quantified the marginal WTP with increases in marginal BI in Dutch inpatient care and could therefore be seen as a Dutch application of the earlier work by Claxton et al [8,23].

The second method, described by Geenen et al., used the Dutch reimbursement decision on sofosbuvir (Hepatitis C drug) to derive an influence of WTP on BI [24]. This hepatitis C case is informative as reimbursement was explicitly limited due to high BI estimations whilst the ICER was below the Dutch WTP threshold of €80,000 [17]. A linear function with a coefficient of –0.000896 was derived, meaning that for each €1 and €1116 of BI, WTP would reduce with €0.000896 and €1, respectively. For methods 2 and 3, only BI above €2.5 million per year is influenced by the coefficient as ZIN disregards BI below 2.5 million [24,40,41].

The third method is an arbitrary coefficient of -0.0004 which results in a €1 reduction in WTP for an annual BI of €2500 [24]. This coefficient results in a WTP that is lowered to €40,000 by an annual BI of €102.5 million (which would be considered very high in the Netherlands). The resulting WTP would still be higher than the lowest Dutch WTP threshold (€20,000) which is designated for innovations treating disease of low severity [40].

We acknowledge that methods 2 and 3 have less empirical foundation than method 1. Still, the fact that they are derived from real-world access decisions and their apparent simplicity makes for a clear and illustrative example of a WTP that is influenced by BI and its influence on ROA-guided decision-making.

Case study selection

We selected Opdivo as a case study based on various characteristics, mainly imposed by the use of the BI prediction model that is only validated for oncology products: The case study should therefore be an oncology product and should have a first BI record after 1 May 2012 [38]. Furthermore, the reimbursement dossier should be published within 6 months of the first BI record to ensure that the dossier was covering the same indication as the indication that generated the BI. As the data cut-off of observed BI data was 1 March 2018 and a minimum of 24 months of observed BI data is deemed to be required for performing ROA, the first BI record should be generated 1 March 2016 the latest.

The availability of a PSA (including scatterplot results) in a Dutch Reimbursement dossier was another inclusion criterion. The aforementioned criteria resulted in Keytruda (pembrolizumab) and Opdivo as potential case studies. The Keytruda base case ICER was however €113,000 and thus higher than the assumed or informal WTP of €80,000 [42]. For Opdivo, a scenario was available with an ICER below the WTP of €80,000. As our methodology bases the investment rule on pNMB (equation 2), a case study with an ICER > WTP is not very informative. Hence, Opdivo was selected.

Expansion of NMB

As we implement ICER and BI as multiple scenarios, aim to incorporate a relationship of BI and WTP and as we use a prediction model that provides predictions for different months and is trained on different amounts of data (t_data), we have to expand the initial pNMB formula (equation 3). Besides, figure 1 presents a schematic overview of all the aforementioned components and their role in assessing pNMB.

Expanding equation 3 with predicted months and ICER and BI scenarios yields equation 4:

$$4.\quad pNMB_{t_{data=td}} = \sum_{i=1}^{n}\left(\sum_{ii=1}^{nn} \begin{bmatrix} (WTP - ICER_{ii}) * \Delta E_{ii} \\ \vdots \\ (WTP - ICER_{nn}) * \Delta E_{nn} \end{bmatrix} * \begin{bmatrix} \frac{BI_{ii(month_i)}}{Treatment\ cost\ per\ patient} \\ \vdots \\ \frac{BI_{nn(month_n)}}{Treatment\ cost\ per\ patient} \end{bmatrix} * p_{ICER} * p_{BI} \right)$$

Where *td* is number of months of t_data, *n* and *i* denote the predicted month, *nn* and *ii* denote a single BI, ICER and ΔE scenario, pICER and p_{BI} imply the probability of each ICER and BI scenario, respectively. As ΔE is part of a specific ICER (i.e., they are not independent), each ICER scenario corresponds to a single ΔE scenario. When then expanding equation 4 with a relationship between BI and WTP, we get:

$$5.\quad pNMB_{t_{data=td}} = \sum_{i=1}^{n}\left(\sum_{ii=1}^{nn} * \begin{bmatrix} (WTP = f(BI_{ii(month_i)}) - ICER_{ii}) * \Delta E_{ii} \\ \vdots \\ (WTP = f(BI_{nn(month_n)}) - ICER_{nn}) * \Delta E_{nn} \end{bmatrix} * \begin{bmatrix} \frac{BI_{ii(month_i)}}{Treatment\ cost\ per\ patient} \\ \vdots \\ \frac{BI_{nn(month_n)}}{Treatment\ cost\ per\ patient} \end{bmatrix} * p_{ICER} * p_{BI} \right)$$

Crucially, BI now has influence on the sign of the pNMB through its influence on the WTP. An example of equation 5 using mock data is presented in equation 6:

$$6.\quad pNMB_{t_{data=1}} = \sum_{i=1}^{45}\left(\sum_{ii=1}^{100} * \begin{bmatrix} (80{,}000 - 30{,}000) * 0.5 \\ \vdots \\ (70{,}000 - 100{,}000) * 0.5 \end{bmatrix} * \begin{bmatrix} \frac{100{,}000}{40{,}000} \\ \vdots \\ \frac{1{,}000{,}000}{40{,}000} \end{bmatrix} * 0.01 * 0.01 \right)$$

From equation 6, one could derive that the monthly BI estimates range from €100,000 to €1,000,000 and the ICER from €30,000 to €100,000. In this example, a monthly BI of €100,000 results in a WTP of €80,000 whilst a BI of €1,000,000 yields a lower WTP of €70,000. Furthermore, as p_{ICER} and p_{BI} are 0.01, a total of 10,000 (100 * 100) BI:ICER combinations are generated per predicted month per t_data.

Implementation of Real Options Analysis

As mentioned in the introduction, ROA aims to quantify the value of waiting with investing compared to investing now or never. In box 1, we discuss key assumptions underlying ROA and whether those are met. In this study, we implement ROA as follows:

Equation 4 or 5 can be used to calculate pNMB. This pNMB consists of multiple BI and ICER scenarios of which some might be positive and some negative. It is however impossible to know which scenario is correct. As during implementation, new information on the ICER is not observed, the true ICER scenario will never be known. Information on BI however does become available and, after one month, the true BI scenario could be observed.

The pNMB where ROA is not used is the pNMB where waiting for more data is not an option. This pNMB, denoted as $pNMB_{standard}$, is thus the outcome of equation 4 or 5. The pNMB using ROA, called $pNMB_{option}$, does allow for observation of BI data. The starting point for $pNMB_{option}$ is always $pNMB_{standard}$. However, $pNMB_{option}$ assumes that $pNMB_{standard}$ is not executed but that

As Palmer and Smith described in 2000, ROA can be used for informing investment decisions if the following characteristics have to be present [43]:

1. Uncertainty regarding the future
2. Irreversibility (as sunk cost) of a decision
3. Timing of the investment matters

The relevance of assumption one is clear as uncertainty in the ICER, BI and post marketing changes is clearly important in current decision making [1,11,20,32]. Assumption two would imply that once a reimbursement decision is taken, the decision is not reversed. This is especially relevant as, in order for ROA to function properly, irreversibility should imply a certain amount of sunk cost. Therefore, this assumption is only met when the decision to reimburse cannot be revoked later.

In practice, halting of a reimbursed drug that still has clinical value (so is not obsolete) is complex and difficult due to political and societal sensitivities [25,26,44]. Makady and colleagues have, for example, shown that in the Netherlands the timely (ie, before having incurred large sunk costs) withdrawal of reimbursement is scarce [26]. We therefore believe that this assumption at least partly holds.

Assumption three holds as timing, in the sense that MEAs or CED (coverage with evidence development) are aimed at providing quick access where evidence or uncertainty develops over time. To conclude, the three assumptions postulated by Palmer and Smith still hold and ROA could therefore in theory be used to inform reimbursement decision making.

For further reading, Palmer and smith provide an excellent numerical example of option value [43].

Box 1. Assumptions underlying Real Options Analysis

instead, the true BI scenario of the next month is observed. Inevitably, the pNMB that is generated by $pNMB_{standard}$ in this month cannot be generated by $pNMB_{option}$ and is thus lost in $pNMB_{option}$. In return, $pNMB_{option}$ gets to observe the true BI scenario that constitutes the underlying pNMB.

If the observed BI scenario has a negative pNMB, there will be no investment. So, these negative BI scenarios would yield an pNMB of €0 (equal to not investing) instead of their negative counterpart in $pNMB_{standard}$. If the observed BI scenario has a positive pNMB, it will be invested in. These BI scenarios will thus yield the same pNMB as they would in $pNMB_{standard}$. The ability to observe negative BI scenarios and avert investment in these negative scenarios, at the expense of the pNMB proceeds of a single month, is thus what drives $pNMB_{option}$.

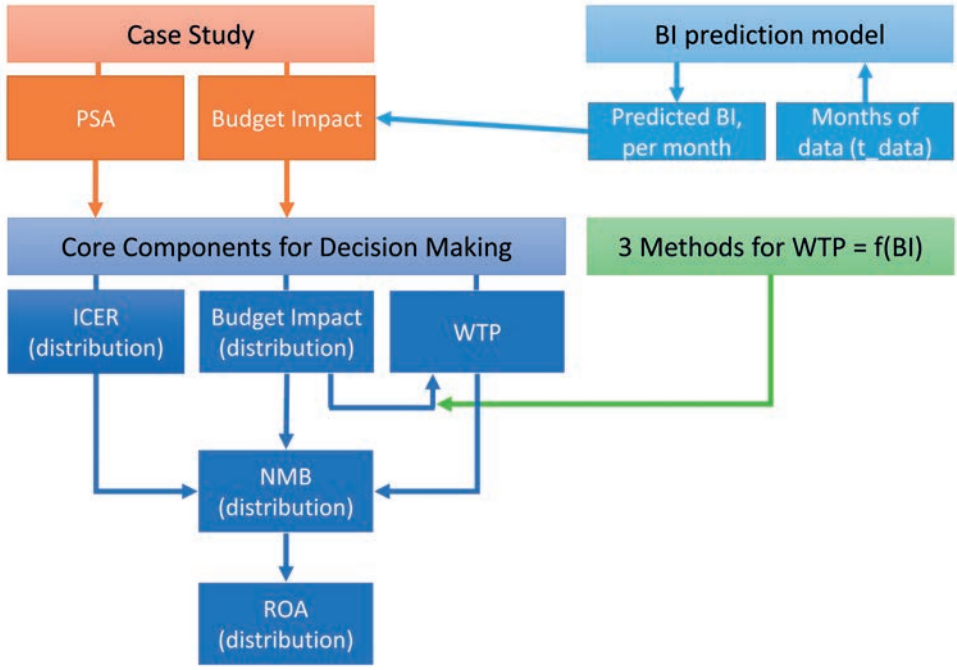


Figure 1. A schematic overview of model components. BI = Budget Impact, NMB = population Net Monetary Benefit, ROA = Real Options Analysis, PSA = probabilistic sensitivity analysis, WTP = Willingness to Pay.

The option value (OV) calculated as:

$$7. \quad Option\,Value = sNMB_{option} - sNMB_{standard}$$

A positive OV thus means $pNMB_{option} > pNMB_{standard}$ which implies that waiting one month is more valuable than immediate investment. A negative OV implies that waiting is less valuable than immediate investment. If $pNMB_{standard}$ is positive and OV is negative, investment should be initiated.

When using this methodology, observed BI is necessary to allow for ROA-based decision making which may sound counter-intuitive given that BI is typically only generated when reimbursement has been granted. In various healthcare systems, there is indeed no access prior to the reimbursement decision so postponing the decision would be a useless endeavour as no data would be observed [38]. In the Netherlands however, BI data is frequently available prior to a final reimbursement decision because of conditional reimbursement schemes or temporary

schemes where the manufacturer supplies a product for free. In the case of Opdivo, 4 months of BI data were available prior to publication of the reimbursement dossier. A recent study has shown that this is the case for many more oncology drugs in the Netherlands, therefore enabling the described ROA implementation [38].

RESULTS

Relationship between WTP and BI

The outcomes of the three *WTP = f(BI)* methods are displayed in figure 2. Each dot is a single estimate from the BI prediction model generated for the case-study using t_data = 0. As this included 45 predicted months, 100 BI and 100 ICER (and related ΔE) scenarios, the total BI estimations per assumption was 45 * 100 * 100 = 450,000.

The BI estimates were identical for the three displayed assumptions. The density plot of the BI estimates, shown in grey, is thus the same for each assumption. This density plot shows that the majority of BI estimates is in the range of €0 – €2.5 million per month. The main plot area shows that estimates up to €35 million per month are present (although the density plot shows that their relative frequency is small).

The main plot as well as the density plot indicate that, for the POINT method, WTP is very insensitive to a change in BI. The main difference between a fixed WTP of €80,000 (which would appear as a line in the main plot with a density spike at €80,000) and the POINT method appears to be that the latter has a (nearly fixed) WTP around €72,500.

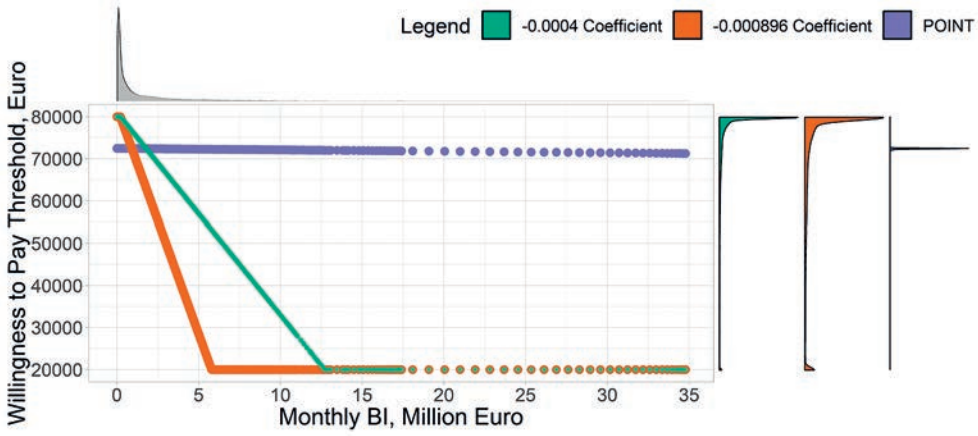


Figure 2. WTP distribution using the three *WTP = f(BI)* methods. Including density plots for both axes. The BI density plot is for the same for the 3 methods. Each method consists of 450.000 datapoints.

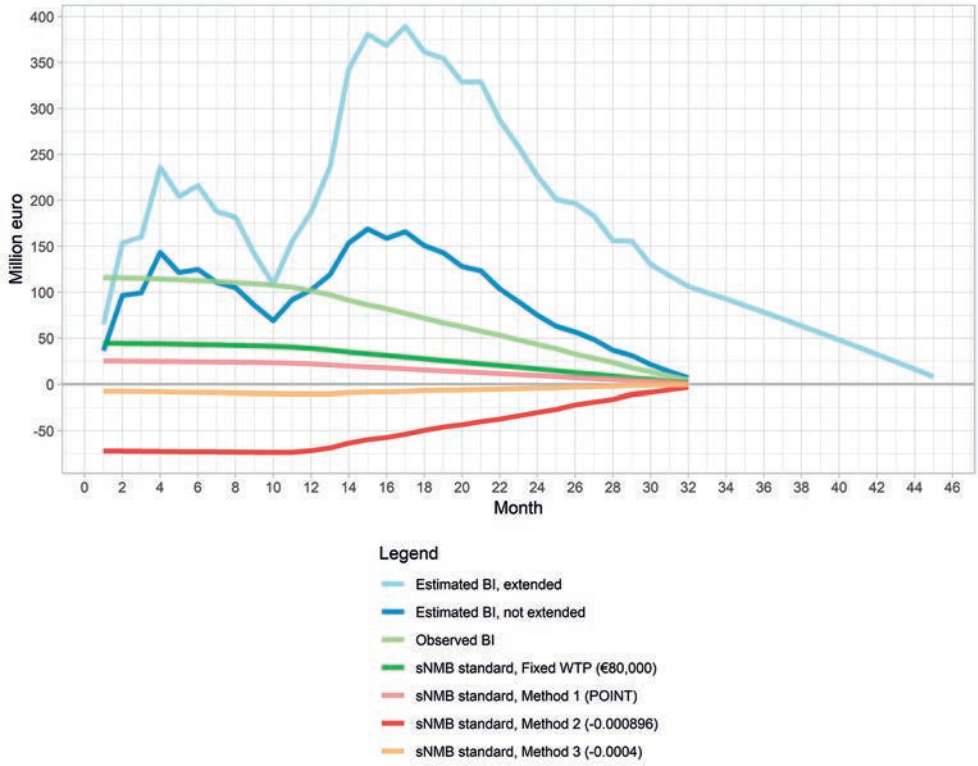


Figure 3. Observed pNMB and observed and estimated BI.

Method 2, with a coefficient of -0.000986, creates a rather large spread in WTP as is indicated in the main plot as well as the density plot. This coefficient even leads to some accumulation at the minimum WTP threshold of €20,000. The density plot of method 3 (coefficient of -0.0004) shows less sensitivity of WTP to BI and results in only very minor accumulation at the minimum WTP.

