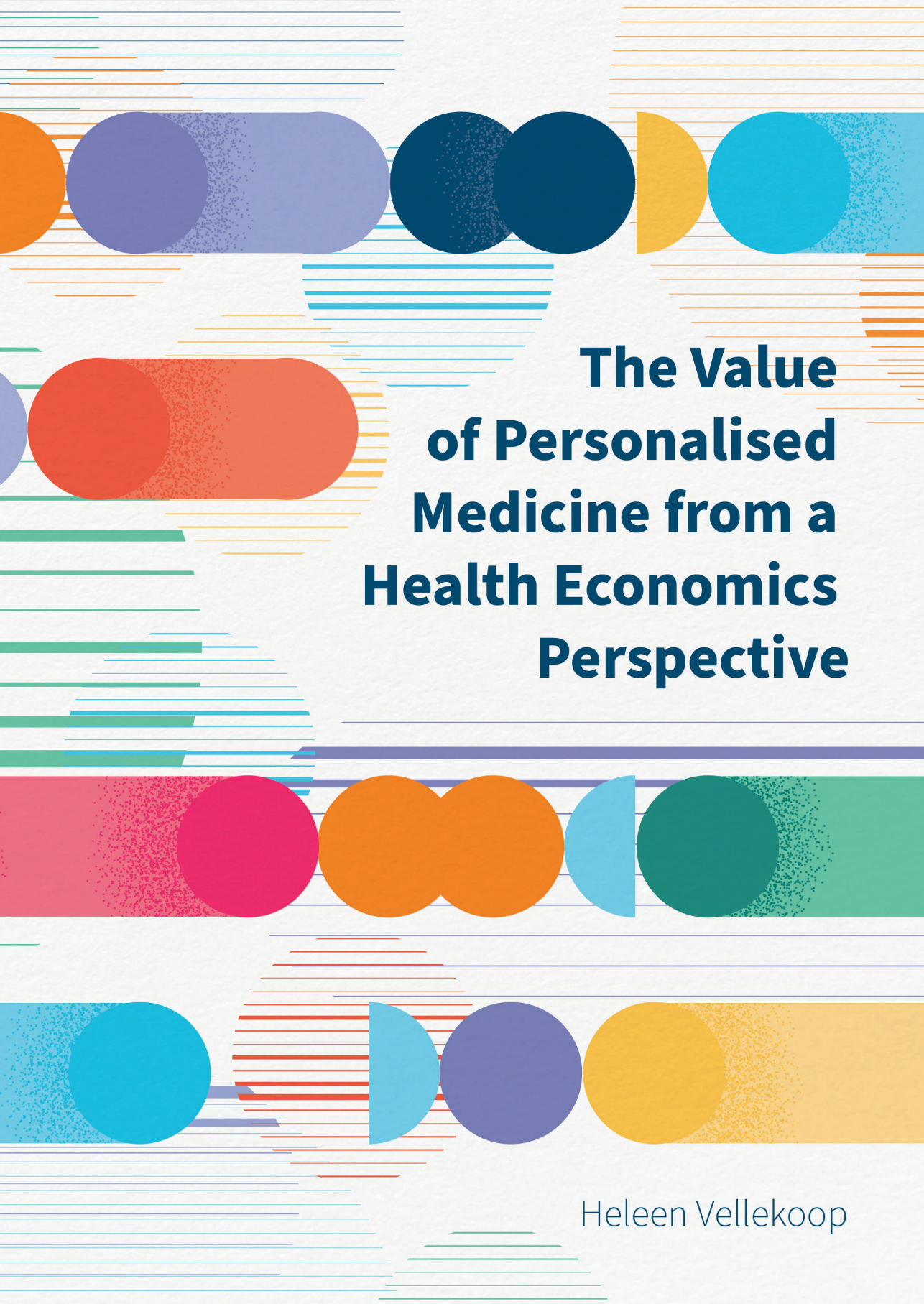


The Value of Personalised Medicine from a Health Economics Perspective

Heleen Vellekoop

The Value of Personalised Medicine from a Health Economics Perspective

Heleen Vellekoop

Colofon

Printing: Ridderprint, ridderprint.nl

Layout and design: Ninke van Wiggen, persoonlijkproefschrift.nl

Funding: The research presented in this thesis was part of the HEcoPerMed project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824997.

The Value of Personalised Medicine from a Health Economics Perspective

De waarde van gepersonaliseerde geneeskunde vanuit een gezondheidseconomisch perspectief

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. A.L. Bredenoord
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 9 oktober 2024 om 15:30 uur

door

Heleen Elianne Vellekoop
geboren te Liemeer

Erasmus University Rotterdam

Promotiecommissie

Promotor	Prof.dr. M.P.M.H. Rutten-van Mölken
Overige leden	Prof.dr. P.M. Bossuyt Prof.dr. V.P. Retel Prof.dr. R.H.N. van Schaik
Copromotoren	Dr. S.A. Huygens Dr. M.M. Versteegh

Contents

CHAPTER 1

General introduction

The research presented in this PhD thesis was conducted within a large (Horizon 2020-funded) project called "Health economics for personalised medicine", abbreviated to "HEcoPerMed". Indeed, the research that we carried out sat right at the junction of the fields of *health economics* on the one hand and *personalised medicine* on the other. Before outlining the main aims and the structure of this thesis, we will take a deep dive into the history and the meaning of personalised medicine and discuss some of the key concepts in health economics.

What is personalised medicine?

The term "personalised medicine" is associated with the idea that the same approach does not work for everyone: healthcare should be tailored according to our individual characteristics. Doing so may have many positive consequences – such as improved treatment effectiveness, less adverse drug reactions, more diseases prevented – that in turn could lead to improved health outcomes and/or a reduction in wasteful healthcare spending.

The historical context

The term "personalised medicine" first appeared in a 1999 article in The Wall Street Journal entitled "New Era of Personalised Medicine: Targeting Drugs for Each Unique Genetic Profile".[1] The article described a consortium that had just been established between ten major pharmaceutical companies and a number of academic research partners in the US and the UK, with the aim of doing genetic research. From the get-go, personalised medicine was a future-oriented concept, with the article describing the *hope* "to create a map of genetic landmarks that will become a potent new tool" and the *hope* of pharmaceutical companies to develop "drugs that can more precisely target the variety of biological quirks that underlie each major disease". Elliott Sigal, vice president of applied genomics at Bristol-Myers at the time, was quoted describing the planned research efforts as "a grand experiment" and "step 1 in a long research process".

Two important scientific developments contributed to the emergence of personalised medicine. First, from the 1970s onward various new technologies were developed in the field of molecular biology.[2] Several techniques for the sequencing of DNA and RNA were introduced, allowing for genetic differences across humans to be explored. Prior to these developments, the pharmaceutical industry had been characterised by Fordism, with standardised products targeted at large patient populations.[3] Hedgecoe & Martin describe the pharmaceutical industry during this period as having a "profound reluctance to admit the extent of genetic variation and its effect on drug response".[4] They posit the "unwillingness of the industry to admit to the huge scale and very real risks posed by genetically based adverse drug reactions" as a possible reason.[4] Heterogeneity in

treatment effects was seen as "nuisance"[5] that was often countered by only including a relatively homogeneous subset of the patient population in clinical trials: usually young white males.[6] The technological advancements in molecular biology, however, created opportunities to engage with patient heterogeneity in a different manner. The pharmaceutical industry welcomed the new possibilities enthusiastically, with The Wall Street Journal article describing the goal of pharmaceutical companies as "a cornucopia of personalised medicines that will produce huge profits into the next century".[1]

Second, the latter half of the twentieth century saw large improvements in information technology leading to much larger capacity for data storage and processing, as well as the development of many new statistical methods. Perhaps counterintuitively, a personalised approach requires large amounts of aggregate data. Think of websites like Amazon, for example, where the products that are suggested to the individual user ("recommended for you") are largely based on the browsing and buying behaviour of many other users. Statistical tools are used to move between aggregate data and predictions at the individual level.[7] The Framingham Heart Study, a long-term cohort study initiated in 1948 to establish incidence rates of heart disease, was used to develop one of the first prediction models in medicine. Around the 1960s, statistical measures were introduced that could link specific biomarkers, behaviours, and other patient characteristics to the probability of developing heart disease, later resulting in the official Framingham Risk Score model.[7] As statistical methods are improving and as data collection expands, scientists can make increasingly precise claims about which individual attributes are associated with which diseases and treatment outcomes.[8]

Present-day debates about the definition of personalised medicine

Since the first coining of the term, personalised medicine has garnered many supporters and attracted large amounts of (research) funding.[9] Some have described the general sentiment around personalised medicine as "buzz" and "hype".[10–12] The popularity of personalised medicine notwithstanding, the meaning of the term is somewhat diffuse. Discourse has emerged around the definition of personalised medicine, which can be divided into two main debates: one about the scope of the context (which types of healthcare should be included), and one about the most appropriate name for what is being referred to as personalised medicine.[13]

What is included in "personalised medicine"?

The headline of The Wall Street Journal article includes the phrase "Targeting Drugs for Each Unique Genetic Profile", indicating that the initiators of the term "personalised medicine" intended for it to refer to pharmaceutical treatments (as opposed to other forms of healthcare) and for the treatments to vary according to genetic differences (as

opposed to other relevant differences between people).[1] However, the scope of the term appears to have expanded since.

The US Food and Drug Administration (FDA) defines personalised medicine as

> "an innovative approach to tailoring disease prevention and treatment that takes into account differences in people's genes, environments, and lifestyles".[14]

The European Commission (EC) defines personalised medicine as

> "a medical model using characterisation of individuals' phenotypes and genotypes (e.g. molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention".[15]

It seems that the focus on pharmaceuticals has been broadened to the more general "disease prevention and treatment" (FDA) and "therapeutic strategy" (EC). Beyond genetic information, "environments and lifestyles" (FDA) and "phenotypes" (EC) are also emphasised as variables to be considered in the tailoring of healthcare. These broadened definitions raise questions about the seemingly large overlap between personalised medicine and "regular" medicine. Why would we need a new term for a practice that has existed for decades already? Indeed, Schleidgen et al., who conducted a systematic literature review on the various uses of the term personalised medicine, argue that a *new* term should be used to indicate *new* developments.[11]

Is "personalised medicine" the right name?