In figure 3, we display the observed and estimated BI as well as the observed pNMB per *WTP = f(BI)* method. The observed pNMB is based on the observed BI and uses the manufacturer's base case ICER (€62,277) to keep results as interpretable as possible. All data is cumulative, meaning that it includes the BI or pNMB of the remaining (i.e., future) months. The predictions of the estimated cumulative BI are updated monthly to reflect an increase in t_data, where t_data equals x – 1 for month x. The final observed Opdivo BI is in month 32 but the final estimated month is extended to 45. Therefore, not all BI predictions can be compared to observed data. For the option value calculations, we use the extended BI estimations. To compare the BI prediction model accuracy with observed BI, the non-extended estimated BI is most informative.

The BI estimations using 0 t_data was €36 million and thus overestimated observed BI (€116 million). From t_data 1 to 14, BI prediction could be considered as quite accurate. The pNMB, based on 0 t_data, was €44 million, €25 million, -€72 and -€8 million for a fixed WTP, method 1, method 2 and method 3, respectively.

Option Value Results

Figure 4 displays the results for the OV for a fixed WTP (4a) as well as the results for the 3 WTP = f(BI) methods (4b – 4d). As in figure 3, pNMB, $pNMB_{standard}$ and $pNMB_{option}$ are cumulative as they include pNMB to be gained in future months. $pNMB_{standard}$ is updated monthly when new data arrives to provide the value of a 'now or never' investment decision, given an amount of t_data. $pNMB_{option}$ is therefore also updated each month. The final observed Opdivo BI is in month 32 so $pNMB_{option}$ cannot be calculated from month 33 onwards.

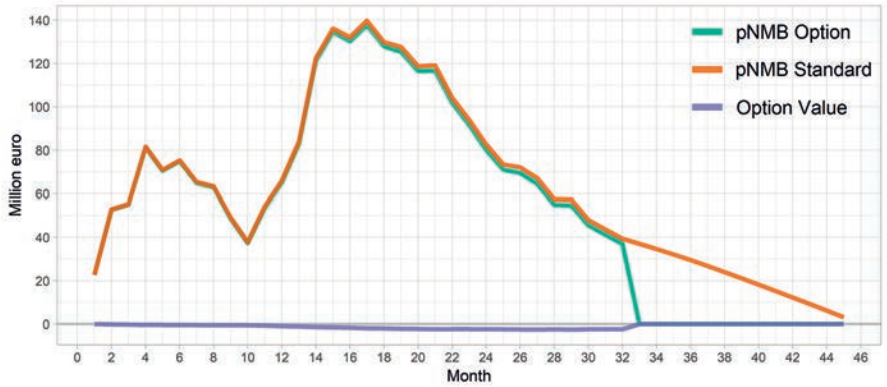


Figure 4a. Base case option value using fixed WTP (€80,000)

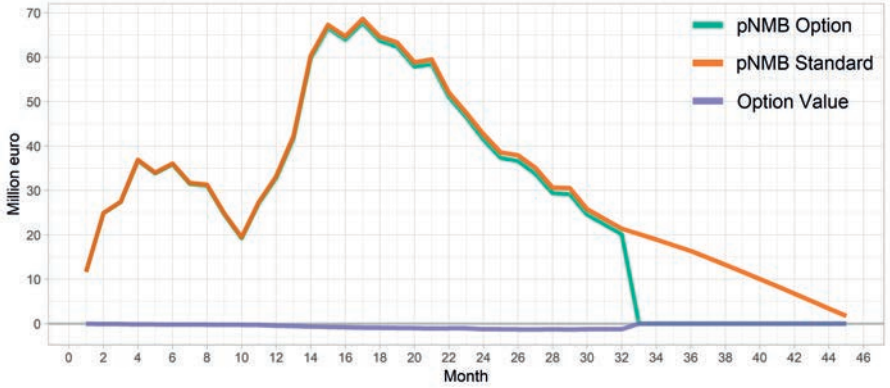


Figure 4b. Base case option value using method 1 (POINT) for *WTP = f(BI)*

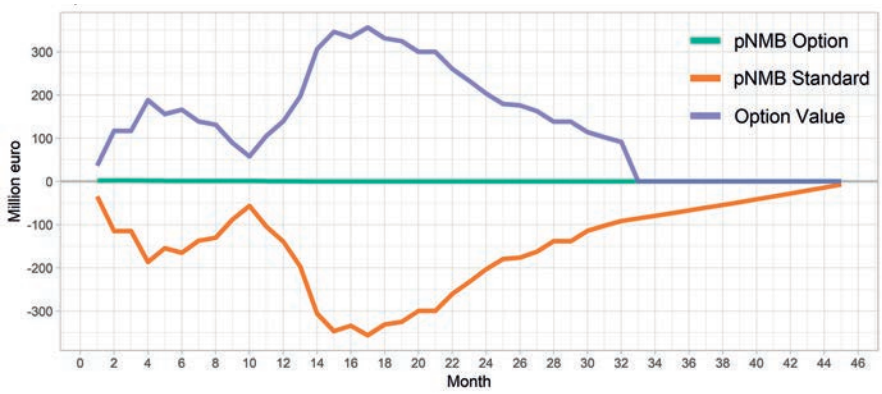


Figure 4c. Base case option value using method 2 (-0.000896 coefficient) for *WTP = f(BI)*

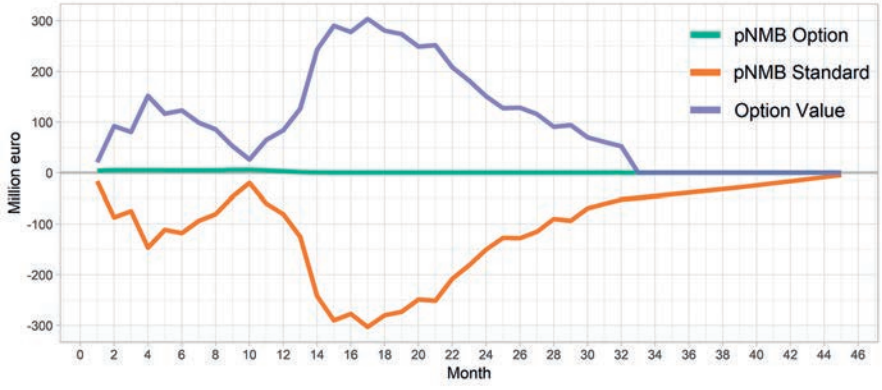


Figure 4d. Base case option value using method 3 (-0.0004 coefficient) for *WTP = f(BI)*

In figures 4a and 4b, $NMB_{standard}$ is positive for all months of data. The general shape of the positive NMB curve is caused by the prediction model results and by the number of months remaining. in these 2 figures, WTP is not (for fixed WTP) or hardly (for method 1) influenced by BI. Therefore, the sign of the pNMB is solely determined by the unobserved ICER distribution. BI scenarios are then either all negative or all positive (as in the case of 4a and 4b, given the positive $pNMB_{standard}$ and $pNMB_{option}$). So, the dip at 10 months (being t_data = 9) is caused by lower by BI estimate for months 10 – 45. From 17 months onwards, a downward trend is visible as the number of months in which pNMB can be generated declines.

Based on the same reasoning regarding the role of the ICER and BI on the sign of the pNMB, it is logical that figures 4a and 4b display a slightly negative OV. For $pNMB_{option}$, observing the true BI scenario does not lead to the exclusion of negative scenarios as there are no negative BI scenarios. Therefore, it does not provide any benefit, but it does cost 1 month of (certain) positive pNMB, hence the negative OV. Figures 4a and 4b are further proof that the POINT tool

could be regarded as a fixed WTP with a value of €72,500 as, if WTP had been more sensitive to BI, it would have generated OV.

Both figure 4c and 4d display negative $pNMB_{standard}$ for all timepoints, indicating that investment should never be carried out. The difference between 4a + 4b (fixed WTP) and 4c + 4d (dynamic WTP) is profound, so a dynamic WTP has great influence on pNMB. pNMB is more negative in 4c compared to 4d, indicating that the stronger influence of BI on WTP from method 2 (figure 4c) compared to method 3 (4d), leads to a greater loss. In 4c and 4d, $pNMB_{option}$ is close to 0, indicating that very few positive BI scenarios remain.

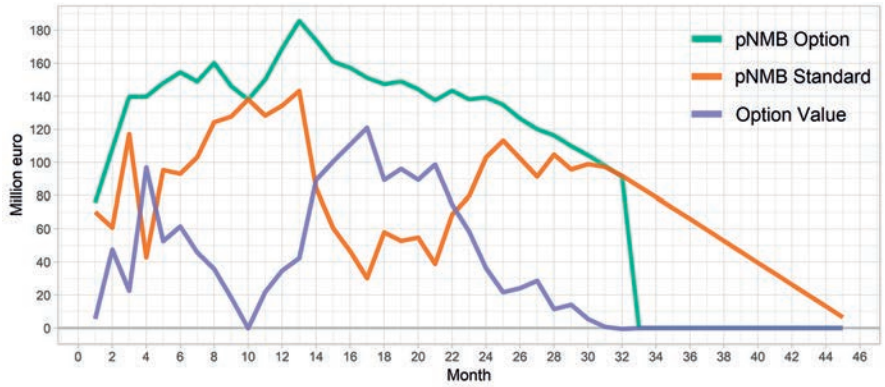


Figure 5a. Option value using method 3 for *WTP = f(BI)*, base WTP = €150,000, minimum WTP = €20,000

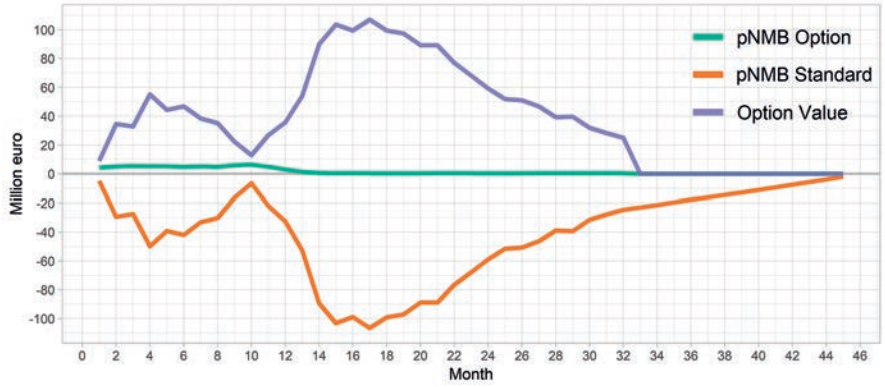


Figure 5b. Option value using method 3 for *WTP = f(BI)*, base WTP = 80,000, minimum WTP = 50,000

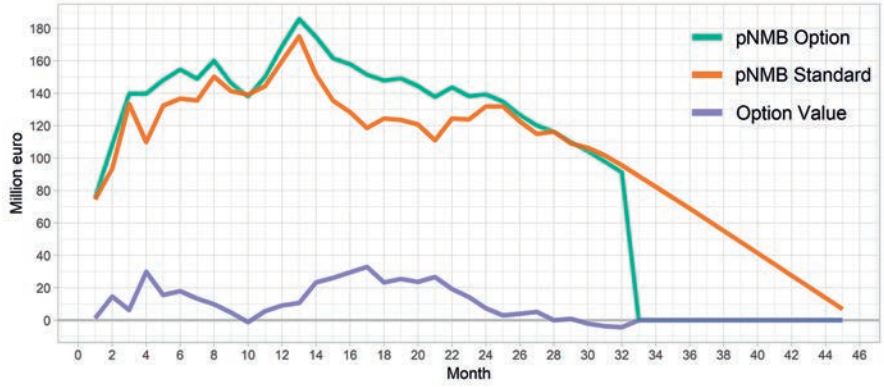


Figure 5c. Option value using method 3 for *WTP = f(BI)*, base WTP = €150,000, minimum WTP = €50,000

5.2

Option Value Results with alternative *WTP = f(BI)* parameters

Method 2 and 3 use, beside the aforementioned coefficient, base case WTP and minimum WTP to calculate WTP per BI. To investigate the influence of these 2 latter parameters, we generated results using various values for minimum WTP and base WTP. In figure 5, the results of these analyses are presented. All results presented in figure 5 are generated using the coefficient of method 3 (-0.0004).

In figure 5a, the base WTP is increased to €150,000. $pNMB_{standard}$ is positive and OV is very high, indicating that uncertainty (by means of negative BI scenarios) has quite a profound effect on $pNMB_{standard}$. On the other hand, high $pNMB_{option}$ indicates that the investment decision has a lot of potential value. According to the decision rule, one should wait for investing until month 10 as OV becomes negative with a positive $pNMB_{standard}$. When comparing this €150,000 base WTP to the €80,000 base WTP shown in figure 4d, it becomes evident that the base WTP has profound influence on $pNMB_{standard}$, $pNMB_{option}$ and resulting OV.

The influence of a higher (€50,000) minimum WTP is shown in figures 5b and 5c. When comparing 5b to 4d and 5c to 5a, it is apparent that a higher minimum WTP increases $pNMB_{standard}$ and that it lowers OV. Both these observations are expected as a higher minimum WTP effectively lowers the potential for a loss of $pNMB_{standard}$. Based on 5c, like in 5a, reimbursement should be initiated after waiting and observing until month 10.

DISCUSSION

We aimed to address two aspects that limit the use of CEA for decision-making purposes, being incorporation of timing and a unified assessment of BI and ICER as well as the role of WTP. Using Opdivo as case study, we unified BI, ICER and WTP by means of pNMB and then incorporated timing using ROA. A BI prediction model that was able to adapt to the monthly addition of data. In figures 4 and 5, we used pNMB to compare traditional 'now or never' decisions to having the option to wait more data and demonstrated that waiting can indeed provide more value and therefore potentially lead to better decision making.

pNMB was based on distributions of three main parameters. The WTP proved to be critical for determining the $pNMB_{standard}$ and therefore $pNMB_{option}$ and OV. For example, using POINT WTP (€72,500) instead of the base WTP (80,000) approximately halved the $NMB_{standard}$. In figure 5, a WTP increase from 80,000 to €150,000 changed the $pNMB_{standard}$ from deeply negative to very positive. This would not be a problem if the true value of the WTP would be known with certainty and if decision making would then be based on this WTP. Literature has however led to a wide variety of possible WTP thresholds [8–10,22].

Evidence furthermore suggests that achieving a single threshold is impossible and that a single threshold should never be used in practice, thereby suggesting that it should be related to the budget available and/or the BI of the respective innovation [45–47]. Therefore, it is not illogical that in practice, the ICER and WTP have not been the sole deciding factors for reimbursement and innovations with an ICER above the WTP have still received positive recommendations [1,3,45]. For our analyses, it is thus probably impossible and unwanted to use a single fixed WTP threshold. Instead, we should use a more dynamic approach, such as we have demonstrated with $WTP = f(BI)$ method 2 and 3 that are derived from real-world decision-making.

A second parameter used for NMB calculation was BI. The traditional BI point estimates provide insufficient data on the underlying probability distribution and they cannot easily be updated based on observed BI so a previously developed BI prediction model was used. In figure 3, we showed that predictions were rather accurate. Still, as this prediction model was only validated for the Netherlands and for oncology drugs, widespread use of this data source for ROA is limited. We have however demonstrated that such a method for BI estimation could serve as valid input data for ROA.

The third and final pNMB parameter was the ICER. As PSA typically is used to depict (parameter) uncertainty and as we were able to use the raw PSA data from the reimbursement file, we believe we have used a valid and representative data source for the ICER. We did however use model-outcomes that were based on the manufacturers' assumptions and parameters,

yielding a base-case ICER of €62,277. ZIN however used alternative assumptions which they deemed more representative of the Netherlands, resulting in a base-case ICER of €133,848. If we had used the latter ICER, all scenarios given current Dutch WTP thresholds would have had a profoundly negative pNMB. We believe that it would be less insightful and less informative to demonstrate ROA with an investment decision that would always be negative and therefore resorted to the manufacturers' scenario.

When using ROA, waiting for the arrival of data is a crucial aspect. This waiting should however be sensible in the sense that one should expect to achieve more accurate data in the future and that this increased accuracy should deliver value. When relating this to the case study, the BI prediction model validation showed that the prediction error decreases with increasing t_data (see Appendix 1). A high prediction error means that the point estimate of the BI prediction is distributed over 100 BI scenarios with a greater range of BI values than a low BI error would generate. If we then assume some form of $WTP = f(BI)$, a higher prediction error would result in lower WTPs amongst various BI scenarios (given they exceed 2.5 million annually). A lower WTP results in a lower pNMB (given the same ICER). So, a high prediction error with $WTP = f(BI)$ results in a lower pNMB. Given the statements above, waiting for more data is a sensible option as a lower error should yield a higher (and more accurate) pNMB.

Our study has the following limitations. First, the presented case study is, given all the assumptions, still more a theoretical and technical example of an implementation of ROA than it is guidance for real-world decision-making. We furthermore acknowledge that real-world decision-making is driven by many other factors than BI, ICER, WTP and their evidence levels such as disease severity, incidence, medical need and various socio-political aspects. ROA is however a proven technique for informing investment decisions in corporate environments and ROA has previously also been used in academic settings [31,33,34,43]. We have demonstrated that ROA can be used to implement the value of waiting for the gradual arrival of more data in a reimbursement decision making setting. This ROA approach would of course be more informative if additional data on the ICER would become available over time, from for example a registry or trial, and would be incorporated in the option valuation. Although including development of the ICER uncertainty is beyond the scope of this study, we believe the herein presented ROA implementation could serve as a foundation for future work and could currently serve as a tool to identify and manage BI uncertainty and its development during the period in which a reimbursement decision is to be taken.