Schleidgen et al. find that most uses of "personalised medicine" in the scientific literature refer to the stratification of patients into subgroups with the help of (newly discovered or newly developed) biomarkers.[11] Treatments are adjusted according to the subgroup patients are in, while previously all patients in the same larger group would have received the same treatment. Although they offer a definition of personalised medicine based on their findings, Schleidgen et al. conclude by proposing we use the term "stratified medicine" instead. The World Health Organisation (WHO), too, has stated that "stratified medicine" is most appropriate, arguing that it reflects the clinical reality most accurately.[16] Nonetheless, "stratified medicine" is used much more scarcely than "personalised medicine". Arguably this is because of the latter's rhetorical force. Indeed, advocates for person-centred care, who strive toward more individualised care using a

holistic view of patients (as opposed to a biomarker-focused perspective)[17] have accused the field of pharmacogenomics of co-opting a term with strong positive associations among the general public (who wouldn't want their healthcare personalised?) for its own advancement.[18–20] And within the field of pharmacogenomics, too, there have been doubts whether "personalised medicine" is the right term, with "precision medicine" frequently being used as an alternative.[21]

De Grandis & Halgunset suggest that a single definition of "personalised medicine" is unlikely, as the concept concerns a diverse range of stakeholders, operating in different contexts.[13] They also argue that the term, future oriented as it is, describes a cluster of visions for the future of healthcare more than an existing reality. To a large extent, they say, the purpose of the term (and the hype around it) is to mobilise funds and to build networks. The meaning of personalised medicine is therefore fluid and may shift over time, depending on which technologies actualise and political developments such as "who gets involved", "who pays, and for what".[13]

Personalised medicine in clinical practice

In line with the original coining of the term, in this PhD thesis personalised medicine is understood to revolve around the use of genetic information. So far, most developments in the use of genetic information in clinical practice have been in the field of oncology. A major reason for this is the nature of cancer: tumours develop because of errors in DNA replication, and they can contain a multitude of DNA "mistakes" in advanced disease stages. Given the central role of genetic changes in cancer, the profiling and targeting of various genetic markers may be more valuable in cancer treatment than in other disease areas. Cancer also has a huge burden of disease and large unmet need. Every year millions of people worldwide receive a new cancer diagnosis and millions of people die from cancer.[22] Many of the cancer treatments that have come onto the market over recent years are intended for patients who have little to no alternatives left and who are expected to die relatively soon. To give terminally ill patients access to newly developed cancer drugs as soon as possible, pharmaceuticals are frequently allowed onto the market through fast-track programmes with looser requirements for clinical evidence, such as "accelerated approval" (FDA) and "conditional market approval" (European Medicines Agency [EMA]).[23,24] In several countries, cancer drugs are also allowed to be more expensive than other pharmaceuticals, as cancer drugs are evaluated through separate processes, such as the pan-Canadian Oncology Drug Review (pCODR) and the Cancer Drugs Fund in the UK.[25,26] This has increased the profitability of developing cancer drugs for pharmaceutical companies, further encouraging the focus for research and development to be on the field of oncology.[27]

Nonetheless, in other disease areas too, personalised medicine interventions have become part of clinical practice. In a systematic literature review of personalised medicine (which they interpret as "stratifying care by genetic characteristics"), Hatz et al.[28] identify four main categories of gene profiling:

1. Screening for a disease or a genetic marker in an asymptomatic population
2. Identifying patients who may experience adverse drug reactions.
3. Gaining information about the prognosis of a disease
4. Distinguishing patients who are likely to respond to a certain treatment from those who are not.

An example of category 1 is BRCA-testing. BRCA-tests can identify variants in the BRCA1 and BRCA2 genes that are associated with an increased risk of developing breast cancer and ovarian cancer.[29] BRCA-testing may be offered to family members of patients for whom pathogenic BRCA-variants have already been identified. Carriers of pathogenic BRCA-variants are subsequently offered risk-reducing surgery. An example outside of oncology is the screening of family members for gene variants associated with hypertrophic cardiomyopathy – a common heart disease that in rare cases can lead to sudden cardiac death.[30]

An example of a test in category 2 is DPYD-profiling prior to administering fluoropyrimidines, which are commonly used cancer drugs.[44,45] Certain variants of the DPYD-gene are associated with a higher risk of developing severe adverse drug reactions. According to the variants identified, patients may be given an adjusted dose of the fluoropymidine drug. In other cases, such as patients with HIV carrying the HLA-B*57:01 allele, patients may be given a different treatment altogether.

Oncotype DX is an example of a test in category 3.[46] The test is used in patients with early-stage breast cancer. Based on a sample of the cancer tumour, the test analyses the activity level of 21 genes and subsequently calculates a risk score indicating the likelihood of metastasis. Patients with a high risk score may subsequently be given preventive chemotherapy.

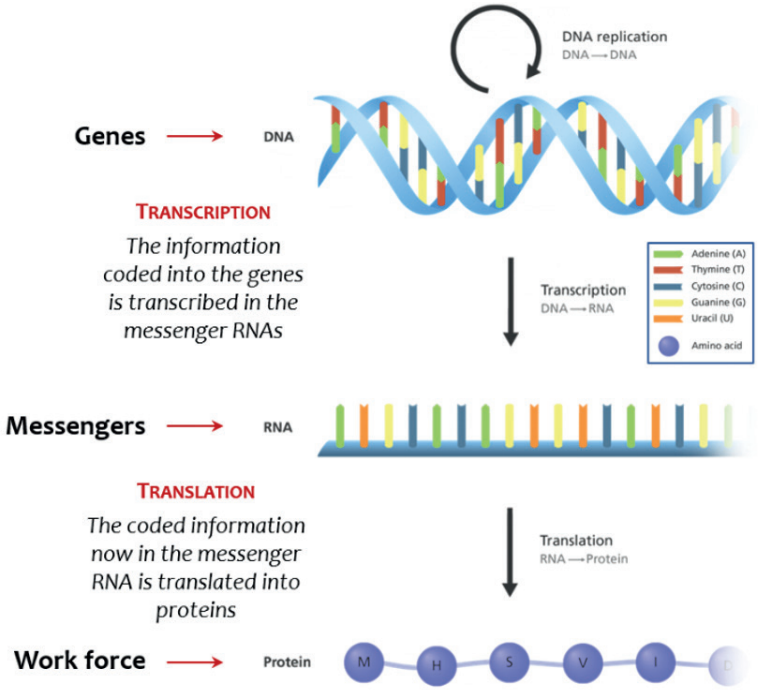
Category 4 tests are common in cancer care, where various new pharmaceuticals have been developed for patients with specific genetic markers. An example is the testing for mutations in the EGFR-gene among patients with non-small cell lung cancer (NSCLC). Mutations in the EGFR-gene can cause cells to grow and divide faster than normal, leading to tumour growth. NSCLC-patients testing positive for (oncogenic) EGFR-mutations are given pharmaceuticals that specifically block the functioning of the EGFR-gene. NSCLC-patients without EGFR-mutations are given a different treatment.

DNA explainer

Deoxyribonucleic acid (DNA) contains the instructions for life and its processes. DNA is made up of two strands that contain nucleobases (the fundamental units of genetic code) and coil around each other to form a double helix.[31] Human DNA is spread over (for most people) 23 chromosome pairs that together constitute the genome, which can be found in most cells in the body.[32]

From gene to protein

The human genome contains thousands of genes. An estimated 20,000 genes code for protein molecules, which determine the structure and functioning of cells, tissues, and organs.[33] The instructions contained within genes are converted into proteins by first transcribing the information encoded in DNA to a messenger RNA (mRNA) molecule and subsequently translating the mRNA into the amino acid sequence of a protein.[34]



Source: [35]

Germline versus tumour DNA

All humans are born with their own genome determining their personal characteristics. The DNA that people are born with is called germline (or constitutional) DNA. The cells that make up a human body are constantly renewed, by replicating the DNA and subsequently dividing into two identical daughter cells.[31] Errors in DNA replication can cause faults in cell division, leading to tumour growth. The DNA found in tumour cells is called tumour (or somatic) DNA. Sometimes, the phrase "genetic testing" is used to refer to the profiling of germline DNA, while "genomic testing" is used to refer to the profiling of tumour DNA. In this PhD thesis, however, the two are used interchangeably.

Variants and mutations

Specific regions of the genome that differ between genomes are called "variants" or "mutations". Certain variants may be associated with (an increased risk of) specific diseases, in which case they are referred to as pathogenic variants. While "mutations" was intended to be a neutral term synonymous to "variant", in practice it is mostly used to indicate pathogenic variants.[36,37] Pathogenic variants may be inherited (i.e. appear in the germline DNA), or they may occur later in life because of errors in DNA replication. Pathogenic variants in cancer specifically are also called "oncogenic variants".

Oncogenic variants

Hundreds of genes are involved in the process of cell division, with some promoting cell division and others keeping cell division under control or repairing mistakes in a cell's DNA.[38,39] An oncogenic variant can cause the delicate balance between the activity of the various genes to become disturbed, resulting in cells growing out of control hence tumour growth. Various genetic mechanisms may cause oncogenic gene variants. Among these are point mutations (such as base substitutions, deletions, and insertions) and chromosomal rearrangements (where sections of the DNA are moved from one location to another).[40,41]

Fusion genes

Fusion genes are hybrid genes that combine parts of two separate genes. They are usually the result of chromosomal rearrangements.[42] Fusion events can contribute to cancer growth through various mechanisms, for example by disrupting the functional domains of tumour suppressor genes.[43] A type of fusion event that is particularly important in the context of this PhD thesis is the fusion of a gene with a strong promoter on the one hand and on the other hand a gene with a functional domain. The strong promoter can lead to the continuous transcription and translation of the fusion gene, meaning that the protein that the gene encodes for is overexpressed compared to what would be a normal level of protein expression.[41] The natural production of a protein in a particular cell type is called "wild-type protein expression", while protein production after a gene alteration is referred to as "mutation protein expression" or, in the case of a fusion event, "fusion protein expression".

Most tests in category 4 (as well as categories 1-3) serve to stratify a subgroup of patients (e.g. patients with NSCLC) into smaller subgroups (e.g. NSCLC-patients with EGFR-mutations, and NSCLC-patients without EGFR-mutations). However, a recent development has been the introduction of so-called "histology-independent" or "tumour-agnostic" cancer drugs. Here, all patients with a specific genetic marker in the tumour DNA receive the same treatment. This means that cancer patients who previously would have fallen in different treatment groups, based on primary site (the location in the body where the cancer started) and histology (tissue type, e.g. carcinoma vs. sarcoma), are now lumped into one treatment group. Histology-independent therapies therefore do not necessarily reflect an increased level of stratification (rather a different manner of stratification), but they have been classified as personalised medicine nonetheless. This appears to lend support to De Grandis and Halgunset's argument that the definition of personalised medicine may shift according to which technologies actualise.[13] Similar reasoning applies to gene therapies, which are innovative treatments where a person's genes are modified.[47] Gene therapies are not necessarily used to stratify patients into smaller subgroups either, yet they are still seen as personalised medicine. Examples of gene therapies are Luxturna[48,49] and Zolgensma[50], both used to treat rare genetic disorders, and CAR-T cell therapy, where T cells are modified and subsequently infused back into the patient to harness the immune system's ability to recognise and destroy cancer cells.[51]

In this thesis, to best capture the available technologies so far, personalised medicine is defined as "a medical model that bases therapeutic choice on the result of gene profiling or aims to correct pathogenic gene mutations".

Why use health economics?

After investigating the concept of personalised medicine, the field of health economics is discussed below. First, some of the main concepts and tools used by health economists are outlined. Then, we explore how health economics may provide valuable insights in the context of personalised medicine.

Crash course in health economics

Health economics emerged as a field in the second half of the twentieth century, and stemmed from the belief that healthcare markets have a combination of distinct features that warrant an analytical lens other than the neo-classical framework of supply and demand.[52,53] These features are, among others, the fact that people usually do not choose when they are in ill-health (i.e. when they have a demand for healthcare services), that patients often do not choose their treatment but rely on the advice of a clinician instead,

and that in many healthcare systems patients do not (fully) pay for healthcare services at the point of use.