Second, aggregating ICER and BI samples in 100 scenarios inevitably causes some loss of data and potentially reduces the influence extreme values. In other words, the tails of the distributions are condensed into the first and last scenarios. We do not believe that this is a major concern as these extreme values typically play a limited role in decision making and as 100 scenarios are able to represent a wide variety of ICER and BI values.

Third, BI data recorded prior to a reimbursement decision might not be representative of post-decision BI. The BI prediction model was however trained using pre- as well as post-decision data and the published validation showed that the resulting accuracy was adequate [38]. Although the factors driving pre- and post-reimbursement BI are different, we believe that pre-decision BI provides sufficient information on (potential) post-decision BI and is therefore adequate for informing access decisions.

Fourth, our time horizon is limited to 45 months. This causes $pNMB_{standard}$ to attenuate to 0 over time as less time remains in which pNMB can be generated. This leads to some bias as it encourages waiting as risk as well as potential benefits are now artificially reduced as time passes. ZIN however uses a time horizon of 3 years when predicting BI and also disregards BI past this 3-year timeframe [40]. In that regard, our 45-month time horizon is still imperfect but better than what is currently used in practice.

CONCLUSION

We showed that our ROA based conceptual method can be used to inform the timing of reimbursement decisions. A relationship between WTP and BI is however required to generate option value and thus for ROA to be useful. The parameters describing this WTP and BI relationship also have great influence on the outcomes. Accurate data on these parameters is therefore a prerequisite for implementation of ROA but the Dutch healthcare system currently does not provide this data. Recent literature however acknowledges the need for a more dynamic approach to WTP that includes available budget and BI. We indeed believe that BI and its influence on WTP should first be clearly defined as it is clear that in practice, BI definitely has had influence on access decisions. When the relationship between BI and WTP would eventually be defined as suggested, we have shown that our method could be suitable for providing guidance on flexible and adaptive reimbursement decisions, deemed essential in the current landscape of ever higher uncertainty at market access of new costly drugs.

ACKNOWLEDGEMENTS

We would like to thank Prof. dr. Manuela Joore for presenting this work at LoLaHESG 2019 and the useful comments she and the attendees raised. We furthermore acknowledge FarmInform for providing the budget impact data.

REFERENCES

1. Ghijben P, Gu Y, Lancsar E, Zavarsek S. Revealed and Stated Preferences of Decision Makers for Priority Setting in Health Technology Assessment: A Systematic Review. PharmacoEconomics. 2018;36:323–40.
2. Cohen JP, Stolk E, Niezen M. Role of Budget Impact in Drug Reimbursement Decisions. J Health Polit Policy Law. 2008;33:225–47.
3. Simoens S. Use of Economic Evaluation in Decision Making: Evidence and Recommendations for Improvement. Drugs. 2010;70:1917–26.
4. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. Health Technol Assess [Internet]. 2009 [cited 2019 Apr 24];13. Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta13290/
5. Carrera P, IJzerman MJ. Are current ICER thresholds outdated? Valuing medicines in the era of personalized healthcare. Expert Rev Pharmacoecon Outcomes Res. 2016;16:435–7.
6. Shiroiwa T, Sung Y-K, Fukuda T, Lang H-C, Bae S-C, Tsutani K. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ. 2010;19:422–37.
7. 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;24:1–7.
8. Adang E, Stadhouders N, Parsons C, Wammes J, Govaert P, Determann D, et al. Verdringingseffecten binnen het Nederlandse zorgstelsel [Internet]. 2018. Available from: https://www.iqhealthcare.nl/media/124601/eindrapport-verdringing-2018.pdf
9. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the NICE cost effectiveness threshold [Internet]. Centre for Health Economics, University of York; 2013 Jan. Report No.: 081cherp. Available from: https://econpapers.repec.org/paper/chyrespap/81cherp.htm
10. Ryen L, Svensson M. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. Health Econ. 2015;24:1289–301.
11. Flume M, Bardou M, Capri S, Sola-Morales O, Cunningham D, Levin L-A, et al. Approaches to manage 'affordability' of high budget impact medicines in key EU countries. J Mark Access Health Policy. 2018;6:1478539.
12. Ghabri S, Mauskopf J. The use of budget impact analysis in the economic evaluation of new medicines in Australia, England, France and the United States: relationship to cost-effectiveness analysis and methodological challenges. Eur J Health Econ. 2018;19:173–5.
13. Niezen MGH, Bont A de, Busschbach JJV, Cohen JP, Stolk EA. Finding legitimacy for the role of budget impact in drug reimbursement decisions. Int J Technol Assess Health Care. 2009;25:49–55.
14. Vooren K van de, Duranti S, Curto A, Garattini L. A Critical Systematic Review of Budget Impact Analyses on Drugs inhe EU Countries. Appl Health Econ Health Policy. 2014;12:33–40.
15. Barua S, Greenwald R, Grebely J, Dore GJ, Swan T, Taylor LE. Restrictions for Medicaid Reimbursement of Sofosbuvir for the Treatment of Hepatitis C Virus Infection in the United States. Ann Intern Med. 2015;163:215.
16. Niezen MGH, Stolk EA, Steenhoek A, Uyl-De Groot CA. Inequalities in oncology care: Economic consequences of high cost drugs. Eur J Cancer. 2006;42:2887–92.
17. Geenen JW, Boersma C, Klungel OH, Hövels AM. Accuracy of budget impact estimations and impact on patient access: a hepatitis C case study. Eur J Health Econ [Internet]. 2019 [cited 2019 May 2]; Available from: https://doi.org/10.1007/s10198-019-01048-z
18. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget

Impact Analysis Good Practice II Task Force. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2014;17:5–14.

19. Mauskopf JA, Sullivan SD, Annemans L, Caro J, Mullins CD, Nuijten M, et al. Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices—Budget Impact Analysis. Value Health. 2007;10:336–47.
20. Pauwels K, Huys I, Vogler S, Casteels M, Simoens S. Managed Entry Agreements for Oncology Drugs: Lessons from the European Experience to Inform the Future. Front Pharmacol [Internet]. 2017 [cited 2019 Apr 4];8. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2017.00171/full
21. Dakin HA, Devlin NJ, Odeyemi IAO. "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. Health Policy Amst Neth. 2006;77:352–67.
22. Lomas J, Claxton K, Martin S, Soares M. Resolving the "Cost-Effective but Unaffordable" Paradox: Estimating the Health Opportunity Costs of Nonmarginal Budget Impacts. Value Health. 2018;21:266–75.
23. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess Winch Engl. 2015;19:1–503, v–vi.
24. Geenen JW, Boersma C, Vreman RA, Knies S, Klungel OH, Hövels AM. Joint appraisal of cost-effectiveness, affordability and associated uncertainty in reimbursement decision-making: an integrated conceptual framework combining net monetary benefit and a dynamic willingness-to-pay threshold. Manuscr Prep Submiss. 2019;
25. van de Wetering EJ, van Exel J, Brouwer WBF. The Challenge of Conditional Reimbursement: Stopping Reimbursement Can Be More Difficult Than Not Starting in the First Place! Value Health. 2017;20:118–25.
26. Makady A, van Veelen A, de Boer A, Hillege H, Klungel OH, Goettsch W. Implementing managed entry agreements in practice: The Dutch reality check. Health Policy Amst Neth. 2019;123:267–74.
27. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with Evidence Development, Only in Research, Risk Sharing, or Patient Access Scheme? A Framework for Coverage Decisions. Value Health. 2012;15:570–9.
28. Carlson JJ, Gries KS, Yeung K, Sullivan SD, Garrison LP. Current Status and Trends in Performance-Based Risk-Sharing Arrangements Between Healthcare Payers and Medical Product Manufacturers. Appl Health Econ Health Policy. 2014;12:231–8.
29. Pouwels XGLV, Grutters JPC, Bindels J, Ramaekers BLT, Joore MA. Uncertainty and Coverage With Evidence Development: Does Practice Meet Theory? Value Health [Internet]. 2019 [cited 2019 May 1]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S1098301518363113
30. Vreman RA, Bouvy JC, Bloem LT, Hövels AM, Mantel-Teeuwisse AK, Leufkens HGM, et al. Weighing of Evidence by Health Technology Assessment Bodies: Retrospective Study of Reimbursement Recommendations for Conditionally Approved Drugs. Clin Pharmacol Ther. 2019;105:684–91.
31. Favato G, Baio G, Capone A, Marcellusi A, Saverio Mennini F. A Novel Method to Value Real Options in Health Care: The Case of a Multicohort Human Papillomavirus Vaccination Strategy. Clin Ther. 2013;35:904–14.
32. Minnema LA, Giezen TJ, Gardarsdottir H, Egberts TCG, Leufkens HGM, Mantel-Teeuwisse AK. Post-marketing dosing changes in the label of biologicals. Br J Clin Pharmacol. 2019;85:715–21.
33. Dixit AK, Pindyck RS. Investment Under Uncertainty. Princeton University Press; 1994.
34. Grutters JPC, Abrams KR, de Ruysscher D, Pijls-Johannesma M, Peters HJM, Beutner E, et al. When to wait for more evidence? Real options analysis in proton therapy. The Oncologist. 2011;16:1752–61.
35. Attema AE, Lugnér AK, Feenstra TL. Investment in antiviral drugs: a real options approach. Health Econ. 2010;19:1240–54.

36. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015.

37. Ross SA. Uses, Abuses, and Alternatives to the Net-Present-Value Rule. Financ Manag. 1995;24:96–102.

38. Geenen JW, Belitser SV, Vreman RA, Bloois, van M, Klungel OH, Boersma C, et al. A novel method for predicting the budget impact of innovative medicines: validation study for oncolytics. Manuscr Prep Submiss. 2019;

39. Zorginstituut Nederland. Advies Zorginstituut: longkankermiddel nivolumab tegen huidige prijs niet in basispakket - Nieuwsbericht [Internet]. 2015 [cited 2018 May 1]. Available from: https://www.zorginstituutnederland.nl/actueel/nieuws/2015/12/08/advies-zorginstituut-longkankermiddel-nivolumab-tegen-huidige-prijs-niet-in-basispakket

40. Zorginstituut Nederland. Kosteneffectiviteit in de praktijk - Rapport [Internet]. 2015 [cited 2019 Apr 24]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk

41. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg - Publicatie [Internet]. 2016 [cited 2018 Oct 23]. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg

42. Zorginstituut Nederland. Advies: longkankermiddel pembrolizumab tegen huidige prijs niet in basispakket - Nieuwsbericht [Internet]. 2016 [cited 2018 May 1]. Available from: https://www.zorginstituutnederland.nl/actueel/nieuws/2016/12/19/advies-longkankermiddel-pembrolizumab-tegen-huidige-prijs-niet-in-basispakket

43. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. J Health Econ. 2000;19:755–66.

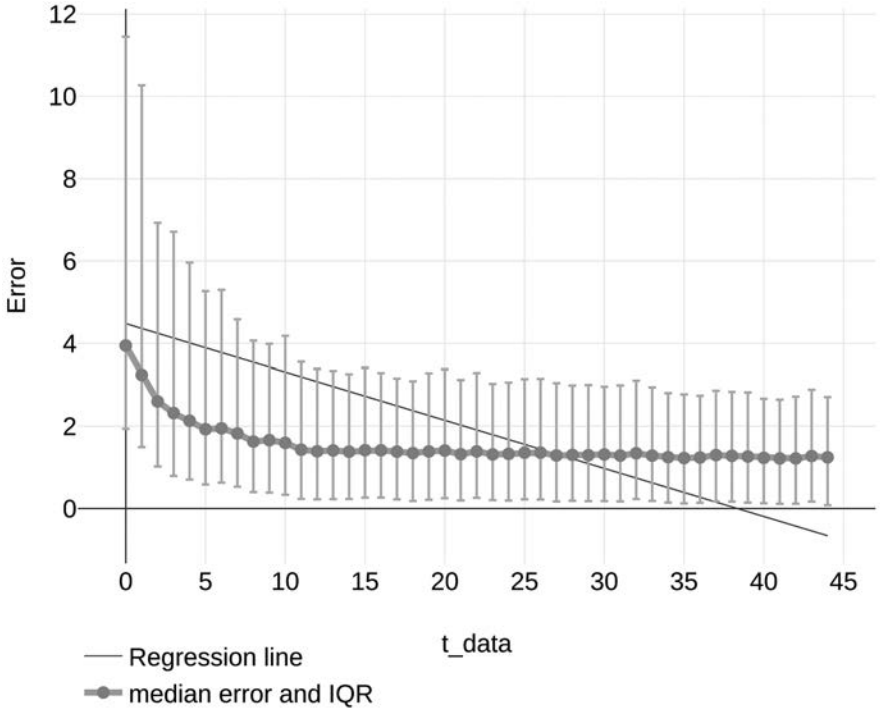
44. Mortimer D, Li JJ, Watts J, Harris A. Breaking up is hard to do: the economic impact of provisional funding contingent upon evidence development. Health Econ Policy Law. 2011;6:509–27.

45. Cleemput I, Neyt M, Thiry N, De Laet C, Leys M. Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. Int J Technol Assess Health Care. 2011;27:71–6.

46. Neumann PJ, Cohen JT, Weinstein MC. Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold. N Engl J Med. 2014;371:796–7.

47. Bertram MY, Lauer JA, De Joncheere K, Edejer T, Hutubessy R, Kieny M-P, et al. Cost–effectiveness thresholds: pros and cons. Bull World Health Organ. 2016;94:925–30.

APPENDICES



Appendix 1. Median error aggregated per t_pred. Median error aggregated per t_pred, including error bars indicating the interquartile range and the regression line. Coefficient = -0.096, se = 0.0035, p<0.0001

6

General discussion

This thesis sought to integrate affordability, cost-effectiveness and uncertainty in a single decision-making framework. Decision-makers typically appraise affordability and cost-effectiveness separately whilst these two aspects both drive the risk as well as potential benefit of innovations in healthcare. Especially the uncertainty in these two aspects needs to be interlinked and assessed as a single entity because the extent of uncertainty in ICER and BI synergistically influences the value or risk of innovations to society.

In this thesis further evidence has been provided that BI estimates are inaccurate and are therefore an important source of uncertainty to decision-makers. This uncertainty should explicitly be accommodated for, since it was demonstrated in this thesis that a relationship between BI and WTP is pivotal with potential impact on healthcare decision-making. The result of this potential relationship is that BI, as opportunity costs, influences the WTP. This link to WTP implicitly links affordability (BI) with cost-effectiveness (ICER and WTP) and was quantified as population Net Monetary Benefit (pNMB). This pNMB thereby provides a major step towards integration of affordability, cost-effectiveness and uncertainty in a single decision-making framework.

The uncertainty in BI could potentially be managed or reduced by the proposed prediction modelling approach which paved the way for integrating the aspect of time on evidence and uncertainty, culminating in the Real Options Analysis driven approach where timing of uncertainty is explicitly integrated in decision-making. By means of ROA it was demonstrated that temporality can be integrated in the unified pNMB approach, thereby providing the final step towards a unified appraisal of affordability, cost-effectiveness and the associated uncertainty and timeliness of these aspects.

Hereafter, various aspects of this thesis will be discussed in more detail, before implications and possibilities for future research will be conveyed.

sNMB TO INTEGRATE COST-EFFECTIVENESS, AFFORDABILITY & UNCERTAINTY

The balancing of cost-effectiveness & affordability, set out in chapter 5.1, is a longstanding challenge to decision-makers. Although, many publications address this topic, none have so far explicitly incorporated the joint influence of uncertainty in both cost-effectiveness and affordability [1–3]. This approach faces challenges before it can be implemented but can yield great advantages to the healthcare system.

Especially when BI has a great influence on marginal cost-effectiveness, then even a small probability on a very high BI poses a great risk to a healthcare system. Indeed, risk could be defined as probability * impact, described by for example Klinke & Renn [4]. As previously

described, the 'probability' aspect of affordability is currently omitted in BI analyses by the lack of BI probability distributions. For this, chapter 4.1 provides a first step towards a potential solution. The 'impact' aspect of BI is as of yet unclear as there is a mismatch between empirical evidence on the influence of BI on marginal cost-effectiveness and the influence of BI on real-world decision-making. So, current decision-making practice is partly aimed at preventing budgetary risk, but is ignorant of the underlying probabilities and types of impact.

For cost-effectiveness, adequate methods exist to quantify cost-effectiveness and the associated uncertainty and we believe that these methods are properly employed in the Netherlands to yield scientifically valid reimbursement dossiers.

When assessing the risk associated with a reimbursement decision, it is clear that having reimbursed a medicine with an ex-post unfavourable ICER and an ex-post high BI has resulted in greater losses than a medicine with a low ex-post BI. Loss in this sense can pertain to net expenditure as well as opportunity costs. This logic also holds for potential gains in the event of a favourable ICER, again indicating the intertwined nature of affordability, cost-effectiveness and uncertainty. The only way then for proper decision-making, managing risks as well as potential benefits, is a combined appraisal. In chapter 5, the pNMB approach is portrayed which aims to solve this issue.

Potential & Pitfalls of pNMB

The presentation of the pNMB approach in chapter 5.1 is somewhat synthetic as simulated BI data was used. Besides, the ICER estimate that was used was supplied by the manufacturer but was rejected by the Dutch Health Care Institute (ZIN). As this chapter is critical to answering this thesis' main objective, we wish to elaborate on some of the aspects of this study that are especially relevant to potential implementation by decision-makers.