In its present-day form, health economics is a broad field of study. Research topics range from understanding the causes for health inequalities to evaluating risk management in health insurance programmes, and from estimating the effects of public health policies to designing pharmaceutical funding models. Although health economics concerns itself with methodological quandaries like any other scientific field, health economics is also a highly applied discipline, with findings of health economic studies frequently used by policymakers.[54] A subfield of health economics that is employed to inform policy decisions particularly often is the field of health technology assessment (HTA). As suggested by its name, HTA is used to assess health technologies, by investigating their properties and their effects in a systematic and transparent way.[55] "Technology", or its synonym "intervention", can be used to describe a range of things, including medical devices, vaccines, medicines, software applications, and even procedures and organisational systems. In practice, HTA is mostly used to evaluate newly developed pharmaceuticals.

The concept of opportunity cost is key in understanding the purpose of HTA. While not everyone might be familiar with the term, its meaning is intuitively clear to most people, even at a young age. When given some coins to go buy biscuits, children already understand that they will not be able to buy everything available at the shop and that they will have to choose between several enticing options. When settling on chocolate chip cookies, the child is no longer able to buy Oreos or shortbread. The other options that the child could have gone for represent the opportunity cost of spending their coins on chocolate chip cookies. The healthcare sector, too, is faced with opportunity cost. More and more healthcare interventions are becoming available – some of them at steep prices -, while healthcare budgets are finite. By evaluating different healthcare interventions with the same methodological approach and the same criteria, HTA provides insights into the value for money that different healthcare interventions provide relative to each other and hence can support decision making on resource allocation.

Value for money?

To evaluate an intervention's value for money, or "cost-effectiveness", its associated costs and health effects are estimated. When interventions are being assessed, for example because they are being considered for inclusion in a health insurance package, they are usually compared to the alternative interventions for the same patient population. Often, new interventions bring health improvements compared to their alternatives, while also being more expensive. The estimated cost increase ("incremental cost") and the estimated health improvement ("incremental benefit") are combined into an incremental

cost-effectiveness ratio (ICER).[56] The ICER reflects how much money we need to spend on the intervention to gain 1 quality-adjusted life year (QALY).

In many countries with national HTA agencies, a cost-effectiveness threshold has been set to determine which ICERs are deemed acceptable. The threshold value indicates the maximum amount of money that is to be paid for a health increase of 1 QALY. If the ICER of an intervention is below the threshold value of the cost-effectiveness threshold, the intervention is deemed good value for money, or "cost-effective".[56] A similar approach is to combine incremental cost, incremental benefit, and the threshold value to calculate net monetary benefit (NMB).[56,57] If NMB is above zero, the intervention is considered cost-effective. **Chapter 2** will provide more details on the calculation of an intervention's ICER and NMB and the relative uses of the two measures.

Although the basic concepts in HTA are widely agreed upon, several differences of opinion exist about the details. An important point of contention is the scope of an HTA, that is, which costs and health benefits should be included, and which should be left out? Some prescribe a healthcare perspective, where only outcomes that affect the healthcare budget are included in the analysis.[56] The main argument is that decision-makers often only have control over resource allocation within a fixed healthcare budget, not over the total government budget. When deciding on how the healthcare budget is spent, decision-makers are therefore faced with healthcare-specific opportunity cost. Costs and benefits that fall outside of the healthcare sector are outside of their remit. Others favour a societal perspective, where further consequences of an intervention – such as changes in the time caregivers spend caring and changes in patients' labour force participation – are also included.[58] Proponents of the societal perspective argue that the overall aim of healthcare decision-making (as well as all other government decision-making) should be to maximise total welfare in society, not just to maximise the health of patients within the healthcare sector.[59,60] In this PhD thesis we do not take a stand on the preferred perspective but rather follow the national HTA guidelines for each country setting.

Another common point of contention is the conceptualisation of the cost-effectiveness threshold: what exactly does the threshold reflect, and how is it measured? Some argue that the threshold value should reflect society's willingness-to-pay for a QALY, while others argue that doing so may lead to health losses and that the threshold should be based on the prevailing marginal opportunity cost in the healthcare system (i.e. how much additional money would we have to spend on ongoing healthcare activities to gain a QALY). More on the two different understandings of the cost-effectiveness thresholds, with the former referred to as "*v*-threshold" and the latter as "*k*-threshold", will follow in **Chapter 2**.

Estimating cost and health outcomes

When an intervention's costs are estimated, both immediate and downstream healthcare expenses associated with the intervention are generally included.[56] The immediate costs of a healthcare intervention usually include the purchasing cost of the intervention itself, the labour costs of healthcare professionals providing the intervention, the cost of materials needed when administering treatment (e.g. syringes, gauzes), the costs of treating side effects, and more.[56] In countries where HTA guidelines prescribe a societal perspective, additional costs may be included (e.g. costs related to productivity losses, costs of informal caregiving, out-of-pocket payments).[58] National variation also exists in the downstream costs that are considered. Some guidelines prescribe that only future medical costs related to the disease in question (e.g. follow-up visits for monitoring) are included, while others prescribe the inclusion of future medical costs that are not directly linked to the disease under evaluation and/or future non-medical costs.[61–64]

Costs can be expressed as a monetary value. The measurement of health, on the other hand, is less straightforward. There are many components to health, and indicators of improvements in health vary across disease areas. Yet, to be able to decide on the allocation of national healthcare budgets, we need a single measurement unit to express health. The most common measure of health in HTAs is the quality-adjusted life year, abbreviated to QALY.[65] The QALY captures the effects that healthcare interventions have on both length of life and quality of life. It is expressed as a utility value between 0 (dead) and 1 (perfect health), where the utility values for various health states are estimated using preference elicitation studies.[56,65]

Economic models are used to combine input data on costs and health benefits from various sources, and to extrapolate from the available short-term data (e.g. clinical trials with a limited observation period) to make predictions about long-term (usually lifetime) consequences for the patient.[56] Because the predictions from such economic models come with a degree of uncertainty, an HTA normally also includes an uncertainty analysis to assess how much a change in the input data or in other elements of the model would affect the estimated long-term consequences.[56,66,67]

Health economics and personalised medicine

The emergence of personalised medicine has raised questions for health economists. A key question concerns the added value of personalised medicine. Alongside excitement about the possibilities of personalised medicine, there have been worries about the high costs of some of the personalised medicine interventions that have become available for clinical practice. Can societies carry the burden of these expensive treatments? And are they worth it?

Another important question is a methodological one. Personalised medicine, with its heavy focus on genetic markers and innovative approaches, differs from much of the healthcare that came before. Concerns have arisen as to whether the methodological toolkit that has been used in HTA for the past few decades is sufficiently suitable for evaluations of personalised medicine interventions. Perhaps new or updated approaches are needed?

Thesis outline

The main objectives of this thesis are to provide insights into the added value of personalised medicine and to investigate methodological challenges in evaluating personalised medicine.

Chapter 2 and **Chapter 3** are both based on a systematic literature review of economic evaluations of personalised medicine. In **Chapter 2** the estimated incremental costs and health effects of personalised medicine interventions (compared to non-personalised medicine interventions) are used to provide insights into the net benefit of personalised medicine. Regression analysis is performed to better understand which types of personalised medicine might offer more added value than others. In **Chapter 3** we investigate the methodological approaches that have been used in economic evaluations of personalised medicine so far. Findings from the literature are combined with a series of expert interviews, resulting in methodological guidance for economic evaluations of personalised medicine. **Chapters 4-6** show a practical example of how a personalised medicine intervention might be assessed. We evaluate the cost-effectiveness of testing for oncogenic neurotrophic tropomyosin kinase receptors (NTRK) gene fusions among cancer patients and subsequently providing the histology-independent treatment entrectinib to those with a positive test result. We first estimate the expected effect of NTRK gene fusions on disease prognosis in **Chapter 4**. In **Chapter 5** the cost-effectiveness of treating patients with NTRK gene fusions with entrectinib is evaluated for the Netherlands. The strategy for NTRK gene fusion-testing is assumed to be as recommended by a Dutch expert group. In **Chapter 6** we expand on Chapter 5 by adding two additional countries (Hungary and the UK) and by comparing several different NTRK-testing strategies.

References

1. Langreth R, Waldholz M. New Era of Personalized Medicine: Targeting Drugs For Each Unique Genetic Profile. ***The Oncologist***. 1999;4(5):426-427. doi:10.1634/theoncologist.4-5-426
2. Morange M. ***A History of Molecular Biology***. Harvard University Press; 2000.
3. Tutton DR. ***Genomics and the Reimagining of Personalized Medicine***. Ashgate Publishing, Ltd.; 2014.
4. Hedgecoe A, Martin P. The Drugs Don't Work: Expectations and the Shaping of Pharmacogenetics. ***Soc Stud Sci***. 2003;33(3):327-364. doi:10.1177/03063127030333002
5. Kosorok MR, Laber EB. Precision Medicine. ***Annu Rev Stat Its Appl***. 2019;6:263-286. doi:10.1146/annurev-statistics-030718-105251
6. Corrigan OP. 'First in Man': The Politics and Ethics of Women in Clinical Drug Trials. ***Fem Rev***. 2002;72(1):40-52. doi:10.1057/palgrave.fr.9400055
7. Phillips CJ. Precision Medicine and Its Imprecise History. ***Harv Data Sci Rev***. 2020;2(1). doi:10.1162/99608f92.3e85b56a
8. Steyerberg EW. ***Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating***. Springer Science & Business Media; 2008.
9. Nimmesgern E, Norstedt I, Draghia-Akli R. Enabling personalized medicine in Europe by the European Commission's funding activities. ***Pers Med***. 2017;14(4):355-365. doi:10.2217/pme-2017-0003
10. Maughan T. The Promise and the Hype of 'Personalised Medicine.' ***New Bioeth***. 2017;23(1):13-20. doi:10.1080/20502877.2017.1314886
11. Schleidgen S, Klingler C, Bertram T, Rogowski WH, Marckmann G. What is personalized medicine: sharpening a vague term based on a systematic literature review. ***BMC Med Ethics***. 2013;14(1):55. doi:10.1186/1472-6939-14-55
12. Precision medicine in acute myeloid leukemia: Hope, hype or both? - ScienceDirect. Accessed September 2, 2023. https://www.sciencedirect.com/science/article/pii/S014521261630162X?casa_token=g_qAkMs7BXYAAAAA:z4jmmhIPhG2mlG8kFjEXXG9wQmm1d5UN5zsHf5qRiSDE65aFWKokKv-DRoR1gsuyy4XJOAhf
13. De Grandis G, Halgunset V. Conceptual and terminological confusion around personalised medicine: a coping strategy. ***BMC Med Ethics***. 2016;17(1):43. doi:10.1186/s12910-016-0122-4
14. Health C for D and R. Precision Medicine. FDA. Published August 18, 2023. Accessed August 27, 2023. https://www.fda.gov/medical-devices/in-vitro-diagnostics/precision-medicine
15. Personalised medicine. Published July 19, 2023. Accessed August 27, 2023. https://health.ec.europa.eu/medicinal-products/personalised-medicine_en
16. Jørgensen JT. Twenty Years with Personalized Medicine: Past, Present, and Future of Individualized Pharmacotherapy. ***The Oncologist***. 2019;24(7):e432-e440. doi:10.1634/theoncologist.2019-0054
17. El-Alti L, Sandman L, Munthe C. Person Centered Care and Personalized Medicine: Irreconcilable Opposites or Potential Companions? ***Health Care Anal***. 2019;27(1):45-59. doi:10.1007/s10728-017-0347-5
18. Cherny NI, de Vries EGE, Emanuel L, et al. Words Matter: Distinguishing "Personalized Medicine" and "Biologically Personalized Therapeutics." ***JNCI J Natl Cancer Inst***. 2014;106(12):dju321. doi:10.1093/jnci/dju321
19. Heusser P. Towards Integration of 'Personalised' and 'Person-Centred' Medicine: The Concept of 'Integrative and Personalised Health Care.' In: ***The Ethics of Personalised Medicine***. Routledge; 2015.
20. Burke W, Trinidad SB, Press NA. Essential Elements of Personalized Medicine. ***Urol Oncol***. 2014;32(2):193-197. doi:10.1016/j.urolonc.2013.09.002
21. Juengst E, McGowan ML, Fishman JR, Settersten Jr. RA. From "Personalized" to "Precision" Medicine: The Ethical and Social Implications of Rhetorical Reform in Genomic Medicine. ***Hastings Cent Rep***. 2016;46(5):21-33. doi:10.1002/hast.614
22. Roser M, Ritchie H. Cancer. ***Our World Data***. Published online July 3, 2015. Accessed August 27, 2023. https://ourworldindata.org/cancer
23. Beaver JA, Howie LJ, Pelosof L, et al. A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics: A Review. ***JAMA Oncol***. 2018;4(6):849-856. doi:10.1001/jamaoncol.2017.5618