BI data and the BI distribution were simulated but in reality, especially the latter is unknown. Efforts to fill this knowledge gap are set out in chapter 4.1 but, as will be described later in this discussion, requires further work. The simulated approach, where the BI probability distribution is based on empirical evidence (e.g., Keeping et al), would still be superior to the current nondescript ranges surrounding BI estimates [5]. Therefore, the assumed BI distributions should not be a major hurdle for implementation.

The current mismatch between empirical evidence on the relationship between WTP and BI and the influence of BI on decision-making is however a significant hurdle. When BI has as little influence on WTP as suggested by Adang and Lomas, affordability should not be considered for reimbursement decision-making and decisions should be solely based on ICER and WTP [6,7]. This is in stark contrast with decision-making practice. The relationship between WTP and BI has major influence on the outcomes of chapters 5.1 and 5.2 and requires elucidation before

these methods can be implemented. Given the resources that are devoted to health-care and the impact of reimbursement decisions, it is required to further research this relationship and attempt to realign the current mismatch.

The existence of any form of a dynamic WTP is however of paramount importance to chapter 5 as if WTP would be completely static, BI would have no role in a pNMB guided reimbursement decision. This existence of a dynamic WTP can however be deduced convincingly: As was set out in the general introduction, the WTP can determined by a combination of striving for equity as well as efficiency. Affordability concerns, which in the Netherlands influence reimbursement decisions, must indeed be caused by maintaining or striving for some level of efficiency and therefore imply that opportunity cost must play a role and must therefore influence the WTP. Therefore, it was assumed that in any healthcare system, at least some efficiency is strived for so that for any system, opportunity costs and marginal benefits are relevant. These assumptions are in line with literature, and therefore the assumption that WTP is dynamic (in chapters 5.1 and 5.2) is justified [6–9].

In the presented pNMB approach, the number of treated patients is derived from the total treatment cost per patient. As is shown in chapter 5.1, this approach underestimates the number of patients treated in the first for treatments with a duration of multiple years. This problem is especially relevant for life-long treatments in for example cystic fibrosis or rheumatoid arthritis. This would be solved when BI data would be combined with data on the number of patients, data sources that are both available to ZIN. The approach for multi-year treatment should of course be validated and until then, the herein presented approach can only be used for treatment durations up to one year. Furthermore, pNMB should ideally be extended to include the first three years (as is the case for BI) instead of the current one-year timeframe [10].

Although these challenges are significant, the potential benefits are also major: an integrated approach could lead to improved decision-making, pave the way for more dynamic access and for new more transparent pharmaceutical pricing policies. In the final part of this discussion, these merits are discussed in detail.

BUDGET IMPACT: ACCURACY AND VALIDITY IN DECISION-MAKING

In 2013, Cha et al. published the aptly named paper ‘Pharmaceutical forecasting: throwing darts?’ where they reported the low accuracy of drug turnover forecasts used by manufacturers or investors [11]. Further work, for example by Broder et al. in 2017, assessed the accuracy of US estimates of sales forecasts and also concluded that accuracy of such estimates is poor [12]. The 2018 publication by Keeping et al. is especially relevant in this regard as they specifically

assessed the accuracy of BI estimates that were used to inform Welsh reimbursement decision-making [5]. Findings from this thesis confirm these conclusions (chapters 3.1 and 3.2).

Therefore, current evidence suggests that BI estimates are generally inaccurate should at best be regarded as crude approximations or "guestimates". This is not due to ignorance, negligence or lack of skill of the parties conducting these BI estimations. To the contrary, it is most likely due to the complexity of the real world that it is extremely difficult to make an accurate a-priori assessment of future BI. This does not exonerate the scientific community from efforts to improve BI estimation methodology but it does require decision-makers to accept that BI estimates are currently merely scientifically substantiated suppositions whose inaccuracy is usually shrouded by means of tables, calculations and (supposed) multiple-digit accuracy.

Inaccuracy of these estimates is not an issue if they merely serve an explorative purpose. In many jurisdictions however (like England and the Netherlands), estimated BI has a formal role in reimbursement decision-making [5,13–19]. Although this role is not as strictly enforced in the Netherlands as it is in England, a multitude of examples (e.g. Sovaldi, chapter 3.1) show that BI does definitely have a role in Dutch reimbursement decision-making [13,20–26].

In the Netherlands, high BI estimates can lead to (temporary) postponement of reimbursement by means of the "lock" (pakketsluis) policy [27]. This restrictive policy is only lifted if negotiations regarding price and/or volume lead to a (confidential) result that is deemed satisfactory to the Minister of Health, Welfare and Sport (MoH) [27]. From the perspective of the marketing authorisation holder (MAH), postponement of access and potential for lower price/volume is unfavourable. The payer however, receives a formal opportunity for price-negotiations, as long as BI is deemed high [16,18,19,28].

Of many products with high BI estimates, (e.g. Sovaldi, Harvoni, Opdivo) the reimbursement dossier states (e.g., Opdivo) or implies that this high BI causes displacement of current (or potential future) more cost-effective care [21,22,29].

Displacement effects have been widely described, for example by Claxton et al. and Lomas et al, who demonstrated and quantified displacement effects (materialised as opportunity costs) [6,7,30–32]. Given that displacement effects are very frequently mentioned as reasons to limit reimbursement and/or initiate negotiations, the evidential foundation for these effects should be solid.

Surprisingly, Lomas & Adang both conclude that BI has very limited influence on marginal cost-effectiveness (see chapter 5.1) [6,7]. In an assumed Dutch BI bandwidth of up to €250 million per year, marginal cost effectiveness is hardly influenced according to these empirical studies.

There thus appears to be no convincing evidentiary basis for the decisions that are based on these (inaccurate) BI estimates, warranting the question whether BI is currently used adequately. Given the high-prices and typically high ICER of newly introduced medicines, it is however understandable that policymakers use the available tools (e.g., price negotiation based on BI) to limit these prices. In the final section of this discussion, implications of these findings will be described.

IMPROVEMENT OF BI ESTIMATION: POTENTIAL FOR DATA-DRIVEN METHODS

As mentioned in the previous paragraph: Revisiting BI estimation is required for BI to have a justified role in reimbursement decision. In chapter 4.1, a new method is presented to potentially secure the future legitimacy of BI, which however should mainly be regarded as fundamental step and not as a finalised method that can readily be implemented.

Currently, the system allows for flexibility in the early HTA setting but the post-marketing authorisation reality is still deemed too rigid [33,34]. The adaptive and dynamic approach described here does therefore not fit in current decision-making practice. The requirement of, for example, approximately 6 – 12 months of BI data before prediction error stabilises results in the need for a system with prolonged temporary reimbursement before a decision would be taken. If such a system were to be devised, it would be worthwhile to include (data on) the ICER in a similar dynamic fashion.

It therefore seems inevitable that HTA in general but also payers move to a system where real-world but also near real-time data play a crucial role. Data-science has recently (disruptively) transformed many different sectors and there is no reason to believe that data won't have influence on the regulation and access of new medicines. The proposed method is a small fundamental step towards integrating BI dynamics (or other real-world data) in decision-making and not a readily implementable solution.

Measures of BI prediction accuracy

There is a wide variety of measures to define prediction accuracy, some of which are described in chapters 3.1, 3.2 and 4.1 and are visualised in chapter 3.2.

When aggregating accuracy data, a practice that is bound to happen if more data-driven approaches are implemented, this is especially relevant. To illustrate this, the accuracy results from the primary analysis set from chapter 3.2, table 2 are used: Mean accuracy was 0.64 but the mean symmetric accuracy was 2.5.

Accuracy (e.g., defined as observed / estimated) is not symmetric which in itself causes bias and is especially troublesome if there are samples with very high and low accuracy. This asymmetry, caused by accuracy of overestimations ranging from 1 – 0 whilst underestimations range from 1 – ∞, results in accuracy measures being predominantly influenced by underestimations. A second limitation of using such outcomes is that when processing accuracy estimates in an additive manner, similar and relatively small deviations (i.e., 100 observed and 90 vs 110 estimated) yield an aggregated accuracy that is nearly 100%.

The latter limitation might be acceptable if a decision-maker's primary concern is to ensure that the total estimated BI is close to the total observed BI, thereby ignoring accuracy of individual products. Still, the former limitation will then introduce bias as over- and underestimations of individual products were significant (chapter 3.2).

The log-transformed measure of accuracy in chapters 3.2 and 4.1 is derived from the formula that Törnqvist et al. postulated in 1985 [35]. They described that Ln (estimated / observed) should be used to relative change as it is symmetric and additive. The resulting outcomes on Ln scale are however hard to interpret. This was solved by exponentiating the absolute log value. Whilst still being additive and symmetric, the sign (i.e., under-or overprediction) is lost (denoted as directionality). This is of course critical information, especially when the ratio of over- vs underestimation is not 1.

In the example from chapter 3.2, overestimations have clearly cancelled out underestimations as 0.64 (or its reciprocal 1.56) are clearly not equal to the symmetric error of 2.50. Clearly, only providing one of these outcomes is inadequate for providing insight in the underlying data.

To describe an average, the mean as well as the median are typically presented. A similar approach for describing (aggregated) accuracy is proposed: provide an asymmetric and non-additive but directional measure (i.e., observed / predicted) as well as the proposed log-transformed unidirectional but fully symmetric and additive measure, potentially supplemented by the ratio of overprediction / underpredictions. This may appear complex but is probably unavoidable as was found that no single (interpretable) number is able to properly convey accuracy outcomes, similar to the fact that using only a mean or only a median is incapable of properly describing the average of a non-normally distributed sample.

TEMPORAL ASPECTS OF UNCERTAINTY AND DECISION-MAKING

In chapter 3, it is shown that timing of decisions respective to market conditions and market dynamics is a crucial factor in BI estimation. This supports the aforementioned need for temporal integration of uncertainty in an early-access setting (2.1, 2.2, 4.1) as well as during access (3.1, 3.2, 4.1).

The distinction between early (pre-access) and standard HTA (peri/ post-access) might however be too large and the flexibility and incorporation of timing appears to be reduced once an initial access decision has been taken. Crucially, a post-decision loss of flexibility might not reflect continued temporal development of uncertainty. Decision-makers' exigency to unite and homogenise uncertainty management as well as opportunities to ameliorate the current deficiencies will now be portrayed.

Early HTA & Conditional access schemes

In early HTA, the existence of a relationship between time and uncertainty is explicitly recognised as early HTA is, at least partly, aimed at identifying main sources of uncertainty that are most valuable or most crucial to resolve within certain timeframes [36,37]. Furthermore, early HTA assumes a degree of flexibility regarding research, development and evidence generation plans as well as regulatory dialogue [36]. The examples presented in chapters 2.1 and 2.2. could, in this regard, be used for guiding research and development pathways as well as pricing strategies. Both examples implicitly incorporate the notion and understanding that time and uncertainty are inherently intertwined.

The impact of time on evidence and uncertainty is also acknowledged by the existence of conditional access schemes. Conditional access, herein described as either a form of conditional marketing authorisation or conditional reimbursement, is aimed at providing flexibility regarding an initial access decision by granting early access whilst accepting higher uncertainty [38–40]. This uncertainty should then resolve over time and, at some point, a final decision should be taken. Although these schemes have not always lead to the timely delivery of adequate evidence, they do highlight the need (and demand) for flexibility regarding access decisions [41–45].

Loss of flexibility once access is granted

Various reasons cause initial flexibility and timeliness to be lost once access is granted. First, in an early HTA setting, the price of an intervention is a variable or even an outcome (Chapters 2.1 & 2.2). Once an initial price for an innovative medicine has been set, pricing generally remain quite stable until patent expiration. In other words, an intervention's price currently changes from a variable to a constant peri-access.

Second, the implied flexibility and timeliness of conditional access schemes might be misleading. Various sources state that obligatory post marketing studies as part of conditional access schemes are frequently delayed or completely omitted [41–45]. Even if adequate evidence is collected, evidence has shown that policy-makers struggle with revoking (conditional) reimbursement once a product has been granted market access [41,46]. It is not surprising that policy-makers have less degrees of freedom for a product to which patients have access as decisions carry increased political and societal significance. Still, these observations lead to the conclusion that the seemingly dynamic and gradual transition from an early to post-authorisation setting is in reality a stark divide. Crucially, this foregoes the continued development of evidence and uncertainty over time.

Third, as mentioned in chapter 5.2, current decision-making lacks the possibility to postpone a decision. This limitation restricts the flexibility of the reimbursement decision itself, and this decision is typically final.

Potential for Real Options Analysis

Given the immense resources that companies invest and the fierce competition in many markets, active management of investments and including timing & uncertainty in investment decisions is an essential aspect of corporate management [47–49]. To manage this, numerous companies use the method of Real Options Analysis (ROA) [47,50]. If it is a proven tool that aids companies to manage investments and prevail amidst global competition, why don't decision-makers use this to manage publicly funded investments such as new medicines?

The potential merits of ROA are described in chapter 5.2 and provide a first (but still imperfect) glimpse of a dynamic ROA driven reimbursement framework. Our proposed integration of affordability & cost-effectiveness requires ICER development over time also to be included but this lacks in our current ROA implementation due to lack of data. If this were to be implemented, we believe ROA could allow for the following:

To make the optimal decision at the right time

The right decision can only be taken when cost-effectiveness, affordability and uncertainty are integrated and jointly appraised for which we have devised our integrated pNMB approach. Timing the actual reimbursement decision, based on pNMB, is then informed using ROA.

Earlier work, by for example Mohseninejad, mainly investigated the required duration of conditional access schemes before a final decision could be taken [51]. Makady et al. however, showed that this type of access scheme had various practical issues and that data was not timely delivered, thereby limiting the use of Mohseninejad's work [41]. Furthermore, ROA has been used in various case-studies but none have used the semi-continuous approach were new data is added monthly [52–56]. Besides, none provide means to integrate cost-effectiveness and affordability like we have presented.

Provide better informed Managed Entry Agreements & more Dynamic Access

Managed entry agreements are discussed in chapter 5.1 and pertain to limiting of uncertainty and risk in affordability and / or cost-effectiveness whilst providing (early) access [41,57,58]. Also, when designing MEAs, the synergy between uncertainty in BI and uncertainty in ICER is what determines value to the payer. In that sense, value is a bivariate distribution of ICER and BI comparable to figures 2 and 3 from chapter 2.1. Therefore, it is very difficult to convey bounds or ranges of one parameter (e.g. BI) whilst not defining the other parameter. Without integrating these and their uncertainty, properly defining ranges or bounds on ICER and/or BI is thus nearly impossible. Informing on these critical but intertwined parameters is therefore crucial for designing and managing MEAs as they should safeguard value for society under risky or uncertain circumstances.

Chapter 5 provides this information in a graphical way. The pNMB results based on a fixed WTP (Chapter 5.1, figure 2) indicated that the distinction between a positive and negative pNMB was defined by a horizontal line at ICER scenario 62, so ICER scenarios 1 – 62 yielded a positive pNMB whilst ICER scenarios 63 – 100 yielded negative pNMB and BI thus does not influence whether pNMB is positive or negative. A MEA with the goal of managing this scenario should therefore only have to focus on maintaining this specific ICER threshold. Affordability (so volume of the product) does not matter in this scenario.

When using the example of the hepatitis-C derived dynamic WTP (specifically chapter 5.1, figure 3), criteria for a positive pNMB are more stringent. Crucially, no single ICER or BI scenario yields a 100% probability for positive pNMB. Contrary to the previous example, a potential MEA should therefore be designed so that it can confer strict restrictions on the ICER but also on BI. This scenario would likely be risky for a payer.

The pNMB results that were generated using an arbitrary relationship between BI and WTP (chapter 5.1, figure 4), would provide a payer with multiple choices regarding potential MEAs. ICER scenarios 0 – 5 yield a positive pNMB, regardless of BI and therefore reflect the MEA of the first (fixed WTP) example: if the manufacturer can guarantee this very specific ICER, there will be no restrictions on BI. Alternatively, a MEA restricting ICER and BI could be proposed.

When this approach is extended over time by monitoring of BI and ICER, it paves the way for more dynamic access schemes. If, for example, observed BI appears to be lower than estimated, the pNMB approach from chapter 5.1 can inform on the ICER range that this allows for. If the lower BI indeed permits a higher ICER, it could allow for (an informed) decision on expanding indications.

As is shown in chapter 3.1, 3.2 and 4.1, market dynamics cannot be predicted as it requires active management of costly interventions, like corporate investments would. So when, for example,

pNMB would turn negative over time, our method could inform on which actions should be taken in terms of BI and ICER.

The need and ability to more closely manage and monitor the value of interventions during their life-cycles will only increase in the future where data is bound to become ever more available. As such, the presented pNMB & ROA methods provide foundations to allow for more dynamic access in the future.

Provide means to guide dynamic pricing

Affordability and cost-effectiveness are intrinsically linked with the price of the intervention. In the aforementioned pNMB & ROA based reimbursement model, the price of the medicines could of course be a considered a variable and the model could thus inform on pricing. Specifically, the pNMB guided ROA approach could pave the way for dynamic and transparent pricing schemes.

pNMB varies based on BI and ICER estimates, including associated uncertainty. The decision-rule states that only a positive pNMB leads to reimbursement. It would also be possible to use this approach to set the price required (or warranted) for the innovation to yield a positive pNMB as outcome, similar to the approach presented in chapter 2.1. Evidence development, where it is assumed that it only lowers uncertainty and ICER and BI estimates stay the same, will then automatically lead to a higher price. This could provide an incentive for manufacturers to indeed undertake additional evidence generation activities and could be more practical and transparent than Value of Information (VOI) analyses.

For the public or for a payer, such a pricing model could also bring opportunities: As the input data for the pNMB calculation & ROA methodology, as well as the underlying formulas are all public (or published after the initial decision), it could pave the way for transparent pricing policies. Especially the public (or tax-payers), who are currently excluded from the confidential status-quo between manufacturers and the MoH, could benefit from a model that is fully transparent.

From the perspective of a payer (or the institution tasked with undertaking price negotiations), such a model could also help to convey drug pricing or (managed) reimbursement policies to the public: The graphical presentation of pNMB results (chapter 5.1) clearly indicate the combinations of ICER and BI that are required for a product to beneficial to a health system. This information can then inform on the actions (e.g. on price, volume, cost-effectiveness or uncertainty) needed to ensure that a product is indeed beneficial.

LIMITATIONS

For the presented studies, several limitations and challenges for implementation are to be addressed.

In chapter 2, the primary limitations are due to parameter uncertainty. For chapter 2.1, the deterministic sensitivity analyses showed that especially the incidence of angiotensin converting enzyme inhibitor (ACEi) induced angioedema, the incidence of intensive care admission and mortality related to this adverse event and the price difference between ACEis and the alternative treatment had a relatively large influence on the ICER. Of these, the estimate of mortality was the most uncertain due to very little and low-quality evidence on lethal cases. Incidence of ACEi induced angioedema, which had the largest relative influence on the ICER, was based on a large meta-analysis and was therefore assumed to be relatively accurate.

In chapter 2.2, the presence of parameter uncertainty is evident given the phase I / II setting of this study. Main limitations are related to resource use estimated by expert opinion, the suboptimal nature of reconstructing individual patient data from published survival curves and uncertainty regarding the relative efficacy of acalabrutinib (modelled as a hazard ratio). Furthermore, an inherent limitation of partitioned survival models is that underlying events are not modelled, thereby potentially leading to biased long-term survival estimates.

The transformation of BI data, composed of volume multiplied by price, to the number of patients treated can be regarded as the main limitation of chapter 3.1. In chapter 5.1, a similar limitation is described which pertains to an underestimation of patients treated when treatment durations exceed one year. The hepatitis C treatments discussed in chapter 3.1, all have treatment durations shorter than one year.

Chapter 3.2 is limited by a relatively small number of products that were included in the primary analysis. Second, off-label use and indication extensions for which no BI analysis was published were not incorporated in the analysis, thereby potentially influencing the results. Both these limitations were primarily caused by lacking information on the indications for which the BI of the included products was generated (i.e., the indication for which the products were prescribed), thereby necessitating the assumptions made in this study.

Methodologically, various issues would need to be resolved before the BI prediction modelling approach set out in chapter 4.1 would suitable to use in practice. As the approach was completely data-driven and did not consider the influence of specific covariates, the influence of specific drug characteristics is unknown. Besides, there is no information on the magnitude of the fixed effects compared to the magnitude of the random effects. Furthermore, a full leave-one-out cross validation should be performed to quantify the influence of individual products on the estimation error, thereby increasing the chance that the results were not due to overfitting.

Lastly, it is described that at some point in time, the model structure itself should be updated but guidance currently lacks as to when this should be done. Of these issues, the last two (overfitting and time to model-structure update) are probably the most prudent and definitely require more insight.

The main limitation of chapter 4.2 is associated with bridging the gap between the presented methodology and implementation in practice. Familiarising decision-makers with the underlying technology (R and Rstudio) as well as the PSA-ReD plots (e.g., definition of the axes, interpreting the colours) will be critical for adoption. The supplied script, GitHub repository with a readme and an example datafile as well as the technical appendix are all aimed at supplementing the presented study in order to facilitate implementation.

As was covered earlier in this discussion, the main limitations pertaining to chapter 5.1 are uncertainty regarding the role of BI on WTP, use of simulated BI data and underestimation for number of treated patients for treatments that span multiple years. Similarly, the main limitations of chapter 5.2 have partly been set out in the previous sections. In short, these refer to the use of the BI prediction model, dependency on the BI and WTP relation discussed in chapter 5.1 and the lacking implementation of development of the evidence and uncertainty of the ICER.

Major challenges for the herein presented pNMB and ROA approach (i.e., chapter 5), and therefore targets for future research, are:

1. Further elucidating the previously discussed role of WTP in decision-making and the relationship between BI and WTP.
2. Incorporating the temporal development of evidence on the ICER in the ROA framework.
3. Familiarising decision-makers with the pNMB method as well as with ROA to identify further potential hurdles.

CONCLUSIONS

In this thesis, the following was shown:

- The two early HTA case studies were able to highlight critical parameter values required for cost-effectiveness and highlight limitations of the current reimbursement frameworks. They also highlight the need for timeliness and flexibility of decision-making.
- Dutch BI estimates, used for oncology and hepatitis C drugs, were inaccurate. The reported (in) accuracy was largely in line with evidence from other jurisdictions.
- A new method for estimating BI was developed and the validation showed that this method is superior regarding insight in uncertainty, the ability to update predictions and the number of future months for which predictions were made, whilst providing predictive accuracy that appears to be superior to currently used BI estimations. The novel method for displaying

results from probabilistic sensitivity analysis provided more information than the traditional scatterplot, potentially improving model validity and interpretation.

- The pNMB concept manages to integrate affordability, cost-effectiveness and uncertainty and was expanded to include timing and timeliness of uncertainty using Real Options Analysis.

These findings can lead to the following implications:

1. Making the optimal decision at the right time

When policy-makers require affordability to be part of decision-making and not merely for price negotiations, affordability must be appraised in conjunction with cost-effectiveness and uncertainty. The pNMB approach is the only method that currently achieves this and can include time and flexibility using ROA. Not only can better decisions be taken by properly integrating probability-weighed risks and benefits, the right time to take a decision can also be established.

2. Revisiting Budget Impact: take it or leave it

A revisited role of BI in decision-making is proposed: The current evidence-base on the inaccuracy of BI estimates, supplemented by the work presented in this thesis, unequivocally requires decision-makers to either accept that either BI in its current form can be no more than a tool for price-negotiations. However, BI by means of opportunity costs, definitely has a role in driving reimbursement decisions. In order to warrant such a role for BI, BI estimation techniques should be improved. This thesis provides the foundation of a new BI estimation paradigm that does justice to the opportunity costs associated with public funding of medicines.

3. Moving towards a continuous access paradigm

The bivariate distribution of pNMB using BI and ICER can be a very beneficial tool for designing managed entry agreements for novel medicines. Combined with incorporation of timeliness by for example ROA, this approach could lead to an actively managed more flexible and continuous access paradigm here the distinction between early access and regular is finally seamless.

4. Medicine Pricing

Medicine prices can be used as a variable in our presented pNMB approach. Then using a threshold approach, the maximum price given the current evidence on cost-effectiveness and BI, including uncertainty, could be generated. This fully transparent method can easily be updated over time and can thus deliver a new pricing model where effectiveness, quality of evidence and affordability transparently coalesce into a product's maximum price.

To conclude, an integrated approach of cost-effectiveness, affordability and the associated uncertainty has been developed: The herein presented pNMB-driven method that integrates these aspects allows decision-makers to conduct a single, integrated appraisal of all possible BI and ICER scenarios that are driven by the inevitable uncertainty that accompanies these

outcomes. This pNMB method is therefore able to provide decision-makers with information on the potential risk and value of medicines that was not possible before. By combining pNMB with ROA, the timing of decisions and the timeliness of evidence and uncertainty can be integrated within this framework, allowing for this integrated appraisal to be available during the entire product lifecycle. Combination of ROA and pNMB could therefore pave the way for a more continuous access paradigm, more transparent and dynamic medicine pricing and crucially, improved reimbursement decision-making.

REFERENCES

1. Pearson SD. The ICER Value Framework: Integrating Cost Effectiveness and Affordability in the Assessment of Health Care Value. Value Health. 2018;21:258–65.
2. Chhatwal J, Kanwal F, Roberts MS, Dunn MA. Cost-effectiveness and budget impact of hepatitis C virus treatment with sofosbuvir and ledipasvir in the United States. Ann Intern Med. 2015;162:397–406.
3. Birch S, Gafni A. The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold. J Health Serv Res Policy. 2006;11:46–51.
4. Klinke A, Renn O. Systemic Risks as Challenge for Policy Making in Risk Governance. Forum Qual Sozialforschung Forum Qual Soc Res [Internet]. 2006 [cited 2019 Jul 2];7. Available from: http://www.qualitative-research.net/index.php/fqs/article/view/64
5. Keeping S, Deslandes PN, Haines KE, Routledge PA. Estimated Versus Observed Expenditure Associated with Medicines Recommended by the All Wales Medicines Strategy Group. PharmacoEconomics - Open [Internet]. 2019 [cited 2019 Jan 22]; Available from: https://doi.org/10.1007/s41669-019-0116-5
6. Lomas J, Claxton K, Martin S, Soares M. Resolving the "Cost-Effective but Unaffordable" Paradox: Estimating the Health Opportunity Costs of Nonmarginal Budget Impacts. Value Health. 2018;21:266–75.
7. Adang E, Stadhouders N, Parsons C, Wammes J, Govaert P, Determann D, et al. Verdringingseffecten binnen het Nederlandse zorgstelsel [Internet]. 2018. Available from: https://www.iqhealthcare.nl/media/124601/eindrapport-verdringing-2018.pdf
8. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess Winch Engl. 2015;19:1–503, v–vi.
9. Sculpher M, Claxton K, Pearson SD. Developing a Value Framework: The Need to Reflect the Opportunity Costs of Funding Decisions. Value Health. 2017;20:234–9.
10. Zorginstituut Nederland. Format Budget Impact Analyse GVS (voorheen kostenconsequentieraming) - Publicatie [Internet]. 2016 [cited 2018 Mar 23]. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/04/06/format-budget-impact-analyse-gvs-voorheen-kostenconsequentieraming
11. Cha M, Rifai B, Sarraf P. Pharmaceutical forecasting: throwing darts? Nat Rev Drug Discov. 2013;12:737–8.
12. Broder MS, Zambrano JM, Lee J, Marken RS. Systematic bias in predictions of new drugs' budget impact: analysis of a sample of recent US drug launches. Curr Med Res Opin. 2017;1–9.
13. Zorginstituut Nederland. Pakketadvies in de praktijk: wikken en wegen voor een rechtvaardig pakket - Rapport [Internet]. 2017 [cited 2019 May 29]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2017/09/06/rapport-pakketadvies-in-de-praktijk-wikken-en-wegen-voor-een-rechtvaardig-pakket
14. National Institute for Health and Care Excellence (NICE). Guide to the processes of technology appraisal [Internet]. 2018 [cited 2019 May 13]. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf
15. National Institute for Health and Care Excellence (NICE). Assessing resource impact process manual: technology appraisals and highly specialised technologies [Internet]. 2017 [cited 2019 Jun 25]. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/Into-practice/assessing-resource-impact-process-manual-ta-hst.pdf
16. Zorginstituut Nederland. Beoordeling dure specialistische geneesmiddelen [Internet]. 2016 [cited 2018 Oct 4]. Available from: https://www.zorginstituutnederland.nl/

over-ons/werkwijzen-en-procedures/adviseren-over-en-verduidelijken-van-het-basispakket-aan-zorg/beoordeling-van-geneesmiddelen/beoordeling-dure-specialistische-geneesmiddelen

17. Zorginstituut Nederland. Beoordeling van nieuwe zorg [Internet]. 2016 [cited 2019 Jun 25]. Available from: https://www.zorginstituutnederland.nl/over-ons/werkwijzen-en-procedures/adviseren-over-en-verduidelijken-van-het-basispakket-aan-zorg/beoordeling-van-nieuwe-zorg
18. Zorginstituut Nederland. Pakketbeheer specialistische geneesmiddelen - Rapport [Internet]. 2013 [cited 2019 Jun 25]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2013/12/03/pakketbeheer-specialistische-geneesmiddelen
19. Zorginstituut Nederland. Procedure beoordeling extramurale geneesmiddelen - Rapport [Internet]. 2018 [cited 2019 Jun 25]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2016/09/09/procedure-beoordeling-extramurale-geneesmiddelen
20. Zorginstituut Nederland. Grote budgetimpact effectief longkankermiddel Durvalumab - Nieuwsbericht [Internet]. 2019 [cited 2019 May 15]. Available from: https://www.zorginstituutnederland.nl/actueel/nieuws/2019/04/03/grote-budgetimpact-effectief-longkankermiddel-durvalumab
21. Zorginstituut Nederland. GVS-advies sofosbuvir (Sovaldi®) bij chronische hepatitis C - Advies [Internet]. 2014 [cited 2018 May 3]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2014/05/20/gvs-advies-sofosbuvir-sovaldi-bij-chronische-hepatitis-c
22. Zorginstituut Nederland. GVS-advies sofosbuvir/ledipasvir (Harvoni®) bij chronische hepatitis C - Advies [Internet]. 2015 [cited 2018 Mar 26]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2015/05/18/gvs-advies-sofosbuvir-ledipasvir-harvoni-bij-chronische-hepatitis-c
23. Ministerie van Algemene Zaken. Kamerbrief over vergoeding geneesmiddelen pertuzumab en pembrolizumab - Kamerstuk - Rijksoverheid.nl [Internet]. 2017 [cited 2019 Jun 25]. Available from: https://www.rijksoverheid.nl/documenten/kamerstukken/2017/06/09/kamerbrief-over-vergoeding-geneesmiddelen-pertuzumab-en-pembrolizumab
24. Zorginstituut Nederland. GVS-advies alirocumab (Praluent®) bij primaire hypercholesterolemie - Advies [Internet]. 2016 [cited 2019 Jun 25]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2016/03/30/gvs-advies-alirocumab-praluent-bij-primaire-hypercholesterolemie
25. Zorginstituut Nederland. GVS-advies evolocumab (Repatha®) bij hypercholesterolemie en gemengde dyslipidemiehomozygote familiaire hypercholesterolemie - Advies [Internet]. 2015 [cited 2019 Jun 25]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2015/12/14/gvs-advies-evolocumab-repatha-bij-hypercholesterolemie-en-gemengde-dyslipidemiehomozygote-familiaire-hypercholesterolemie
26. Zorginstituut Nederland. Sacubitril/valsartan (Entresto) bij symptomatisch chronisch hartfalen met verminderde ejectiefractie - Rapport [Internet]. 2016 [cited 2019 Jun 25]. Available from: https://www.zorginstituutnederland.nl/publicaties/rapport/2016/04/05/sacubitril-valsartan-entresto-bij-symptomatisch-chronisch-hartfalen-met-verminderde-ejectiefractie
27. Zorginstituut Nederland. Sluis voor dure geneesmiddelen [Internet]. 2018 [cited 2019 Jun 25]. Available from: https://www.zorginstituutnederland.nl/over-ons/programmas-en-samenwerkingsverbanden/horizonscan-geneesmiddelen/sluis-voor-dure-geneesmiddelen
28. Zorginstituut Nederland. Richtlijnen voor economische evaluatie [Internet]. 2016 [cited 2018 Feb 8]. Available from: https://www.zorginstituutnederland.nl/over-ons/werkwijzen-en-procedures/adviseren-over-en-verduidelijken-van-het-basispakket-

aan-zorg/beoordeling-van-geneesmiddelen/richtlijnen-voor-economische-evaluatie

29. Zorginstituut Nederland. Pakketadvies nivolumab (Opdivo®) bij gevorderde (plaveiselcel) niet-kleincellige longkanker - Advies [Internet]. 2015 [cited 2019 Jun 19]. Available from: https://www.zorginstituutnederland.nl/publicaties/adviezen/2015/12/08/pakketadvies-nivolumab-opdivo

30. Ryen L, Svensson M. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. Health Econ. 2015;24:1289–301.

31. Shiroiwa T, Sung Y-K, Fukuda T, Lang H-C, Bae S-C, Tsutani K. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ. 2010;19:422–37.

32. 32. 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;24:1–7.

33. Eichler H-G, Baird LG, Barker R, Bloechl-Daum B, Børlum-Kristensen F, Brown J, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. Clin Pharmacol Ther. 2015;97:234–46.

34. Leyens L, Brand A. Early Patient Access to Medicines: Health Technology Assessment Bodies Need to Catch Up with New Marketing Authorization Methods. Public Health Genomics. 2016;19:187–91.

35. Tornqvist L, Vartia P, Vartia YO. How Should Relative Changes Be Measured? Am Stat. 1985;39:43–6.

36. IJzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. PharmacoEconomics. 2017;35:727–40.

37. Ijzerman MJ, Steuten LMG. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. Appl Health Econ Health Policy. 2011;9:331–47.

38. Commission E. Commission Regulation (EC) No 507/2006 of 29 March 2006 on the conditional marketing authorisation for medicinal products for human use falling within the scope of Regulation (EC) No 726/2004 of the European Parliament and of the Council. Off J Eur Union. 2006;50:6–9.

39. Boon W, Martins L, Koopmanschap M. Governance of conditional reimbursement practices in the Netherlands. Health Policy. 2015;119:180–5.

40. Franken M, Polain M le, Cleemput I, Koopmanschap M. SIMILARITIES AND DIFFERENCES BETWEEN FIVE EUROPEAN DRUG REIMBURSEMENT SYSTEMS. Int J Technol Assess Health Care. 2012;28:349–57.

41. Makady A, van Veelen A, de Boer A, Hillege H, Klungel OH, Goettsch W. Implementing managed entry agreements in practice: The Dutch reality check. Health Policy Amst Neth. 2019;123:267–74.