24. EMA. Support for early access. European Medicines Agency. Published September 17, 2018. Accessed August 26, 2023. https://www.ema.europa.eu/en/human-regulatory/overview/support-early-access
25. Shin G, Kwon HY, Bae S. For Whom the Price Escalates: High Price and Uncertain Value of Cancer Drugs. ***Int J Environ Res Public Health***. 2022;19(7):4204. doi:10.3390/ijerph19074204
26. Dixon P, Chamberlain C, Hollingworth W. Did It Matter That the Cancer Drugs Fund Was Not NICE? A Retrospective Review. ***Value Health***. 2016;19(6):879-884. doi:10.1016/j.jval.2016.04.001
27. Michaeli DT, Yagmur HB, Achmadeev T, Michaeli T. Valuation and Returns of Drug Development Companies: Lessons for Bioentrepreneurs and Investors. ***Ther Innov Regul Sci***. 2022;56(2):313-322. doi:10.1007/s43441-021-00364-y
28. Hatz MHM, Schremser K, Rogowski WH. Is Individualized Medicine More Cost-Effective? A Systematic Review. ***PharmacoEconomics***. 2014;32(5):443-455. doi:10.1007/s40273-014-0143-0
29. Gori S, Barberis M, Bella MA, et al. Recommendations for the implementation of BRCA testing in ovarian cancer patients and their relatives. ***Crit Rev Oncol Hematol***. 2019;140:67-72. doi:10.1016/j.critrevonc.2019.05.012
30. Bonaventura J, Polakova E, Vejtasova V, Veselka J. Genetic Testing in Patients with Hypertrophic Cardiomyopathy. ***Int J Mol Sci***. 2021;22(19):10401. doi:10.3390/ijms221910401
31. Snustad DP, Simmons MJ. ***Principles of Genetics***. John Wiley & Sons; 2015.
32. National Research Council (US) Committee on Mapping and Sequencing the Human Genome. ***Mapping and Sequencing the Human Genome***. National Academies Press (US); 1988. Accessed August 26, 2023. http://www.ncbi.nlm.nih.gov/books/NBK218252/
33. The complete sequence of a human genome | Science. Accessed August 26, 2023. https://www.science.org/doi/10.1126/science.abj6987
34. Translation: DNA to mRNA to Protein | Learn Science at Scitable. Accessed August 26, 2023. https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393/
35. From Gene to Protein - LGMD2i Research Fund | LGMD2i Research Fund. Accessed August 26, 2023. https://www.lgmd2ifund.org/science-basics/from-gene-to-protein
36. Updates in Genetic Terminology: From “Mutation” to “Variant” | Basser Center. Accessed August 26, 2023. https://www.basser.org/resources/updates-genetic-terminology-mutation-variant
37. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. ***Genet Med***. 2015;17(5):405-424. doi:10.1038/gim.2015.30
38. Proto-oncogenes to Oncogenes to Cancer | Learn Science at Scitable. Accessed August 26, 2023. https://www.nature.com/scitable/topicpage/proto-oncogenes-to-oncogenes-to-cancer-883/
39. Cell Division, Cancer | Learn Science at Scitable. Accessed August 26, 2023. https://www.nature.com/scitable/topicpage/cell-division-and-cancer-14046590/
40. Pierotti MA, Sozzi G, Croce CM. Mechanisms of oncogene activation. In: ***Holland-Frei Cancer Medicine. 6th Edition***. BC Decker; 2003. Accessed August 26, 2023. https://www.ncbi.nlm.nih.gov/books/NBK12538/
41. Latysheva NS, Babu MM. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. ***Nucleic Acids Res***. 2016;44(10):4487-4503. doi:10.1093/nar/gkw282
42. Taniue K, Akimitsu N. Fusion Genes and RNAs in Cancer Development. ***Non-Coding RNA***. 2021;7(1):10. doi:10.3390/ncrna7010010
43. Tuna M, Amos CI, Mills GB. Molecular mechanisms and pathobiology of oncogenic fusion transcripts in epithelial tumors. ***Oncotarget***. 2019;10(21):2095-2111. doi:10.18632/oncotarget.26777
44. Lunenburg CATC, Henricks LM, Guchelaar HJ, et al. Prospective DPYD genotyping to reduce the risk of fluoropyrimidine-induced severe toxicity: Ready for prime time. ***Eur J Cancer***. 2016;54:40-48. doi:10.1016/j.ejca.2015.11.008
45. Lunenburg CATC, Henricks LM, van Kuilenburg ABP, et al. Diagnostic and Therapeutic Strategies for Fluoropyrimidine Treatment of Patients Carrying Multiple DPYD Variants. ***Genes***. 2018;9(12):585. doi:10.3390/genes9120585
46. McVeigh TP, Kerin MJ. Clinical use of the Oncotype DX genomic test to guide treatment decisions for patients with invasive breast cancer. ***Breast Cancer Targets Ther***. 2017;9:393-400. doi:10.2147/BCTT.S109847

47. Prado DA, Acosta-Acero M, Maldonado RS. Gene therapy beyond luxturna: a new horizon of the treatment for inherited retinal disease. ***Curr Opin Ophthalmol***. 2020;31(3):147. doi:10.1097/ICU.0000000000000660
48. Maguire AM, Russell S, Wellman JA, et al. Efficacy, Safety, and Durability of Voretigene Neparvovec-rzyl in RPE65 Mutation–Associated Inherited Retinal Dystrophy: Results of Phase 1 and 3 Trials. ***Ophthalmology***. 2019;126(9):1273-1285. doi:10.1016/j.ophtha.2019.06.017
49. Darrow JJ. Luxturna: FDA documents reveal the value of a costly gene therapy. ***Drug Discov Today***. 2019;24(4):949-954. doi:10.1016/j.drudis.2019.01.019
50. Stevens D, Claborn MK, Gildon BL, Kessler TL, Walker C. Onasemnogene Abeparvovec-xioi: Gene Therapy for Spinal Muscular Atrophy. ***Ann Pharmacother***. 2020;54(10):1001-1009. doi:10.1177/1060028020914274
51. Sterner RC, Sterner RM. CAR-T cell therapy: current limitations and potential strategies. ***Blood Cancer J***. 2021;11(4):1-11. doi:10.1038/s41408-021-00459-7
52. Arrow KJ. 21 - UNCERTAINTY AND THE WELFARE ECONOMICS OF MEDICAL CARE. In: Diamond P, Rothschild M, eds. ***Uncertainty in Economics***. Academic Press; 1978:345-375. doi:10.1016/B978-0-12-214850-7.50028-0
53. Culyer AJ, Newhouse JP. ***Handbook of Health Economics***. Elsevier; 2000.
54. Culyer AJ, Newhouse JP. Introduction: The State and Scope of Health Economics. In: Culyer AJ, Newhouse JP, eds. ***Handbook of Health Economics***. Vol 1. Handbook of Health Economics. Elsevier; 2000:1-8. doi:10.1016/S1574-0064(00)80159-6
55. O'Rourke B, Oortwijn W, Schuller T, Group the IJT. The new definition of health technology assessment: A milestone in international collaboration. ***Int J Technol Assess Health Care***. 2020;36(3):187-190. doi:10.1017/S0266462320000215
56. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. ***Methods for the Economic Evaluation of Health Care Programmes***. Oxford University Press; 2015.
57. Paulden M. Calculating and Interpreting ICERs and Net Benefit. ***PharmacoEconomics***. 2020;38(8):785-807. doi:10.1007/s40273-020-00914-6
58. Avşar TS, Yang X, Lorgelly P. How is the Societal Perspective Defined in Health Technology Assessment? Guidelines from Around the Globe. ***PharmacoEconomics***. 2023;41(2):123-138. doi:10.1007/s40273-022-01221-y
59. Wouterse B, van Baal P, Versteegh M, Brouwer W. The Value of Health in a Cost-Effectiveness Analysis: Theory Versus Practice. ***PharmacoEconomics***. 2023;41(6):607-617. doi:10.1007/s40273-023-01265-8
60. Jönsson B. Ten arguments for a societal perspective in the economic evaluation of medical innovations. ***Eur J Health Econ***. 2009;10(4):357-359. doi:10.1007/s10198-009-0173-2
61. Morton A, Adler AI, Bell D, et al. Unrelated Future Costs and Unrelated Future Benefits: Reflections on NICE Guide to the Methods of Technology Appraisal. ***Health Econ***. 2016;25(8):933-938. doi:10.1002/hec.3366
62. van Baal P, Meltzer D, Brouwer W. Future Costs, Fixed Healthcare Budgets, and the Decision Rules of Cost-Effectiveness Analysis. ***Health Econ***. 2016;25(2):237-248. doi:10.1002/hec.3138
63. van Baal PHM, Wong A, Slobbe LCJ, Polder JJ, Brouwer WBF, de Wit GA. Standardizing the Inclusion of Indirect Medical Costs in Economic Evaluations. ***PharmacoEconomics***. 2011;29(3):175-187. doi:10.2165/11586130-000000000-00000
64. van Baal PHM, Feenstra TL, Hoogenveen RT, Ardine de Wit G, Brouwer WBF. Unrelated medical care in life years gained and the cost utility of primary prevention: in search of a 'perfect' cost–utility ratio. ***Health Econ***. 2007;16(4):421-433. doi:10.1002/hec.1181
65. Vergel YB, Sculpher M. Quality-adjusted life years. ***Pract Neurol***. 2008;8(3):175-182. doi:10.1136/pn.2007.140186
66. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. ***Value Health***. 2012;15(6):835-842. doi:10.1016/j.jval.2012.04.014
67. Vreman RA, Naci H, Goettsch WG, et al. Decision Making Under Uncertainty: Comparing Regulatory and Health Technology Assessment Reviews of Medicines in the United States and Europe. ***Clin Pharmacol Ther***. 2020;108(2):350-357. doi:10.1002/cpt.1835

CHAPTER 2

The net benefit of personalised medicine

Vellekoop H, Versteegh M, Huygens S, Ramos IC, Szilberhorn L, Zelei T, Nagy B, Tsiachristas A, Koleva-Kolarova R, Wordsworth S, Rutten-van Mölken M.