42. Vreman RA, Bouvy JC, Bloem LT, Hövels AM, Mantel-Teeuwisse AK, Leufkens HGM, et al. Weighing of Evidence by Health Technology Assessment Bodies: Retrospective Study of Reimbursement Recommendations for Conditionally Approved Drugs. Clin Pharmacol Ther. 2019;105:684–91.

43. Hoekman J, Klamer T T, Mantel-Teeuwisse AK, Leufkens HGM, Bruin MLD. Characteristics and follow-up of postmarketing studies of conditionally authorized medicines in the EU. Br J Clin Pharmacol. 2016;82:213–26.

44. Pease AM, Krumholz HM, Downing NS, Aminawung JA, Shah ND, Ross JS. Postapproval studies of drugs initially approved by the FDA on the basis of limited evidence: systematic review. Bmj-Br Med J. 2017;357:j1680.

45. Banzi R, Gerardi C, Bertele V, Garattini S. Approvals of drugs with uncertain benefit-risk profiles in Europe. Eur J Intern Med. 2015;26:572–84.

46. Van de Wetering EJ, van Exel J, Brouwer WBF. The Challenge of Conditional Reimbursement: Stopping Reimbursement

Can Be More Difficult Than Not Starting in the First Place! Value Health. 2017;20:118–25.

47. Dixit AK, Pindyck RS. Investment Under Uncertainty. Princeton University Press; 1994.

48. Brealey R, Myers S, Marcus A. Fundamentals of Corporate Finance. McGraw-Hill Education; 2014.

49. Julio B, Yook Y. Political Uncertainty and Corporate Investment Cycles. J Finance. 2012;67:45–83.

50. Block S. Are "Real Options" Actually Used in the Real World? Eng Econ. 2007;52:255–67.

51. Mohseninejad L. Uncertainty in economic evaluations: implications for healthcare decisions [PhD Thesis]. University of Groningen; 2013.

52. Grutters JPC, Abrams KR, de Ruysscher D, Pijls-Johannesma M, Peters HJM, Beutner E, et al. When to wait for more evidence? Real options analysis in proton therapy. The Oncologist. 2011;16:1752–61.

53. Attema AE, Lugnér AK, Feenstra TL. Investment in antiviral drugs: a real options approach. Health Econ. 2010;19:1240–54.

54. Favato G, Baio G, Capone A, Marcellusi A, Saverio Mennini F. A Novel Method to Value Real Options in Health Care: The Case of a Multicohort Human Papillomavirus Vaccination Strategy. Clin Ther. 2013;35:904–14.

55. Pertile P, Torri E, Flor L, Tardivo S. The timing of adoption of positron emission tomography: a real options approach. Health Care Manag Sci. 2009;12:217–27.

56. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. J Health Econ. 2000;19:755–66.

57. Klemp M, Frønsdal KB, Facey K. What principles should govern the use of managed entry agreements? Int J Technol Assess Health Care. 2011;27:77–83.

58. Pauwels K, Huys I, Vogler S, Casteels M, Simoens S. Managed Entry Agreements for Oncology Drugs: Lessons from the European Experience to Inform the Future. Front Pharmacol [Internet]. 2017 [cited 2019 Apr 4];8. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2017.00171/full

7

Summary & samenvatting

7.1

Summary

In chapter 1, the scene is set for the main goal of this thesis: To develop an integrated approach of cost-effectiveness, affordability and the associated uncertainty, where uncertainty also pertains to the temporal aspects of evidence and uncertainty.

In **chapter 2**, the role of uncertainty in early HTA is assessed. Furthermore, it aims to investigate whether current assessments are suitable for various early HTA cases. In **chapter 2.1**, this is illustrated with a pharmacogenomic case study. This specific case regarded a potential single nucleotide polymorphism (SNP) that could be used to identify angiotensin converting enzyme inhibitor (ACEi) users who are at high risk for developing ACEi induced angioedema. This adverse drug reaction (ADR) is severe, very frightening and potentially lethal. As ACEi use is highly prevalent, preventing this ADR could provide significant benefit to society. A threshold analysis was conducted to characterise the required specifications of a potential diagnostic based on this SNP. The threshold is achieving cost-effectiveness and the specifications investigated were the price, specificity and sensitivity of the potential diagnostic. The decision-tree model shows that, when testing all or only high baseline risk patients and assuming 100% sensitivity and specificity, the price of this novel technology should be low (< €1.95 and < €7.55, respectively). When accounting for lower than perfect sensitivity and specificity, the required price is even lower and potentially less than €0. Clearly, such a low price is impossible given current prices of diagnostical procedures, potentially foregoing the clinical benefit of this technology to society. The advent of whole genome sequencing (WGS) and resulting reduction in WGS prices could however allow for conductance of WGS for each individual. If this were to happen, all current and future pharmacogenomic markers could deliver value for free (minus the initial WGS investment). The current paradigm of stand-alone cost-effectiveness assessments of these novel technologies does not do justice to their potential value to society.

In **chapter 2.2**, an early HTA study on acalabrutinib is described. This medicine was still in clinical development at the time the study was performed. The indication of interest was relapsed chronic lymphocytic leukaemia (CLL) and the goal was to assist early reimbursement decision making, partly by assessing scenarios to find the impact of critical parameters on cost-effectiveness. A partitioned survival model was constructed for comparing acalabrutinib with ibrutinib and used a UK national health service perspective. Progression-free survival (PFS), post-progression survival (PPS) and death were the selected model states. PFS and overall survival (OS) were parametrically extrapolated from ibrutinib publications. To model acalabrutinib efficacy, a preliminary hazard ratio based on phase I/II data was applied. Deterministic sensitivity analysis (DSA) as well as probabilistic sensitivity analyses (PSA) were performed. Furthermore, 1296 scenarios (based on combinations of various parameter values) were assessed. The base case ICER is £58,899 / QALY, with 3.44 incremental QALYs and incremental costs of £202,861. DSA indicates that survival estimates, utilities and treatment costs of ibrutinib and acalabrutinib and resource use during PFS had the greatest influence on the ICER. PSA indicates that greater efficacy of acalabrutinib would decrease the likelihood of cost-effectiveness (from 69% at no effect to 2% at maximum efficacy). Scenario analyses shows

that a reduction in PFS does not lead to great QALY differences although it does greatly impact costs. For OS, the opposite is true. Acalabrutinib is therefore not likely to be cost-effective compared to ibrutinib under current development scenarios. The conflicting effects of OS, PFS, drug costs and utility during PFS shows that determining cost-effectiveness of acalabrutinib without insight into all parameters complicates decision-making.

Chapter 3 is aimed at investigating the uncertainty of budget impact (BI) estimates by quantifying the accuracy of BI estimates and by identifying determinants for (in)accuracy. In **chapter 3.1**, this is addressed by studying the BI estimates of novel hepatitis C drugs in the Netherlands. This case of hepatitis C is especially relevant as access restrictions were imposed based on high BI estimates. Hepatitis C direct-acting antivirals (DAAs) that were introduced in the Netherlands between January 2014 and March 2018 were therefore selected. Of these products, the BI estimates as presented in the Dutch National Health Care Institute (ZIN) reimbursement dossiers were compared to the observed BI. Total observed BI in that period amounted to €248 million whilst BI estimates ranged from €388 - €510 million. The foreseeable introductions of new products were inadequately incorporated in BI estimations and timing of specific regulatory decisions were inadequately incorporated in estimates which both contributed to BI over-estimation. Furthermore, uncertainty regarding the patient population size and the impact of the final reimbursement decision limited BI estimation accuracy. To conclude, the findings show that BI for this novel drug class was largely overestimated.

The accuracy and timing of BI estimates in oncology drugs is assessed in **chapter 3.2**. Given the potentially life-saving nature of oncology products, access decisions that are wrongly informed by inaccurate BI estimations could cause great harm. We selected oncology products that were granted with European Medicines Agency (EMA) Marketing Authorisation (MA) between 1-Jan-2000 and 1-Oct-2017 and which were designated as a 'New Active Substance' by the EMA. Products were consequently included if a BI estimation was present in a Dutch Health Care Institute (ZIN) reimbursement dossier. These BI estimates were compared with the observed BI in the third year after the publication of the respective reimbursement dossier as BI estimates in the dossiers were aimed to project BI in this third year. Products where the date of publication of the reimbursement dossier deviated from the date of the first BI record by a maximum of 6 months were included in the base-case analysis. The resulting 10 products resulted in BI estimation accuracy of 0.64, where accuracy is defined as observed BI / estimated BI. Accuracy differed dramatically between these products, ranging from 0.14 to 1.08. For these 10 products, a total sum of €141 million BI was estimated whilst only €82 million was observed. Chapter 3.2 therefore shows that BI estimates for oncology products in the Netherlands were in general over-estimated and were associated with considerable inaccuracy. Chapter 3 thus shows that BI estimates in two different settings were quite inaccurate and that using these estimates for informing reimbursement decisions should therefore be carefully considered.

Chapter 4 describes new methodology for managing uncertainty in both affordability and cost-effectiveness and their roles in decision-making. **Chapter 4.1** addresses the deficiencies in BI estimation that were presented in **chapter 3** by providing new methodology for BI estimation and by describing the results of the validation of this method, thereby using oncology products as a case study. Like in **chapter 3.2**, we included oncology products which received EMA MA between 1-Jan-2000 and 1-Oct-2017 and which were designated as 'New Active Substance' by the EMA. For these products, characteristics such as orphan, first-in-class or conditional approval designation were collected, as well as a classification regarding the target tumour site. Furthermore, the monthly observed BI data was collected where observed BI was composed of a product's list price multiplied by monthly volume. This dataset was split in a training set and validation set based on a whether a product's first BI record occurred before or after 1-May-2012, respectively.

Using the training set, a mixed-effects prediction model was constructed which was consequently cross-validated using a rolling forecasting origin. This approach mimics the monthly addition of new data by means of new observed BI but also the addition of newly products to the model. The model was constructed to predict the first 45 months of BI of each product. Error, used as validation outcome, was defined as $e^{|ln(Observed\ BI/Predicted\ BI)|}$. The mean and median errors were 2.94 and 1.57, respectively. Errors were higher with fewer months of observed BI data for a specific product and for more future predictions. Based on this validation, it was concluded that the developed model is valid for predicting BI.

Chapter 4.2 targets uncertainty in cost-effectiveness by setting out a new approach for displaying the results of Probabilistic Sensitivity Analysis (PSA) with the aim of providing more information compared to the traditional scatterplot. This scatterplot visualisation has two major issues: i) overlap of individual estimates in high density areas and, partially extending on this issue ii) the general difficulty of estimating relative density from a scatterplot. To overcome these issues, the Relative Density Plot (PSA-ReD) was developed. In **chapter 4.2**, this development as well as a demonstration using three case studies is presented. The PSA-ReD combines a density plot and a contour plot to display the PSA results and created using R and the corresponding R was made available to other scholars by means of GitHub and an elaborative manual. To construct a PSA-ReD, density is calculated using two-dimensional kernel density estimation, transformed to cumulative probability, which is depicted using a colour gradient. Contours are then plotted over regions with a predefined cumulative probability. The case studies showed that PSA-ReD provides additional visual information such as a very dense area in one case study that was not visible in the original scatterplot. Such information could be used for additional model validation and potentially better-informed reimbursement decisions.

Chapter 5 presents new concepts for managing uncertainty in decision-making. The first concept integrates affordability, cost-effectiveness and uncertainty and the second concept integrates the aspect of time and the timeliness of uncertainty into the first concept. This first

concept is discussed in **chapter 5.1**. The background of this study was the separate appraisal of cost-effectiveness and affordability and the resulting separate assessment of uncertainty of these outcomes, which incorrectly weighs the potential risks and benefits of new interventions. This is especially an issue if, as evidence has shown, Willingness to Pay (WTP) is a determinant of opportunity costs (being BI). **Chapter 5.1** therefore presents a conceptual framework for united appraisal of BI, WTP and ICER and their associated uncertainty where WTP is dynamic and influenced by BI. The lung cancer drug Opdivo (nivolumab) was selected as case study and three different methods were used to quantify the relationship between WTP and BI. BI, ICER and WTP were integrated using population Net Monetary Benefit (pNMB), an outcome derived from Net Monetary Benefit. When WTP was not influenced by BI, only ICER influences whether pNMB is positive or negative and therefore influences the investment decision. These results were also produced by one of the dynamic WTP methods as this method yielded a WTP that was very insensitive to BI. When a stronger relationship is present, as was the case for the relationship that was derived from a real-world reimbursement decision, BI, ICER and their uncertainty have a synergistic influence on pNMB. This new concept allows for truly integrated appraisal of cost-effectiveness, affordability and uncertainty.

Chapter 5.2 extends this concept by including the aspect of time in decision-making and uncertainty: Using Real Options Analysis (ROA), the option of postponing the decision is added as a potential outcome of a decision. ROA is incorporated using aspects from two previous chapters. First, the pNMB framework described in **chapter 5.1** is used to integrate cost-effectiveness (as ICER), affordability (as BI) and uncertainty, thereby also adopting the dynamic WTP approach described in this chapter. Second, the BI prediction model described in **chapter 4.1** is used to incorporate the temporality of evidence and uncertainty. The ROA implementation simulates the monthly arrival of observed BI data, which lead to more accurate BI prediction (as **chapter 4.1** has shown). ROA then values the future reduction in uncertainty against the benefits of immediate reimbursement. It thus compares postponing (for more certainty, but losing some potential benefit) vs immediate access (thereby accepting more risk but benefitting from the intervention). The results indicate that, for the Opdivo (nivolumab) case study, waiting until month 10 before issuing reimbursement was the optimal balance between access and risk. The presented technique is however inflicted with the limitations of **chapters 4.1 and 5.1**, necessitating future research before this technique can be implemented. Still, ROA is a proven technique and could be a suitable methodological tool for providing early guidance on flexible and adaptive reimbursement decisions, deemed essential in the current landscape of ever higher uncertainty at market access of new costly drugs.

In **chapter 6**, the presented findings are discussed including the potential implications, limitations of the conducted studies and potential hurdles for implementation. The work presented herein has culminated in the potential solution to the main objective raised in the introduction of this thesis, as the proposed pNMB and ROA approach can be used to perform a unified appraisal of affordability, cost-effectiveness and the associated uncertainty and timeliness of these aspects.

7.2

Samenvatting

Hoofdstuk 1 introduceert het doel van dit proefschrift: Het ontwikkelen van een geïntegreerde benadering van kosteneffectiviteit, betaalbaarheid en de daarmee geassocieerde onzekerheid. Onzekerheid omvat hierbij ook de invloed van tijd op (de ontwikkeling van) wetenschappelijk bewijs en de bijbehorende onzekerheid.

In **hoofdstuk 2** is de rol van *Health Technology Assessment* (HTA) bij besluitvorming over geneesmiddelen uiteengezet. Daarnaast is in dit hoofdstuk onderzocht of de huidige beoordelingsmethodiek geschikt is voor verschillende typen van vroege HTA. **Hoofdstuk 2.1** illustreert dit aan de hand van een farmacogenetische casus. Deze casus behelst een potentiele *single nucleotide polymorphism* (SNP). Deze SNP kan worden ingezet om patiënten te identificeren die een hoog risico hebben op het ontwikkelen angio-oedeem bij gebruik van angiotensin converting enzyme inhibitors (ACE-remmers). Deze bijwerking kan zelfs tot mortaliteit leiden is daarom ernstig te noemen. Aangezien ACE-remmer gebruik zeer prevalent is kan het voorkómen van deze bijwerking een grote gezondheidswinst opleveren voor de maatschappij. Voor deze mogelijke screeningsmethode is een grenswaarde-analyse uitgevoerd om de voor kosteneffectiviteit vereiste karakteristieken te bepalen. De prijs, sensitiviteit en specificiteit van de screeningsmethode zijn de karakteristieken die zijn onderzocht. De resultaten zijn door middel van een beslisboom gegenereerd. Deze resultaten tonen aan dat wanneer alle ACE-remmer gebruikers worden getest, waarbij uit wordt gegaan van 100% sensitiviteit en specificiteit, de prijs van dit nieuwe diagnosticum zeer laag (< €1.95) moet zijn om als kosteneffectief te worden beschouwd. Wanneer testen wordt beperkt tot enkel patiënten met een verhoogd risico op de betreffende bijwerking is de maximale prijs €7.55. Wanneer uit wordt gegaan van lagere sensitiviteit en specificiteit is de vereiste prijs nog lager en in veel gevallen zelfs lager dan €0. Het is evident dat dermate lage prijzen onmogelijk zijn gezien de huidige prijzen voor diagnostiek. Hierdoor kan de potentiele gezondheidswinst van deze screeningsmethode verloren gaan. De recente ontwikkelingen van *whole genome sequencing* (WGS) en de recente prijsverlagingen van deze procedure kunnen het mogelijk maken om voor ieder individu WGS uit te voeren. Indien dit het geval zou zijn zouden alle huidige maar ook toekomstige farmacogenetische tests gratis (minus de initiële investering in WGS) waarde kunnen opleveren. Het huidige paradigma, waarbij losstaande kosteneffectiviteitsanalyses worden uitgevoerd op deze nieuwe technieken, lijkt in ieder geval geen recht te doen aan hun potentiele maatschappelijke waarde.