The Net Benefit of Personalized Medicine: A Systematic Literature Review and Regression Analysis. Value in Health. 2022;25(8):1428-38.

Abstract

Objectives

Amidst conflicting expectations about the benefits of personalized medicine (PM) and the potentially high implementation costs, we reviewed the available evidence on the cost-effectiveness of PM relative to non-PM.

Methods

We conducted a systematic literature review of economic evaluations of PM and extracted data, including incremental quality-adjusted life-years (ΔQALYs) and incremental costs (Δcosts). ΔQALYs and Δcosts were combined with estimates of national cost-effectiveness thresholds to calculate incremental net monetary benefit (ΔNMB). Regression analyses were performed with these variables as dependent variables and PM intervention characteristics as independent variables. Random intercepts were used to cluster studies according to country.

Results

Of 4,774 studies reviewed, 128 were selected, providing cost-effectiveness data for 279 PM interventions. Most studies were set in the United States (48%) and the United Kingdom (16%) and adopted a healthcare perspective (82%). Cancer treatments (60%) and pharmaceutical interventions (72%) occurred frequently. Prognostic tests (19%) and tests to identify (non)responders (37%) were least and most common, respectively. Industry sponsorship occurred in 32%. Median ΔQALYs, Δcosts, and ΔNMB per individual were 0.03, Int$ 575, and Int$ 18, respectively. We found large heterogeneity in cost-effectiveness. Regression analysis showed that gene therapies were associated with higher ΔQALYs than other interventions. PM interventions for neoplasms brought higher ΔNMB than PM interventions for other conditions. Nonetheless, average ΔNMB in the 'neoplasm' group was found to be negative.

Conclusions

PM brings improvements in health but often at a high cost, resulting in 0 to negative ΔNMB on average. Pricing policies may be needed to reduce the costs of interventions with negative ΔNMB.

Introduction

Personalized medicine (PM), a term often used to describe innovative healthcare interventions that enable improved patient stratification (generally through genetic or genomic testing), may improve health outcomes (eg, by preventing adverse drug reactions in "slow metabolizer" patients) and reduce healthcare costs (eg, by preventing the prescription of treatments to patients who do not benefit from it). Therefore, PM has been subject to high hopes and expectations. Nevertheless, there are concerns about the potentially high costs of (implementing) PM. Among these are worries about the costs of larger-scale gene testing and concerns about the steep pricing of some of the PM interventions that have come onto the market in recent years.[1,2] Although many countries perform economic evaluations to assess the balance between health benefits and costs of individual PM interventions coming onto the market, there is limited knowledge about the average net benefit of PM.

Previous reviews have found that cost savings are relatively rare; most economic evaluations of PM show increased costs and higher health benefits.[3, 4, 5] They have also found large heterogeneity in study methods and cost-effectiveness outcomes, sometimes even across different evaluations of the same intervention.[6, 7, 8] The previous reviews focused on incremental cost-effectiveness ratios (ICERs) or binary "cost-effective yes/no" judgments to assess cost-effectiveness. In this study, we aim to build upon previous research by investigating the net monetary benefit (NMB) of PM interventions instead of their ICERs and by performing regression analyses in which we explore the heterogeneity in the cost-effectiveness of PM interventions.

Methods

Systematic Literature Review

We performed a systematic literature review aiming to identify all published economic evaluations of PM between 2009 and 2019. PM was defined as "a medical model that bases therapeutic choice on the result of gene profiling or aims to correct pathogenic gene mutations," based on a study by Hatz et al.[7] Given the rapid pace of innovation in PM and changes in its costs over time, studies from before the 2009 cutoff were expected to be less insightful about the current value of PM.

The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[9] On March 13, 2019, databases Embase, Google Scholar, Medline Ovid, and Web of Science were searched. An additional search for gray literature was performed on May 16, 2019, and included the Centre for Reviews and Dissemination and EconLit databases and the reimbursement dossier sections on the

websites of the National Institute for Health and Care Excellence and the Institute for Clinical and Economic Review.

A total of 3 groups of search terms were used, combined with the Boolean operator AND: "economic evaluation"; "modelling"; "personalized medicine" (see Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.01.006). Studies were included if they fell within our definition of PM, presented a cost-effectiveness model, provided patient-level cost and quality-adjusted life-year (QALY) outcomes, provided the cost year (so that cost outcomes could be inflated to 2020), extrapolated outcomes beyond short-term clinical trial data, and described an existing (ie, nonhypothetical) intervention. Studies also had to compare a PM intervention with a non-PM intervention, given that PM versus PM comparisons would not allow for the assessment of the added value of PM.

Various data items were collected from the included studies, using a Microsoft Excel–based data extraction form. Study details were recorded (first author, year of publication, country, currency), as well as information about the health technology assessment (HTA) methods used (perspective, time horizon, discount rates, cost-effectiveness threshold). Details about the interventions under evaluation (description of the intervention, description of the comparator, disease class according to the International Classification of Diseases, Tenth Revision [ICD-10]), and cost-effectiveness outcomes (incremental costs [Δcosts], incremental QALYs [ΔQALYs], ICERs) were also captured. If studies evaluated multiple interventions or comparators, all 2-way comparisons between PM and non-PM interventions were recorded.

Assessing Cost-Effectiveness

NMB was used as the measure of cost-effectiveness. Although the ICER measure is widely used to measure cost-effectiveness, NMB is better suited for ranking large numbers of interventions (eg, because of issues around interpreting negative ICERs) and for assessing the magnitude to which an intervention is more (or less) cost-effective than another one.[10] Therefore, the NMB measure was deemed more appropriate for our study.

The incremental NMB (ΔNMB) of each intervention i was calculated with the formula $\Delta NMB_{ij} = \Delta h_{ij} * k_j - \Delta c_{ij}$, where Δh_{ij} = ΔQALYs for intervention i in country j, where k_j = cost-effectiveness threshold in country j, and Δc_{ij} = Δcosts for intervention i in country j. Δcosts were inflated to 2020 prices using country-specific inflation rates and converted to purchasing power parity using conversion factors from the World Bank Global Economic Monitor.[11]

Despite the importance of cost-effectiveness thresholds in cost-effectiveness analysis, limited research has been conducted regarding the appropriate threshold value in each

country. Indeed, in many countries, the standard cost-effectiveness thresholds applied during HTAs are based on little to no data. Part of the reason for the limited research into cost-effectiveness thresholds may be that there exists conceptual disagreement about what the threshold represents and how it should be calculated.[12,13] The 2 main views are that the threshold should reflect (1) society's willingness-to-pay for increases in health (*v*) and (2) the opportunity cost of healthcare spending (*k*). We opted to use *k* thresholds, because we were able to find national estimates for all countries included in our data set (apart from Taiwan), whereas *v* estimates were not available for all countries. In addition, *v* thresholds might not always be appropriate, especially in studies with a healthcare perspective, given that society's willingness to pay for health benefits may not align with available (healthcare) resources. We explicitly chose not to use the thresholds countries have historically used in our base case analysis. This was partly because several countries do not make use of a specific, single threshold and partly because there are likely inconsistencies in how different countries arrived at their thresholds.

Values for national *k* thresholds were mostly taken from a 2016 study by Woods et al,[14] which estimates *k* thresholds for 183 countries. Whenever country-specific studies were available, the estimates from these studies were used. See Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.01.006 for an overview of the national threshold values.

Exploring Heterogeneity

Regression analysis was conducted to explore the heterogeneity in the reported cost-effectiveness of PM in the included studies, aiming to identify characteristics of PM that may be associated with higher (or lower) health benefits, costs, and NMB. Generalized linear mixed models were used, with random intercepts at country level to account for national differences in prescribed HTA methods (such as discount rates for future health benefits and costs), healthcare systems, and epidemiology. Restricted maximum likelihood estimation was applied instead of maximum likelihood estimation to avoid bias in the variance estimation. Analysis of variance (ANOVA) was used to assess the benefit of a random intercept at country level. "Variance explained" was assessed using conditional R2 (following the definitions in Nakagawa and Schielzeth[15]) and compared with the adjusted R2 of a simple linear model with the same specification but without the random intercept. A total of 3 separate models were specified, with ΔQALYs, Δcosts, or ΔNMB as the dependent variables, and all included the same independent variables: "purpose of test," "type of treatment," "gene therapy," "industry sponsorship," and "disease classification."

The independent variable "purpose of test" was based on a previous literature review of economic evaluations of PM, which identified broad categories that tests could be

classed into and found possible differences in (median) cost-effectiveness between the categories.[7] The categories were testing to (1) screen for a disease or a genetic marker in an asymptomatic population (eg, genetic testing for LDL receptor mutations in relatives of patients with familial hypercholesterolemia), (2) gain information about disease prognosis (eg, OncotypeDX), (3) identify likely (non)responders to treatment (eg, testing for NTRK gene fusions so that TRK inhibitors can be provided to cancer patients who test positive), and (4) identify patients who may experience adverse drug reactions (eg, CYP2D6 testing to optimize pharmacotherapy).

The variable "type of treatment" indicates for each intervention whether the treatment is pharmaceutical, nonpharmaceutical, or a combination of both (eg, gene-expression profiling to help diagnose cancers of unclear origin, with subsequent surgical and pharmaceutical treatment). The variable was included based on debate around the affordability and cost-effectiveness of expensive pharmaceuticals in PM.

Literature and initial descriptive analysis showed that genetic therapies tend to be outliers, with sizably higher incremental health benefits and costs than other PM interventions. Therefore, the dichotomous variable "gene therapy" was added to avoid genetic therapies skewing the results for the other variables.

We included "industry sponsorship" as a dichotomous variable to investigate any differences in the reported cost-effectiveness between industry-sponsored and nonindustry-sponsored studies given that previous studies have found that (cost-) effectiveness outcomes tend to be more favorable in industry-sponsored studies than in studies by publicly funded and independent research organisations.[5,16,17]

Finally, we included a dichotomous variable "disease classification," which could take on the value "neoplasm" and "non-neoplasm", depending on whether the intervention was used to treat neoplasms (ie, cancer) or other conditions. This variable accounts for the predictive value of belonging to the largest set of studies found in the literature review. The variable did not further specify the "other" category to avoid multicollinearity with the other predictor variables.

Sensitivity Analysis

In the base case, all studies received equal weight in the regression analysis, only studies with a healthcare perspective were included and ΔNMB was calculated using country-specific *k* thresholds. Sensitivity analyses were performed where (1) studies were weighted according to their quality score in the Tufts Cost-Effectiveness Analysis Registry, (2) ΔNMB was calculated based on the cost-effectiveness threshold stated by the authors of each study, and (3) studies with a societal perspective were also included. We also repeated the base case analysis in a subsample including only interventions

for neoplasms. Finally, because of ambiguity in the definition of genetic therapies, an additional analysis was performed in which the classification of a specific intervention (Spinraza to treat spinal muscular atrophy) was changed from 'genetic' to 'other'.

Results

Study Sample

The systematic search rendered 4,774 articles, whose abstracts were screened. Full-text articles were read for 615 studies, of which 128 met the inclusion criteria and were included for data analysis (see Appendices 3-4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.01.006 for the inclusion flowchart and an overview of included studies). The included articles provided cost-effectiveness outcomes for 279 PM interventions.

Characteristics of Studies, Interventions, and Methods

The distribution of interventions across countries, perspective, disease classes, types of test, types of treatment, and types of funding is presented in Table 1. A total of 23 countries were included in the data set, although most interventions were evaluated in the United States and the United Kingdom (48% and 16%, respectively). All included countries are upper-middle- or high-income economies according to the World Bank country classification,[18] meaning no economic evaluations of PM were identified for lower-middle- or low-income economies. The healthcare perspective was the most common perspective (81%). Most evaluated interventions were in the "neoplasms" category (60%). Pharmaceutical treatments allocated based on markers in the tumor DNA were common in this category. Other frequently occurring interventions were gene assays providing risk assessments regarding the aggressiveness of tumors and screening interventions aiming to identify individuals at risk of developing cancer. Common interventions in the "diseases of the circulatory system" category (19%) were pharmacogenomic testing before anticoagulation therapy and genetic screening to identify patients at risk of various heart conditions. Interventions in the "endocrine, nutritional, or metabolic" diseases category (4%) focused on screening for familial hypercholesterolemia and maturity onset diabetes of the young, whereas interventions in the "mental, behavioral, or neurodevelopmental disorders" category (3%) mostly involved pharmacogenomic testing before starting antidepressants.

Table 1 Descriptive statistics of the included studies

Category	Number of interventions (percent of total)
Country	
Canada	10 (4)
China	15 (5)
Germany	15 (5)
Netherlands	14 (5)
UK	44 (16)
US	135 (48)
Other*	42 (15)
Perspective	
Healthcare	229 (82)
Societal	50 (18)
Disease class	
Diseases of the circulatory system	54 (19)
Endocrine, nutritional or metabolic diseases	11 (4)
Mental, behavioural or neurodevelopmental disorders	9 (3)
Neoplasms	167 (60)
Other**	40 (14)
Purpose of test	
Screening	58 (21)
Info prognosis	54 (19)
Identify responders	103 (37)
Identify adverse drug reactions	64 (23)
Type of treatment	
Pharmaceutical	201 (72)
Non-pharmaceutical	70 (25)
Combination	8 (3)
Gene therapy	
Gene therapy	11 (4)
No gene therapy	268 (96)

Table 1 Descriptive statistics of the included studies (continued)

Category	Number of interventions (percent of total)
Industry sponsorship	
Industry sponsorship	90 (32)
No industry sponsorship	189 (68)

*Included in "Other" are Australia (3 interventions assessed), Austria (3), France (5), Hong Kong (1), Italy (3), Japan (5), Malaysia (2), New Zealand (2), Puerto Rico (1), Singapore (4), Slovenia (1), South Korea (2), Spain (3), Sweden (1), Switzerland (2), Taiwan (3), Thailand (5).

** Included in "Other" are Adverse drug reactions (7), Diseases of the digestive system (3), Diseases of the immune system (6), Diseases of the musculoskeletal system or connective tissue (6), Diseases of the nervous system (5), Diseases of the respiratory system (2), Diseases of the visual system (2).

Notably, 21% of the evaluated interventions fell in the "screening" category, 19% in "info prognosis," 37% in "identify responders," and 23% in "identify adverse drug reactions"; 72% of the evaluated interventions were pharmaceuticals, 25% nonpharmaceutical, and 3% a combination of both. Nonpharmaceutical interventions consisted mostly of gene tests to determine the appropriate screening interval (eg, increased screening frequency for patients at increased risk of hypertrophic cardiomyopathy or colorectal cancer) and gene tests to determine whether surgery is necessary (eg, preventive surgery for patients with BRCA mutations). Only 4% of evaluated interventions were gene therapies, of which 6 were in "neoplasms" and 5 in "non-neoplasms." Moreover, 32% of interventions were evaluated in industry-sponsored studies.

Estimated Cost-Effectiveness

The median amount of ΔQALYs of PM interventions relative to their non-PM comparators was 0.03, whereas the mean was 0.26. As can be seen in Table 2, most ΔQALY values are just above 0, with 0.00 ΔQALY and 0.16 ΔQALY at the 25th and 75th percentile, respectively. Nonetheless, several interventions had larger benefits. Sixteen interventions (6%) rendered > 1 ΔQALY.

Table 2 Quantiles ΔQALYs, Δcosts, ΔNMB

Variable	Min	5%	25%	50%	75%	95%	Max	Mean
ΔQALYs	-0.76	-0.10	0.00	0.03	0.16	1.08	11.8	0.26
Δcosts (2020 Int$)	-34,062	-7,233	-338	575	3,233	282,080	8,095,744	99,777
ΔNMB (2020 Int$)	-7,997,236	-91,832	-2,665	18	3,538	21,615	406,277	-77,072

Δcost indicates incremental cost; ΔNMB, incremental net monetary benefit; ΔQALY, incremental quality-adjusted life-year; Max, maximum; Min, minimum.

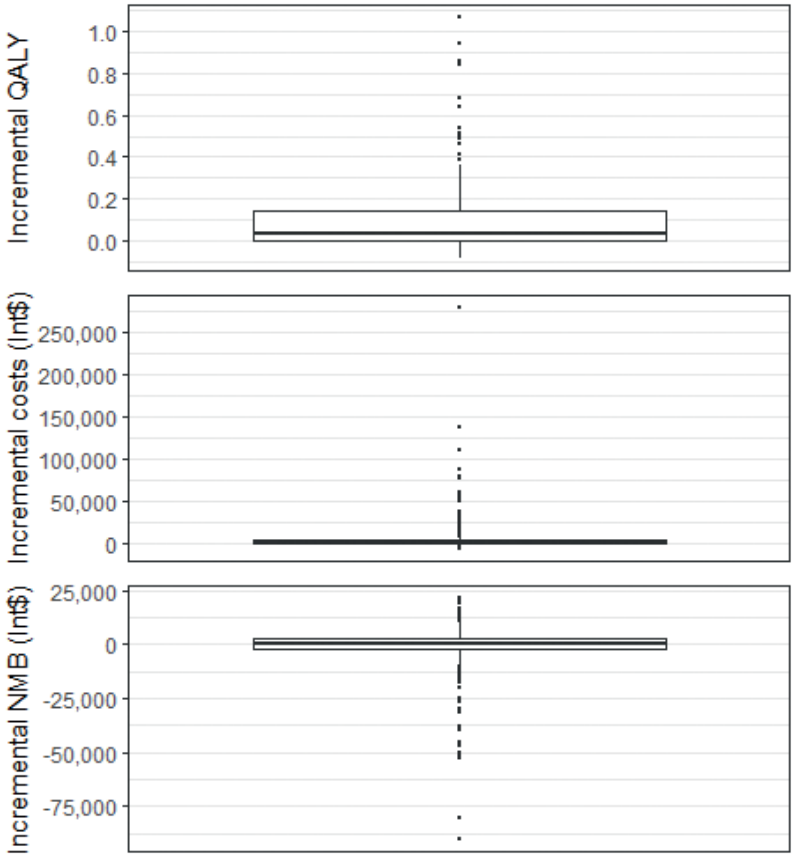


Figure 1 Boxplots ΔQALY, Δcost, ΔNMB

For each boxplot, the bottom and top 5% of the distribution were excluded from the figure. ΔQALY indicates incremental quality-adjusted life-year; Δcost, incremental cost; ΔNMB, incremental NMB.

Median Δcosts were Int$ 575, whereas mean Δcosts were Int$ 99,777. Figure 1 shows that a small number of interventions have notably higher Δcosts than the rest.

Median ΔNMB across the included interventions was Int$ 18, and mean ΔNMB was Int$ −77,072. ΔNMB centers around 0, with a value of Int$ −2,665 at the first quantile and Int$ 3,538 at the third quantile. As can be seen in Figure 1, extreme negative values are more common than extreme positive values for ΔNMB. Nonetheless, there are also some positive outliers, with a maximum ΔNMB of Int$ 406,277. Details of the interventions that are among the top 5% in terms of ΔNMB are presented in Table 3.[19,20,21,22,23,24,25,26,27,28,29,30]

Table 3 Details for interventions among top 5% of ΔNMB

ΔNMB	Intervention	Comparator	Country	Perspective	Disease class	Purpose of test	Type of treatment	Gene therapy	Industry sponsor	ΔQALY	Δcosts	Ref
406,277	CAR T-cell therapy (tisagenlecleucel) for paediatric patients with relapsed or refractory B-cell acute lymphoblastic leukemia	Clorafibine	US	Health-care	Neoplasms	Identify (non-) responders	Pharm	yes	no	7.18	346,163	[19]
265,848	The addition of gene-expression profiling to usual care to identify the tissue of origin and guide therapy choice in adults with metastatic cancer of uncertain origin	Therapy choice based on usual care (including IHC, blood tests, physical examination)	US	Health-care	Neoplasms	Identify (non-) responders	Pharm	no	yes	2.65	11,863	[20]
208,766	CAR T-cell therapy (tisagenlecleucel) for paediatric patients with relapsed or refractory B-cell acute lymphoblastic leukemia	Blinatumomab; clofarabine, cyclophosphamide, and etoposide combination therapy; and clofarabine monotherapy	US	Health-care	Neoplasms	Identify (non-) responders	Pharm	yes	yes	5.17	333,033	[21]

Table 3 Details for interventions among top 5% of ΔNMB (continued)

ΔNMB	Intervention	Comparator	Country	Perspective	Disease class	Purpose of test	Type of treatment	Gene therapy	Industry sponsor	ΔQALY	Δcosts	Ref
95,175	Gene profiling (mutational load-based) in patients with non-dysplastic Barrett's esophagus. Only patients with high risk score receive ablative therapy	Periodic endoscopic surveillance, ablative therapy when high grade dysplasia or esophageal adenocarcinoma are detected	US	Health-care	Diseases of the digestive system	Info prognosis	Non-pharm	no	yes	0.86	-5,260	[22]
78,312	Gene-expression profiling (Mammaprint™) in patients with node-negative, estrogen receptor-positive breast cancer. Only patients with high risk score receive adjuvant chemotherapy	All patients receive adjuvant chemotherapy	NL	Health-care	Neoplasms	Info prognosis	Pharm	no	no	1.20	-12,170	[23]
62,705	Gene profiling (mutational load-based) in patients with nondysplastic Barrett's esophagus. All patients who test positive for mutational load (high- or low-risk) receive ablative therapy	Periodic endoscopic surveillance, ablative therapy when high grade dysplasia or esophageal adenocarcinoma is found	US	Health-care	Diseases of the digestive system	Info prognosis	Non-pharm	no	yes	0.54	-6,115	[22]