In **hoofdstuk 2.2** is een vroege kostenutiliteitsanalyse van acalabrutinib beschreven. Dit middel, dat nog in ontwikkeling was ten tijde van het uitvoeren van deze studie, is beoordeeld in het kader van gebruik voor de behandeling van chronisch lymfatische leukemie (CLL). Het doel van deze analyse was bepalen of- en in hoeverre vroege vergoedingsbeslissingen kunnen worden geïnformeerd, mede door het in kaart brengen van parameters die kritisch zijn voor de kosteneffectiviteit van dit geneesmiddel. Acalabrutinib werd vergeleken met ibrutinib waarbij gebruik werd gemaakt van een parametrisch overlevingsmodel en het perspectief van de Britse Nationale Gezondheidsdienst (NHS). Het model bestond uit drie stadia, namelijk progressievrije overleving (PFS), overleving na progressie (PPS) en algehele

overleving (OS). Voor het modelleren van de effectiviteit van acalabrutinib werd gebruik gemaakt van een voorlopige *hazard ratio* (HR) die was gebaseerd op fase I/II data. Zowel een deterministische sensitiviteitsanalyse (DSA) als een probabilistische sensitiviteitsanalyse (PSA) werden uitgevoerd. Verder werden 1296 scenario's, gebaseerd op verschillende combinaties van parameterwaarden, geanalyseerd. De *base case* incrementele kosteneffectiviteitsratio (IKER) werd berekend op £58,899 / *Quality Adjusted Life Year* (QALY), bestaande uit 3.44 incrementele QALY's en £202,861 aan incrementele kosten. De DSA laat zien dat parameters voor overleving, utiliteiten en kosten van behandeling met ibrutinib en acalabrutinib de grootste invloed hadden op de IKER. Uit de PSA blijkt dat een hogere effectiviteit van acalabrutinib de kans op kosteneffectiviteit verkleint (van 69% bij gelijke effectiviteit tot 2% bij maximale effectiviteit). De scenarioanalyses tonen aan dat een vermindering in PFS niet leidt tot grote verschillen in QALY's terwijl dit wel een grote impact heeft op kosten. Voor OS is het tegendeel waar. Gezien deze bevindingen is het onwaarschijnlijk dat acalabrutinib, vergeleken met ibritunib, in dit stadium kosteneffectief is. De tegengestelde invloeden van OS, PFS en geneesmiddelkosten en utiliteiten gedurende PFS op de IKER maken het vaststellen van kosteneffectiviteit moeilijk wanneer er onvolledig inzicht is in alle parameters.

Hoofdstuk 3 is gericht op het onderzoeken van de onzekerheid van *budget impact* (BI) schattingen door het kwantificeren van de nauwkeurigheid van deze schattingen en door het identificeren van determinanten voor de precisie van BI-schattingen. **Hoofdstuk 3.1** behandelt dit door het bestuderen van Nederlandse BI-schattingen van een nieuwe generatie geneesmiddelen voor hepatitis C. Deze hepatitis C casus is vooral interessant gezien de ingestelde vergoedingsrestricties voor deze geneesmiddelen als gevolg van hoge BI-schattingen. Hepatitis C *direct-acting antivirals* (DAA's) die in Nederland zijn geïntroduceerd tussen januari 2014 en maart 2018 zijn geselecteerd voor deze studie. Van deze geneesmiddelen zijn de door het Zorginstituut Nederland (ZIN) gepubliceerde BI-schattingen vergeleken met de daadwerkelijk waargenomen BI. De totaal waargenomen BI in die periode was €248 miljoen terwijl in totaal €388 - €510 miljoen aan BI was voorspeld. De nieuwe introducties van DAA's, die voorzien hadden kunnen worden, werden niet of onvolledig meegenomen in BI-schattingen. Daarnaast werd de timing van regulatoire beslissingen niet juist ingebed in deze schattingen, welke samen met de hiervoor genoemde reden hebben bijgedragen aan de overschatting van BI. Verder zijn ook onzekerheid over de grootte van de hepatitis C populatie en de impact van de uiteindelijke vergoedingsbeslissing waarschijnlijk debet geweest aan de beperkte precisie van de BI-schattingen. De bevindingen in dit hoofdstuk tonen aan dat de BI van deze nieuwe geneesmiddelklasse grotendeels is overschat.

De precisie en timing van BI-schattingen voor oncologie geneesmiddelen is geëvalueerd in **hoofdstuk 3.2**. Omdat deze geneesmiddelen van levensbelang kunnen zijn voor patiënten kunnen vergoedingsbeslissingen die gebaseerd zijn op niet precieze BI-schattingen grote schade veroorzaken. Voor deze studie zijn oncologiemiddelen geselecteerd die tussen 1 januari

2000 en 1 oktober 2017 van het Europees Geneesmiddelen Agentschap (EMA) marktoelating verkregen en daarbij werden aangemerkt als een '*New Active Substance*'. De geneesmiddelen werden geïncludeerd als er voor een product een BI-schatting beschikbaar was in het door ZIN gepubliceerde vergoedingsdossier. Deze BI-schattingen zijn vervolgens vergeleken met daadwerkelijk waargenomen BI in het derde jaar na publicatie van het vergoedingsdossier. Deze termijn is gekozen omdat BI-schattingen doorgaans voor het derde jaar introductie worden opgesteld. Enkel de producten waarbij de datum van publicatie van het vergoedingsdossier maximaal 6 maanden afweek van de eerste maand waarin BI werd waargenomen zijn geïncludeerd in de *base case* analyse. Dit resulteerde in de inclusie van 10 producten waarbij de gemiddelde BI precisie (gedefinieerd als waargenomen BI / geschatte BI) werd berekend op 0.64. De precisie van de BI-schatting voor individuele geneesmiddelen verschilde sterk (van 0.14 tot 1.08). Voor de 10 *base case* producten werd een gezamenlijke BI van €141 miljoen geschat terwijl slechts €82 miljoen werd waargenomen. Hoofdstuk 3.2 toont daarmee aan dat BI voor deze oncologiemiddelen in Nederland in het algemeen wordt overschat en dat BI-schattingen aanzienlijke onzekerheid met zich meebrengen. Hoofdstuk 3 laat dus zien dat BI-schattingen in twee verschillende indicatiegebieden relatief weinig precies waren en dat het gebruik van deze schattingen voor het informeren van vergoedingsbeslissingen daarom zorgvuldig moet worden overwogen.

Hoofdstuk 4 beschrijft nieuwe methodologie voor het omgaan met onzekerheid rondom betaalbaarheid en kosteneffectiviteitsuitkomsten en de rol daarvan in de besluitvorming. **Hoofdstuk 4.1** gaat in op de tekortkomingen in BI-schattingen die zijn gepresenteerd in **hoofdstuk 3,** door nieuwe methodologie voor het schatten van BI te beschrijven. In **hoofdstuk 4.1** wordt de ontwikkeling van deze nieuwe methode beschreven en wordt deze methode gevalideerd aan de hand van oncologieproducten. Net als in **hoofdstuk 3.2** zijn oncologiemiddelen geselecteerd die tussen 1 januari 2000 en 1 oktober 2017 door de EMA zijn toegelaten tot de Europese markt en die door de EMA als 'New Active Substance' zijn aangemerkt. Van deze producten zijn verschillende kenmerken verzameld (bijvoorbeeld mogelijke weesgeneesmiddelstatus) evenals een classificatie met betrekking tot de fysieke locatie van de tumor. Vervolgens zijn de maandelijks waargenomen BI-gegevens verzameld, waarbij de waargenomen BI bestond uit de officiële lijstprijs van een geneesmiddel vermenigvuldigd met het maandelijkse volume. Deze dataset is opgesplitst in een trainingsset en validatieset op basis van de classificatie van de eerste waargenomen BI van een product respectievelijk vóór of na 1 mei 2012 plaatsvond.

Op basis van de trainingsset is een *mixed effects* model geconstrueerd waarmee de BI voorspeld kan worden. Vervolgens is dit model gevalideerd aan de hand van de validatieset waarbij gebruik is gemaakt van een *rolling forecasting origin*. Deze techniek bootst de maandelijkse toevoeging van nieuwe gegevens na door middel van het toevoegen van nieuw waargenomen BI, maar ook door de toevoeging van nieuwe producten aan het model. Het model is gebouwd om de BI in de eerste 45 maanden van elk product te voorspellen. Voorspellingsafwijking, gebruikt als

validatieresultaat, is gedefinieerd als *e^|ln(waargenomen BI / voorspelde BI)|*. De gemiddelde en mediane voorspellingsafwijking waren respectievelijk 2.94 en 1.57. Afwijkingen waren hoger wanneer er voor geneesmiddelen minder BI-data beschikbaar was in het model en voor voorspellingen verder in de toekomst. Op basis van deze validatie is geconcludeerd dat het ontwikkelde model valide is voor het voorspellen van BI.

Hoofdstuk 4.2 richt zich op onzekerheid in kosteneffectiviteit door een nieuwe methode voor het weergeven van de resultaten van Probabilistic Sensitivity Analysis (PSA) te presenteren. Het doel hiervan is om meer informatie over te brengen dan mogelijk is met de traditionele *scatter plot*. Deze *scatter plot* heeft twee belangrijke gebreken: i) De overlap van individuele schattingen in gebieden met een hoge dichtheid en, gedeeltelijk voortbordurend op deze kwestie, ii) de moeilijkheid om (verschillen in) de relatieve dichtheid in een *scatter plot* juist te interpreteren. Om deze beperkingen op te lossen, is de *Relative Density Plot* (PSA-ReD) figuur ontwikkeld. In **hoofdstuk 4.2** wordt deze ontwikkeling gepresenteerd, evenals een demonstratie met behulp van drie casussen. De PSA-ReD combineert een *density plot* en een *contour plot* om de PSA-resultaten weer te geven en wordt gegenereerd door middel van R. De methode en het bijbehorende R-script is publiekelijk beschikbaar gesteld middel van GitHub en een uitgebreide handleiding. Om een PSA-ReD figuur te construeren, wordt de dichtheid berekend met behulp van tweedimensionale *kernel density estimation*, omgezet in cumulatieve waarschijnlijkheid, die wordt afgebeeld met een kleurgradiënt. Contouren worden vervolgens weergegeven over gebieden met een vooraf gedefinieerde cumulatieve waarschijnlijkheid. De gebruikte casussen toonden aan dat PSA-ReD figuren aanvullende visuele informatie bieden, zoals het bestaan van een klein gebied met zeer hoge relatieve dichtheid in één casus dat niet zichtbaar was in de oorspronkelijke *scatter plot*. Dergelijke informatie kan worden gebruikt voor aanvullende of verbeterde modelvalidatie en mogelijk beter geïnformeerde vergoedingsbeslissingen.

Hoofdstuk 5 presenteert nieuwe concepten voor het omgaan met onzekerheid in besluitvorming omtrent geneesmiddelvergoeding. Het eerste concept integreert betaalbaarheid, kosteneffectiviteit en onzekerheid en het tweede concept integreert het aspect van tijd en de relatie tussen tijd en onzekerheid in het eerste concept. Dit eerste concept wordt besproken in **hoofdstuk 5.1**. De achtergrond van deze studie is de afzonderlijke beoordeling van kosteneffectiviteit en betaalbaarheid en de resulterende afzonderlijke beoordeling van de onzekerheid van deze twee uitkomsten. Hierdoor worden de potentiële risico's en voordelen van nieuwe geneesmiddelen onvolledig en onjuist beoordeeld. Dit is met name een probleem als, zoals uit meerdere studies is gebleken, *Willingness to Pay* (WTP) mede bepaald zou moeten worden door opportuniteitskosten (ofwel BI). Hoofdstuk 5.1 presenteert daarom een conceptueel raamwerk voor een gezamenlijke beoordeling van BI, WTP en IKER en de bijbehorende onzekerheid waarbij WTP dynamisch is en wordt beïnvloed door BI. Het longkanker geneesmiddel Opdivo (nivolumab) is in deze studie geselecteerd als casus en drie verschillende methoden zijn gebruikt om de relatie tussen WTP en BI te kwantificeren. BI, IKER en WTP zijn geïntegreerd met behulp van *population Net Monetary Benefit* (pNMB), een resultaat afgeleid van *Net Monetary Benefit*. Wanneer WTP

niet wordt beïnvloed door BI, beïnvloedt alleen de IKER of pNMB positief of negatief is en beïnvloedt daarmee eenzijdig de vergoedingsbeslissing. Dit effect, waarbij BI geen rol heeft op de vergoedingsbeslissing, was ook aanwezig bij gebruik van één van de dynamische WTP-methoden omdat deze methode een WTP opleverde die zeer ongevoelig was voor BI. Wanneer een sterkere relatie tussen BI en WTP aanwezig is, zoals het geval was voor de relatie die was afgeleid van een vergoedingsbeslissing uit de praktijk, hebben BI, IKER en hun onzekerheid een synergetische invloed op pNMB. Dit nieuwe conceptuele raamwerk maakt een volledig geïntegreerde beoordeling van kosteneffectiviteit, betaalbaarheid en onzekerheid mogelijk.

Hoofdstuk 5.2 breidt dit concept uit met het aspect van tijd in besluitvorming en onzekerheid: met behulp van *Real Options Analysis* (ROA) wordt de optie om de beslissing uit te stellen toegevoegd als een mogelijke uitkomst van een beslissing. ROA is geïmplementeerd met behulp van aspecten uit twee voorgaande hoofdstukken. Ten eerste wordt het pNMB-raamwerk zoals beschreven in **hoofdstuk 5.1** gebruikt om kosteneffectiviteit (als IKER), betaalbaarheid (als BI) en onzekerheid te integreren, waarbij ook de dynamische WTP-aanpak wordt toegepast die in dit hoofdstuk wordt beschreven. Ten tweede wordt het BI-voorspellingsmodel uit **hoofdstuk 4.1** gebruikt om de tijdigheid van wetenschappelijk bewijs en onzekerheid te implementeren. Deze ROA-implementatie simuleert het maandelijks beschikbaar komen van nieuw waargenomen BI-gegevens, die leiden tot een accurate BI-voorspelling (zoals **hoofdstuk 4.1** heeft aangetoond). ROA waardeert vervolgens de toekomstige vermindering van de onzekerheid ten opzichte van de voor- of nadelen van onmiddellijke vergoeding. Het vergelijkt dus het uitstellen (dit zorgt voor meer zekerheid, maar heeft als gevolg het verlies van een gedeelte van de potentiële waarde) met directe toegang (waardoor meer risico wordt aanvaard maar er meteen geprofiteerd kan worden van een nieuw geneesmiddel). Voor de Opdivo casus laten de resultaten zien dat het optimale evenwicht tussen toegang en risico na 10 maanden wordt bereikt. Aangezien **hoofdstuk 5.2** is gebaseerd op **hoofdstukken 4.1 en 5.1** hebben ook de beperkingen van deze twee hoofdstukken betrekking op de ROA-methode. Verder onderzoek is dus nodig voordat ROA geïmplementeerd kan worden. ROA is echter een wetenschappelijk beproefde techniek waardoor het een geschikte en valide methode kan zijn om vroegtijdig advies te geven over flexibele en dynamische vergoedingsbeslissingen. Dit kan als essentieel worden beschouwd in het huidige landschap van steeds grotere onzekerheid bij het verlenen van toegang tot nieuwe dure geneesmiddelen.

In **hoofdstuk 6** worden de in dit proefschrift gepresenteerde bevindingen besproken, inclusief de mogelijke implicaties en beperkingen van de uitgevoerde onderzoeken alsmede mogelijke obstakels wat betreft implementatie. Verder wordt gesteld dat het gepresenteerde werk heeft geleid tot de mogelijke oplossing voor de probleemstelling die in de introductie van dit proefschrift naar voren is gebracht. In dit proefschrift is namelijk aangetoond dat de beschreven pNMB- en ROA-methodiek kan worden gebruikt om een geïntegreerde beoordeling van betaalbaarheid, kosteneffectiviteit mogelijk te maken, inclusief de bijbehorende onzekerheid en tijdigheid van deze aspecten.

8

Dankwoord

List of co-authors

List of publications

8.1

Dankwoord

Het is af, wat een heerlijk gevoel! Met dit boek sluit ik niet alleen mijn promotieonderzoek af maar komt er ook een einde aan een tijd waar ik ontzettend van heb genoten en geen moment spijt van heb gehad. Ik ben er vrij zeker van dat de mensen met wie ik deze vier jaar heb doorgebracht hierin minstens een net zo belangrijke rol hebben gespeeld als het daadwerkelijke onderzoek. Het is dan ook heerlijk om al die mensen in dit stuk, vrij van author guidelines, rondes commentaar en andere wetenschappelijke mores, te bedanken.

Alvorens hieraan te beginnen wil ik echter stilstaan bij een wat serieuzer aspect van mijn vier jaar in het David de Wiedgebouw. Pas als je terugkijkt besef je namelijk pas hoe snel de tijd gaat, hoe gelukkig je met bepaalde zaken moet zijn en de ontwikkelingen die je zelf maar ook anderen doormaken. Vooral ook vanuit het perspectief van mijn promotieteam is dit, zoals hieronder te lezen valt, een waardevolle invalshoek.

Beste Jan, helaas werd redelijk snel na de start van mijn onderzoek duidelijk dat jij dit voltooide werk waarschijnlijk nooit zou zien. Tóch bleef je ongelofelijk betrokken, enthousiast en altijd positief. Zelfs wanneer je niet naar het DDW kon komen gaf je via de mail feedback en wist je me te inspireren. Die inspiratie, die al begon bij jouw aanstekelijke colleges die ik als student volgde, heeft een erg belangrijke rol gespeeld in mijn keuze om bij jou en Anke een promotietraject te starten. Hiermee heeft jouw enthousiasme voor altijd een belangrijke impact op mijn leven.

Beste Anke, wat heb ik ongelofelijk geluk gehad met jou als copromotor en dagelijkse begeleider! Onze samenwerking begon in 2013 toen ik mijn onderzoeksproject bij jou, Geert en Rob uit ging voeren in het Meander. Het onderzoek zelf vond ik erg leuk en interessant maar de persoonlijke klik was doorslaggevend toen jij, ergens in 2015, vroeg of ik bij jou en Jan wilde komen promoveren. Hier heb ik zoals gezegd geen enkel moment spijt van gehad en mede door jouw persoonlijke en altijd optimistische aanpak heb ik zo van deze vier jaar kunnen genieten. Jij stond altijd achter mijn ideeën en plannen, steunde me in mijn enthousiasme en had altijd het beste met me voor. Tel daarbij de lunches, diners en barbecues die jij en Daan bij je thuis organiseerde en natuurlijk de gezelligheid op de borrels en congressen en je begrijpt waarom ik zo'n fan van je ben. Ook voor jou is er een hoop veranderd afgelopen jaren en volgens mij heb je het nu perfect voor elkaar! Anke, ontzettend bedankt voor alles! Op een mooie toekomst!

Beste Cornelis, ik ben heel blij dat jij in 2016 aansloot bij het team. De combinatie van jouw scherpe en kritische blik, je brede perspectief en je internationale ervaring was namelijk zeer waardevol. Verder heb ik jouw openheid en rechtdoorzee mentaliteit heel erg gewaardeerd. Ik wil je echter niet alleen bedanken voor je bijdragen aan mijn proefschrift maar ook voor je hulp en tips over de carrièremogelijkheden na een PhD.

Beste Olaf, toen ik in de zomer van 2015 begon was jij nog niet heel lang professor en ik herinner me jouw oratie dan ook nog goed. Destijds kon ik je wat lastig peilen en lag mijn onderwerp ook wat af van jouw expertisegebied. Sindsdien weet ik echter dat je 's nachts ook poule-les geeft,

na een borrel in een draaimolen springt en er zeker van houdt om met promovendi wat biertjes te drinken maar dat je daarnaast ook een ongelofelijk aardige, meedenkende en no-nonsense promotor bent. Ik heb met jou heel veel geluk gehad als promotor en je bent voor mij echt een voorbeeld geworden voor wat een goede en prettige leidinggevende maakt. Naast dit alles waren jouw scherpe inhoudelijke commentaren cruciaal, heel erg bedankt voor alles!

Ook de Leescommissie, bestaande uit prof. dr. Lieven Annemans, prof. dr. Ton de Boer, prof. dr. Bert Leufkens, prof. dr. Hans Severens en prof. dr Maarten Postma, wil ik bedanken voor de tijd en moeite die ze hebben genomen om mijn proefschrift te beoordelen.

Zonder de data van Farminform waren vele studies onmogelijk geweest. Ik wil Farminform, en in het bijzonder Léon Walenbergh en Martijn van Bloois, dan ook hartelijk danken voor hun tijd en voor het beschikbaar stellen van de budget impact data.

Beste Saskia Knies & Joost Enzing, ik heb oprecht genoten van onze samenwerking. Zonder jullie inzet en ideeën waren de laatste twee studies van dit proefschrift zeker niet zo sterk geworden, heel erg bedankt!

Toen ik mezelf vier jaar geleden met ietwat te lang haar op de vierde verdieping installeerde kwam ik terecht tussen vijf parels van collega's. Joris en Corinne, jullie waren toen jullie proefschriften aan het afronden en waren daarom iets minder enthousiast voor zaken als bureaustoel-roeien of een klein wijntje na de lunch. Bij Richelle en Mariette ging dit er echter in als zoete koek, hetgeen dan met zowel lichte scepsis als plezier werd aanschouwd door Hedy. Lieve Corinne, Joris, Richelle, Hedy en Mariette: dankzij jullie voelde ik me meteen thuis en ging ik meteen met plezier naar de Uithof. Wat zijn er daarnaast een hoop prachtige verhalen en hilarische momenten ontstaan. Gelukkig zien we elkaar nog regelmatig om die te koesteren, ik hoop van harte dat we dat voortzetten.

Richelle, eindelijk mag ik een stuk over jou schrijven. Wat hebben wij in die 3.5 jaar een mooie tijd gehad! We hebben zoveel gelachen om de meest zinloze dingen, geroddeld, geborreld en gefeest. Onze verhalen over de promotiefeestjes van anderen zijn natuurlijk legendarisch, net als de UU-kerstborrels en ICPE in Praag. Verder moet ik nog steeds hardop lachen over onze discussies over zeevruchten, middagdutjes en de ogenschijnlijke multifunctionaliteit van bureaustoelen. Ik ben je echter niet alleen dankbaar voor deze leuke en gezellige dingen maar ook voor jouw vriendschap want ik kon (en kan) altijd bij jou terecht voor serieuze zaken. Lieve Richelle, zonder jou had ik absoluut niet zo positief kunnen terugkijken op mijn PhD. Bedankt voor alles en ik hoop dat ik jou en Michiel nog heel vaak blijf zien!

Rick, eerlijk gezegd moest ik eerst best een beetje aan je wennen. Het was echter al heel snel gezellig en man, wat ben ik jou gaan waarderen als collega maar zeker ook als vriend. Niet alleen

hadden we zowel heel serieuze als erg zinloze gesprekken maar hebben maar we hebben vooral ook op wetenschappelijk gebied veel samengewerkt. Voor mij was het echt super waardevol om een soort extra en informele copromotor binnen handbereik te hebben. Daarnaast waren de vele congressen met jou erg gezellig, alleen jammer dat ik altijd brakker was dan jij. Rick, bedankt voor dit alles ik hoop dat we elkaar nog vaak blijven zien!

Lotte, jouw gezelligheid onze gesprekken ga ik echt missen. Bij jou kon ik altijd terecht om PhD-perikelen te bespreken maar natuurlijk ook voor een casual koffie of thee. Wat je hierna ook gaat doen, er is daar vast plek voor iemand zo lief, attent en optimistisch als jij!

Lourens en Joris; de grote beren van beneden. Lou, we kennen elkaar natuurlijk al vanuit allerlei verbanden maar ook aan de UU was het weer erg gezellig! Joris, ik ben erg blij dat jij ook praktisch fulltime in de parel van de Uithof te vinden was. Mooi dat er met jullie ook goed geborreld kan worden, wintersport en Philadelphia waren mede daardoor zeer geslaagd!

Renske en Geert, hartelijk dank voor jullie gezelligheid en de memorabele borrels die we samen hebben gehad. Lola, ISPOR en Hollandse avonden waren anders absoluut niet zo leuk geweest.

Rianne, wat ben jij een gezellige collega. Roddels, obscure YouTube filmpjes, Woerdense rituelen en natuurlijk je befaamde verhalen over de geneugten van het hebben van twee ovens; jij draait je hand er niet voor om.

Dear Armina, it's great that I can now congratulate you on finishing your PhD! Thanks for your kindness and of course your culinary advice regarding food on Bali, I wish you and your husband all the best.

Beste Svetlana, zonder jouw hulp had dit proefschrift er absoluut niet in de huidige vorm gelegen. Jouw statistische inzichten waren namelijk echt cruciaal bij de ontwikkeling van het predictiemodel. Daarnaast heb jij me essentiële R-skills bijgebracht en was het ook vaak gezellig op borrels. Bedankt!

Marcel, naast een gezellige collega was jij ook mijn onafhankelijke begeleider, een rol die jij goed vervulde aan waarmee je me absoluut hebt geholpen. Bedankt! Patrick, bedankt voor de gezelligheid en de ICT-support! Aukje, jouw open manier van leiding- en invulling geven aan mijn onderwijsaanstelling heb ik enorm gewaardeerd. Daarnaast was jouw oratiefeest natuurlijk fantastisch. Ineke, Suzanne en Anja, bedankt voor al jullie hulp, tijd en de vele dingen die jullie voor de afdeling hebben georganiseerd!

Delphi, mooi dat ook jouw boek af is, gefeliciteerd! Milou, top dat jij verder werkt aan het HTA-onderzoek aan de UU! Wim, bedankt voor de gezelligheid; we gaan elkaar vast nog regelmatig tegenkomen! Marle, jouw immer semi-vage maar altijd gezellige verhalen ga ik zeker missen! Lenneke, dank voor alle gezelligheid en hilarische gesprekken. Pieter, mannen met smaak voor koffie, muziek én een goede haarlijn zijn moeilijk te vinden maar gelukkig bracht jij hoop in deze immense duisternis. Rosanne, Mirjam, Rachel, Nicholas, Gert-Jan, Ard, Debbie, Sander, Doerine, Elise, Jeroen, Ali, Mohammad en alle andere FenF collega's: Jullie ook ontzettend bedankt voor alle leuke momenten, jullie tijd en gezelligheid!

Mark Jut, Vince Sindhunata en Aljosja Karpenko: Als studenten hebben jullie een belangrijke bijdrage geleverd aan mijn proefschrift, dank voor jullie inzet.

Beste Ivo, wij kennen elkaar als sinds de middelbare school en, ondanks dat jij in Groningen farmacie ging studeren, zijn we elkaar regelmatig blijven zien. Naast deze gezamenlijke interesse delen we ook de voor velen onbegrijpelijke liefde voor gamen. Ook was jij tijdens carnaval de perfecte wingman tussen een niet nader te noemen soldaat en kangoeroe; een ingreep waar ik tot op de dag van vandaag erg gelukkig mee ben. Ivo, bedankt voor alle gezelligheid en dat er nog veel mooie momenten mogen komen!

Ingmar, Thijs, Ben, Dedmer, Teun, Frank, Jord, Guus, Patrick, Russel en Pasquale: Wat hebben wij een mooie jaarclub. Met de mooie, absurde en vooral bijzonder gezellige dingen die wij de afgelopen 8 (!) jaar hebben meegemaakt zou ik pagina's kunnen vullen. Schaapcie, de clubweekenden, Van Plaggen, lustrumreis, wintersport, de Rue, het zakelijke succes van Constant Cars en natuurlijk het kleinigheidje met een aap en een telefoon: Een kleine greep uit de stuk voor stuk debiele en hilarische dingen die ik met andere clubs of vrienden nooit had meegemaakt, zonder dit soort zaken was mijn PhD-tijd écht minder leuk geweest. Op naar het volgende lustrum!

Heren van het P.H.B.G: bedankt voor alle fantastische borrels en activiteiten maar vooral ook, zoals het goede apothekers betaamt, voor de uitstekende gesprekken zowel tijdens als voor de borrels. In het bijzonder wil ik kort stilstaan bij de club met wie we de afgelopen jaren honderden gezinnen met een chronisch ziek kind een fantastische dag hebben bezorgd: De Vet Cool Man is ieder jaar weer iets om naar uit te kijken, ga vooral zo door!

Lieve familie, jullie spelen natuurlijk een cruciale rol in mijn leven en zonder jullie liefde, steun en support had ik hier nooit gestaan. Pa en ma, Bert en Irma: Jullie hebben altijd al mijn keuzes gesteund en op ontelbaar veel plekken proberen te helpen. Bedankt voor alles!

Lieve Laurie, op Pubmed heb je me absoluut verslagen maar gelukkig ben ik nog (net) eerder met de doctorstitel. Bedankt voor al je belletjes en de gezelligheid, ik ben super trots op je!

Lieve Koen, met Lotte en bij de Rabo heb jij het top voor elkaar. Nu je zowaar in Utrecht wil gaan wonen wordt het vast nóg gezelliger. Jullie beide bedankt voor alle leuke dingen die we de afgelopen jaren hebben gedaan!

Lieve Emma, van begin tot eind heb jij achter mij en mijn keuze voor deze PhD gestaan en heb je me proberen te helpen waar dat kon. Gelukkig is dit ook voor ons beide een leuke tijd gebleken, al heb ik jou af en toe wel wat stress bezorgd doordat ik me relatief weinig zorgen maakte. Zoals je weet heb ik niet alleen ontzettend veel zin in een nieuwe uitdaging na mijn PhD maar vooral ook in onze hopelijk hele lange toekomst samen. Liefste Em, ik had dit nooit zonder jou willen doen, op ons volgende hoofdstuk!

Joost, oktober 2019

8.2

List of co-authors

M.J. (Maiwenn) Al, PhD
Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, the Netherlands.

F.W. (Folkert) Asselbergs, MD, PhD
Department of Cardiology, Division Heart & Lungs, UMC Utrecht, Utrecht, The Netherlands
Durrer Center for Cardiovascular Research, Netherlands Heart Institute, Utrecht, The Netherlands
Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom

E.V. (Ekaterina) Baranova, MD, PhD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

S.V. (Svetlana) Belitser, MSc
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

M. (Martijn) van Bloois, MSc
Brabers, Gouda, The Netherlands

A. (Anthonius) de Boer, MD, PhD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

C. (Cornelis) Boersma, PharmD, PhD
Health-Ecore, Zeist, The Netherlands
Division of Global Health, Department of Health Sciences, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

J.J. (Joost) Enzing, MSc
The National Healthcare Institute, Diemen, The Netherlands
Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

W.G. (Wim) Goettsch, PhD
The National Health Care Institute (ZIN), Diemen, The Netherlands
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

R.M.T. (Renske) ten Ham, PharmD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

A.M. (Anke) Hövels, PhD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

M. (Mark) Jut, PharmD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

O.H. (Olaf) Klungel, PharmD, PhD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

S. (Saskia) Knies, PhD
The National Healthcare Institute, Diemen, The Netherlands
Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

H.G.M. (Hubert) Leufkens, PhD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands

A.H. (Anke-Hilse) Maitland-van der Zee, PharmD, PhD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands
Department of Respiratory Disease, Academic Medical Center, Amsterdam, The Netherlands

C.N.A. (Colin) Palmer, PhD
Population Pharmacogenetics Group, Biomedical Research Centre, University of Dundee, Ninewells Hospital & Medical School, Dundee, United Kingdom

R.A. (Rick) Vreman, PharmD
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Science, Utrecht University, Utrecht, The Netherlands
The National Healthcare Institute, Diemen, The Netherlands

8.3

List of publications

Accuracy of budget impact estimations and impact on patient access: a hepatitis C case study. **Geenen JW**, Boersma C, Klungel OH, Hövels AM. Eur J Health Econ. 2019 Aug;20(6):857-867. doi: 10.1007/s10198-019-01048-z. Epub 2019 Apr 5.

Early health technology assessments in pharmacogenomics: a case example in cardiovascular drugs. **Geenen JW**, Baranova EV, Asselbergs FW, de Boer A, Vreman RA, Palmer CN, Maitland-van der Zee AH, Hövels AM. Pharmacogenomics. 2017 Aug;18(12):1143-1153. doi: 10.2217/pgs-2017-0063. Epub 2017 Jul 26.

Phase I/II Clinical Trial-Based Early Economic Evaluation of Acalabrutinib for Relapsed Chronic Lymphocytic Leukaemia. Vreman RA, **Geenen JW**, Hövels AM, Goettsch WG, Leufkens HGM, Al MJ. Appl Health Econ Health Policy. 2019 Jul 18. doi: 10.1007/s40258-019-00496-1. [Epub ahead of print]

OTHER PUBLICATIONS

Cost of illness of metastatic prostate cancer: a perspective of costs for new treatment options in The Netherlands. Ter Heine R, Frederix GW, **Geenen JW**, Hövels AM, van Vulpen M, Kooistra A, De Klerk JM, Bloemendal HJ. J Comp Eff Res. 2017 Oct;6(7):575-581. doi: 10.2217/cer-2017-0026. Epub 2017 Sep 13.

Whole-exome sequencing in intellectual disability; cost before and after a diagnosis. Vrijenhoek T, Middelburg EM, Monroe GR, van Gassen KLI, **Geenen JW**, Hövels AM, Knoers NV, van Amstel HKP, Frederix GWJ. Eur J Hum Genet. 2018 Nov;26(11):1566-1571. doi: 10.1038/s41431-018-0203-6. Epub 2018 Jun 29.

ABOUT THE AUTHOR

Joost Geenen (1990)

Joost Geenen was born in Tilburg, the Netherlands. After moving to Goes, he graduated at the Goese Lyceum in 2008. Hereafter, he started studying pharmacy at Utrecht University and obtained his bachelor's degree in 2012. Whilst pursuing his pharmacy master's, Joost performed research internships at the Meander Medical Centre in Amersfoort and the GlaxoSmithKline global headquarters in London. He consequently graduated as a pharmacist (MSc, PharmD) in 2015.

After his graduation, Joost stayed at Utrecht University to pursue a PhD degree at the Division of Pharmacoepidemiology and Clinical Pharmacology of the Utrecht Institute for Pharmaceutical Sciences (UIPS). He was supervised by prof. dr. J.A.M. Raaijmakers†, prof dr. O.H. Klungel, dr. A.M. Hövels and dr. C. Boersma and aimed to address the roles of budget impact, cost-effectiveness and the associated uncertainty in reimbursement decision-making. His studies involved the development of new methodology and modelling techniques using R.

Besides his PhD research, Joost was a part-time lecturer, developed pharmacoeconomic courses, was a member of the division's social committee and co-founded and chaired the Utrecht student chapter of ISPOR. Since 2013, Joost has been actively involved in the 'Stichting Geluk en Vrijheid', a charity that organises events for families with chronically ill children.