Table 3 Details for interventions among top 5% of ΔNMB (continued)

ΔNMB	Intervention	Comparator	Country	Perspective	Disease class	Purpose of test	Type of treatment	Gene therapy	Industry sponsor	ΔQALY	Δcosts	Ref
48,518	Gene profiling in children with neonatal diabetes. Patients switch from insulin to sylfonylurea when mutations in the KCNJ11 or ABCC8 genes are found	Insulin treatment for all patients	US	Societal	Endocrine, nutritional or metabolic diseases	Identify (non-) responders	Pharm	no	no	0.32	-14,983	[24]
41,885	Gene profiling to identify long-QT syndrome in relatives of index patients. Relatives who test positive are treated with β-blockers	β-blockers for all first-degree relatives	US	Societal	Diseases of the circulatory system	Screen asymptomatic	Pharm	no	no	0.41	1,081	[25]
37,277	Gene profiling (CPGx™) in treatment-resistant patients with major depressive disorder to guide therapy choice	Therapy choice based on treatment history, physical examination, lab tests that are part of usual care	US	Societal	Mental, behavioural or neurodevelopmental disorders	Identify (non-) responders	Pharm	no	yes	0.32	-4,161	[26]
35,604	Gene profiling (assessing IL-28B genotype) of patients with chronic hepatitis C to identify patients who would benefit from the addition of protease inhibitors to standard therapy	Pegylated interferon and ribavirin (standard therapy) for all	US	Societal	Certain infectious or parasitic diseases	Identify (non-) responders	Pharm	no	no	0.54	20,986	[27]

Table 3 Details for interventions among top 5% of ΔNMB (continued)

ΔNMB	Intervention	Comparator	Country	Perspective	Disease class	Purpose of test	Type of treatment	Gene therapy	Industry sponsor	ΔQALY	Δcosts	Ref
35,424	Gene profiling (CYP2C19) in patients with acute coronary syndrome. Patients with CYP2C19*2 received prasugrel, all other patients clopidogrel	All patients receive prasugrel	US	Healthcare	Diseases of the circulatory system	Identify ADR	Pharm	no	no	0.01	-34,062	[28]
26,891	Gene profiling (CYP2C19) in patients with acute coronary syndrome. Patients with CYP2C19*2 received prasugrel, all other patients clopidogrel	All patients receive clopidogrel	US	Healthcare	Diseases of the circulatory system	Identify ADR	Pharm	no	no	0.12	-14,629	[28]
24,294	Gene profiling (loss of heterozygosity-based) in patients with low-grade oral dysplasia. Patients with a low risk score return for follow-up every 5 years, patients with an intermediate risk score every 2 years. Patients with a high risk score are referred for surgery	All patients return for follow-ups every 6 months. Patients are referred for surgery when cancer is found	CA	Healthcare	Diseases of the digestive system	Info prognosis	Non-pharm	no	no	0.64	-7,467	[29]

Table 3 Details for interventions among top 5% of ΔNMB (continued)

ΔNMB	Intervention	Comparator	Country	Perspec-tive	Disease class	Purpose of test	Type of treatment	Gene therapy	Industry sponsor	ΔQALY	Δcosts	Ref
24,052	Gene profiling (IDGx) in patients with major depressive disorder who are treatment-naïve or whose depression is inade-quately controlled to guide therapy choice	Therapy choice based on usual care, in which non-responders to treatment iteratively try alternative options	US	Societal	Mental, behavioural or neurode-velopmental disorders	Identify (non-) responders	Pharm	no	yes	0.17	-6,237	[30]

Δcost indicates incremental cost; ΔNMB, incremental net monetary benefit; ΔQALY, incremental quality-adjusted life-year; ADR, adverse drug reaction; CA, Canada; CAR, chimeric antigen receptor; IHC, immunohistochemistry; IL-28B, interleukin-28B; NL, The Netherlands; Pharm, Pharmaceutical; Ref, reference; US, United States.

In Figure 2, median ΔNMB per country (ie, based on all PM interventions evaluated in the country) is plotted against the national *k* threshold and the (median) author-reported threshold. We see that median ΔNMB of PM is generally close to 0, regardless of the national threshold. This implies that any QALY gains of PM interventions tend to be counterbalanced by their costs to the healthcare system.

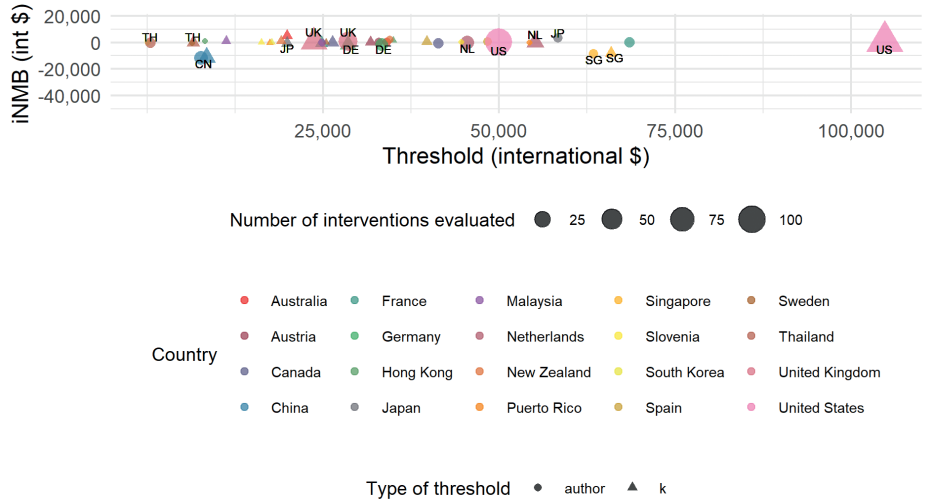


Figure 2 Median ΔNMB plotted against cost-effectiveness threshold, per country

ΔNMB indicates incremental net monetary benefit; CN, China; DE, Germany; JP, Japan; NL, Netherlands; SG, Singapore; TH, Thailand; UK, United Kingdom; US, United States.

Heterogeneity in Cost-Effectiveness

Our data set is inherently heterogeneous, because of the inclusion of studies from many countries and many authors, each with slight methodological differences. This variation is reflected in wide confidence intervals (CIs) (Table 4). Nevertheless, the mean values predicted by the models are close to the observed values.

In Table 4, the regression results for the model with ΔQALYs of PM versus non-PM as the dependent variable are presented first. The regression coefficient for the gene therapy variable is the only coefficient for which 0 is not included in the 95% CI. The coefficient of 3.22 is much larger than any of the other coefficients, suggesting large QALY gains for gene therapies. This may be because most of the gene therapies included in the review focus on early onset conditions with high morbidity and mortality. The conditional R2 of the mixed model for QALYs was 0.47 compared with an adjusted R2 of 0.46 of a simple linear model without random intercepts at country level. ANOVA therefore showed that the use of a random intercept did not improve goodness of fit.

For Δcosts of PM versus non-PM, "gene therapy" and "non-neoplasms" have a 95% CI that does not cross 0. The regression coefficient of 1,179,540 for "gene therapy" implies that, on average, the Δcost for gene therapies is 1,179,540 higher than for PM interventions that are not gene therapies. Similarly, the regression coefficient of 386,325 implies that for PM interventions in "non-neoplasm" the Δcost is 386,325 higher than for interventions in "neoplasm." Conditional R2 of the mixed model for costs was 0.66 compared with an adjusted R2 of 0.33 of a simple linear model without random intercepts at country level. Hence, ANOVA showed that using a random intercept improved goodness of fit of the model.

Finally, in the ΔNMB model of PM versus non-PM, the regression coefficients for "gene therapy" has a negative regression coefficient, with the 95% CI again not crossing 0. The coefficient suggests that, on average, gene therapies bring Int$ 868,759 less net benefit compared with non-PM interventions, despite offering higher QALY gains. This implies that the costs associated to gene therapies are higher than the monetary value of the QALY gains, leading to a net loss. The coefficient for "non-neoplasm" also does not cross 0 and implies that PM interventions for conditions other than neoplasms render Int$ 380,950 less ΔNMB than PM interventions in "neoplasm". In line with the findings from the regression analysis, the average ΔNMB in the observed data was Int$ −1,287,417 (median Int$ −343,379) for gene therapies, whereas it was Int$ −27,394 (median Int$ 49) for the other interventions. Average ΔNMB was Int$ −1,161 (median Int$ −426) for neoplasms and Int$ −190,260 (median Int$ 164) for other interventions. Conditional R2 of the mixed model for NMB was 0.53 compared with an adjusted R2 of 0.23 of a simple linear model without random intercepts at country level. ANOVA showed the model goodness of fit improved by using a random intercept.

Sensitivity Analysis

Results for the sensitivity analysis can be found in Appendix 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.01.006. The size and direction of regression coefficients tend to be similar between the analyses using base case assumptions and analyses using the alternative assumptions. In all analyses, gene therapies are associated with significantly higher QALY gains, and PM interventions for neoplasms are associated with lower costs and higher ΔNMB. Within the neoplasm subsample (n = 167), conclusions are similar to those reported in the main analysis, although the 95% CI for "gene therapy" now crosses 0. Reclassification of genetic therapy Spinraza to nongene therapy increased the coefficient of "gene therapy" for QALYs to 4.5. After the reclassification, the CI for "gene therapy" crossed 0 in the cost model. The coefficient for "gene therapy" in the NMB model was Int$ 777,987. The average ΔNMB for gene therapies remained negative after the reclassification and was Int$ −356,016.

Table 4 Regression results (N = 229)

Variable	Category	Regression coefficient	95% confidence interval	t-value
Dependent variable: ΔQALYs				
Intercept		0.02	[-0.27; 0.32]	0.16
Purpose of test*	Info prognosis	0.10	[-0.27; 0.47]	0.55
	Identify responders	0.22	[-0.13; 0.56]	1.23
	Identify ADR	-0.25	[-0.63; 0.14]	-1.25
Type of treatment†	Pharmaceutical	0.02	[-0.27; 0.31]	0.14
	Combi	0.12	[-0.58; 0.82]	0.33
Gene therapy	Gene therapy	**3.22**	[2.67; 3.75]	12.0
Sponsorship	Industry	-0.15	[-0.38; 0.08]	-1.31
Disease classification‡	Non-neoplasm	0.25	[-0.05; 0.55]	1.64
Dependent variable: Δcosts				
Intercept		-163,055	[-474,256; 148,145]	-1.04
Purpose of test*	Info prognosis	137,639	[-190,299; 465,578]	0.83
	Identify responders	239,144	[-85,158; 563,445]	1.46
	Identify ADR	-183,128	[-524,414; 158,156]	-1.06
Type of treatment†	Pharmaceutical	7,225	[-254,119; 268,571]	0.06
	Combi	-96,374	[-687,345; 494,595]	-0.32
Gene therapy	Gene therapy	**1,179,540**	[732,527; 1,626,554]	5.25
Sponsorship	Industry	-292,337	[-292,337; 107,539]	-0.92
Disease classification‡	Non-neoplasm	**386,325**	[122,244; 650,406]	2.91
Dependent variable: ΔNMB				
Intercept		152,210	[-144,118; 448,539]	1.02
Purpose of test*	Info prognosis	-126,431	[-445,368; 192,505]	-0.78
	Identify responders	-221,146	[-535,623; 93,331]	-1.39
	Identify ADR	176,913	[-156,155; 509,981]	1.06
Type of treatment†	Pharmaceutical	3,479	[-251,022; 257,981]	0.03
	Combi	99,634	[-475,896; 675,166]	0.34
Gene therapy	Gene therapy	**-868,759**	[-1,307,289; -430,228]	-3.94
Sponsorship	Industry	92,109	[-103,308; 287,526]	0.94
Disease classification‡	Non-neoplasm	-380,949	[-638,867; -123,032	-2.94

For values in bold, the 95% confidence interval does not cross 0.

Δcost indicates incremental cost; ΔNMB, incremental net monetary benefit; ΔQALY, incremental quality-adjusted life-year; ADR, adverse drug reaction.

* Reference category is “screening.”

† Reference category is “nonpharmaceutical interventions.”

‡ Reference category is “neoplasms.”

Discussion

Interpretation Results

The median and mean ΔQALYs found in this study (0.03 and 0.26, respectively) are comparable with the QALY gains found by a literature review of cost-utility analyses for all types of healthcare, which identified a median QALY increase of 0.06 (mean 0.31).[31] We found that 6% of interventions had a QALY increase of > 1, while this was 8% in the previous study. This suggests that the health benefits of PM tend to be similar to (or possibly slightly lower than) the health benefits of other (new) healthcare interventions.

The median ΔNMB of PM interventions was close to 0 (Int$ 18), implying that the health benefits rendered by PM interventions are counterbalanced by the increased costs associated with the interventions. It could be that PM interventions are associated with higher costs than non-PM interventions because of the additional testing needed for the improved stratification in PM. Nevertheless, the costs of testing were often not—or only partially—considered in the studies included in our review.

Gene therapies were found to offer high QALY gains in all analyses. Gene therapies were also found to be associated with high costs and had an average ΔNMB of Int$ −1,287,417 (or Int$ −356,016 after reclassifying Spinraza to nongene therapy). It has been argued that the QALY insufficiently captures the value of gene therapies and that additional value elements, such as the value of a cure (ie, nonhealth-related welfare benefits, such as being able to do more future-planning), should also be considered. Although there may indeed be benefits beyond QALYs to being completely cured of a condition, our findings suggest that, on average, the monetary value of these additional benefits would have to be Int$ −1,287,417 per patient for the ΔNMB to no longer be negative, which arguably is implausibly high, especially in the light of a recent study by Reed et al[32] in which much lower figures are found for the "value of hope". The genetic therapies included in our analysis were treatments for spinal muscle atrophy (Spinraza, Zolgensma) and loss of vision because of inherited retinal dystrophy (Luxturna), and CAR-T cell therapies. Spinraza, an antisense oligonucleotide, is referred to as a "genetic therapy," yet oligonucleotides are not classified as a "gene therapy" by the Food and Drug Administration or as advanced therapy medicinal products by European Medicines Agency.[33] After reclassifying Spinraza to a nongene therapy, the outcomes for the ΔQALY and Δcost models were largely the same. However, the coefficient for "gene therapy" in the ΔNMB model changed from a negative to a positive value. Given that the "gene therapy" coefficient is sensitive to single data points, it appears that data are too scarce to draw definitive conclusions on the NMB of gene therapies.

PM interventions in neoplasms were shown to have lower Δcosts and higher ΔNMB. Indeed, average ΔNMB was higher in the "neoplasm" group than in the "non-neoplasm" group (Int$ −1,161 vs. Int$ −190,260). However, median ΔNMB is more similar across the groups, with median ΔNMB even being a little lower for neoplasms (Int$ −426 for "neoplasm" and Int$ 164 for "non-neoplasm"). This suggests large heterogeneity in the "non-neoplasm" group and precludes a simple explanation of why interventions for neoplasms might render more ΔNMB. The number of interventions per "non-neoplasm" condition are too low to perform condition-specific regression analysis. Different research methods may be needed to further explore the heterogeneity in the benefit that interventions for "other" conditions bring. The addition of a random intercept at country level improved goodness of fit for the Δcost and ΔNMB models but not for the ΔQALY model, suggesting that the estimation of QALYs is less affected by differences between countries than the estimation of costs and NMB.

The CIs around the regression coefficients other than "gene therapy" and "neoplasm" are wide and cross 0 for all abovementioned categories. Therefore, no definitive conclusions regarding the association between the different categories and dependent variable ΔNMB can be drawn. Nonetheless, we offer possible explanations for our findings below.

First, a potential explanation for the positive coefficient for "identify ADR" could be that many of the interventions included in this category aim to better stratify patients to existing treatments instead of to new treatments (eg, by offering reduced warfarin dosing to patients with gene mutations that are associated with increased sensitivity to warfarin). Conversely, many interventions in the "identify responders" category stratify toward new treatments. New treatments are generally still patented and may be costly, especially when the target population is small, which has a negative effect on the interventions' ΔNMB. Nonetheless, in the future, the benefit from new innovations may increase. Decisions on whether to invest in products that currently bring more ΔNMB to the healthcare system or in products that may bring more ΔNMB in the future depend on value judgments on the extent to which current versus future (uncertain) QALYs should be prioritized.

The regression coefficient for pharmaceutical interventions was positive in the ΔQALY and costs models and negative in the ΔNMB model. This could mean that although PM pharmaceuticals have higher health gains than nonpharmaceuticals, PM pharmaceuticals come at a higher cost than nonpharmaceuticals, causing lower net value (ΔNMB). The higher cost for PM pharmaceuticals compared with PM nonpharmaceuticals could be because of high prices charged by pharmaceutical companies. Nevertheless, the higher cost for pharmaceuticals could also be caused by other factors, such as the nature of the treated diseases.

The positive coefficient for "industry sponsorship" in the ΔNMB could mean that industry-sponsored studies are more likely to have positive cost-effectiveness outcomes, which is in line with previous studies.[5,16,17] This might be because industry-sponsored studies are often for interventions about to be evaluated for reimbursement and focused on a single promising intervention, whereas nonindustry-sponsored studies may take a wider approach and include several interventions in the evaluation, some of which perhaps less promising. Additionally, authors of industry-sponsored reimbursement studies may have an incentive to limit the included costs (eg, include only testing costs for patients who tested positive as opposed to testing costs for the entire population that received testing) or otherwise make model assumptions to improve cost-effectiveness outcomes. Finally, the issue of publication bias, whereby only studies with positive results are published, may be more prominent for industry- than for nonindustry-sponsored research.

Strengths and Limitations

Previous studies investigating the cost-effectiveness of PM focused on ICERs and descriptive analyses.[3,4,5,6,7,8] Our study expands on these studies, by focusing on NMB as opposed to ICERs, given that NMB has been argued to be more appropriate for comparing large numbers of interventions. Our study also adds to the literature by presenting regression analyses, for each of the variables ΔQALYs, Δcosts, and ΔNMB separately. Our study builds on the work of Hatz et al[7] in particular, by incorporating the types of test they identified into our analysis.

Our definition of PM focuses on genetic and genomic test-treatment combinations. We acknowledge that alternative interpretations of "personalized medicine" exist. Some understand the "personalized" aspect as an increased focus on patient preferences, and some include decision making based on patients' phenotypes in their definition of PM.[34,35] Our decision to focus on patients' genotypes was based on a study by Schleidgen et al,[34] in which a systematic literature review was conducted to understand how the term "personalized medicine" is used in scientific practice. The study finds that most scientific articles on PM focus on the use of gene information in medical decision making. Indeed, Schleidgen et al[34] argue that decision making based on phenotype (eg, weight, sex) and patient preferences is part of traditional healthcare. Including these in the definition of PM would blur the distinction between PM and traditional healthcare. Nonetheless, we acknowledge that this study provides an incomplete overview of the net benefit of PM to those who hold a wider definition of PM.

Although many authors expressed the incremental health benefit of a test-treatment combination "per patient who tested positive," some expressed the health benefit "per patient tested." The health benefit "per patient tested" may be different from the health benefit "per patient who tested positive." For example, when genetic testing has to be

applied to a large number of patients to detect a few patients with a rare mutation that determines eligibility for a particular treatment, the average health benefit per patient tested (and treated) may be diluted compared to the health benefit per patient who tested positive. Nonetheless, if outcomes across patients with different test results are expected to be similar, health benefit "per patient tested" and "per patient who tested positive" may be similar. For example, patients with certain variants in the VKORC1 and CYP2C9 genes require lower warfarin dosing. Once warfarin dosages have been adjusted appropriately, health outcomes for patients who test positive for these gene variants may be similar to health outcomes for patients who test negative for the gene variants and receive normal dosing. Studies of cascade screening often add the health benefit of family members that are identified after an initial index patient tests positive to the health benefit of the index patient, resulting in a single QALY value "per index patient."[36] A single value combining the health gains of multiple patients is likely higher than the benefit "per patient who tested positive." Therefore, there may be some inconsistency in what is captured in the ΔQALY values that we extracted from the selected studies.

Indeed, our sample has a high level of heterogeneity as a result of the wide scope of our literature search and inconsistency in methods across studies. Therefore, the results of our regression analysis are uncertain. Although we attempted to account for country-level differences by clustering studies according to country, more streamlining of methods across studies may be needed to reduce uncertainty around the cost-effectiveness of PM.

Finally, the countries included in our review were mostly high-income countries (and a few higher-middle-income). Therefore, our results do not necessarily apply to low- and (lower-)middle-income countries. Although this was caused by the limited economic evaluations of PM currently available for lower-income countries, we acknowledge the importance of more evidence generation for lower-income countries, given that insights into which interventions offer the highest added value are of critical importance in settings with highly constrained healthcare budgets.

Implications

Our study results show modest health benefits for PM versus non-PM interventions and a median ΔNMB of close to 0. The available evidence therefore seems to contrast the high anticipations that have surrounded PM over recent years. Nonetheless, there are several PM interventions with (very) positive ΔNMB. It appears that the term "personalized medicine" may be too general because it conceals sizable differences in the net benefit of different PM interventions. A more precise division into subcategories of PM may be needed to uncover the most promising areas for further investment.

Despite the general tendency of PM interventions toward a ΔNMB of 0, we identified various interventions with (very) negative ΔNMB, among which several gene therapies. National and international pricing policies may be needed to reduce the costs of these interventions and ensure that societies are not faced with negative added value when implementing them.

Conclusions

This study has provided evidence that PM leads to additional health gains compared with non-PM but its costs tend to result in 0 to negative ΔNMB. Gene therapies offer high QALY gains and render negative net monetary benefit on average, though data scarcity prohibits drawing firm conclusions on their added value. For PM interventions with negative ΔNMB, the benefit to society may be increased if ways can be found to reduce costs.

References

1. O'Sullivan B.P., Orenstein D.M., Milla C.E. Pricing for orphan drugs: will the market bear what society cannot?. *JAMA*. 2013; **310**: 1343-1344
2. Experts in Chronic Myeloid Leukemia. The price of drugs for chronic myeloid leukemia (CML) is a reflection of the unsustainable prices of cancer drugs: from the perspective of a large group of CML experts. *Blood*. 2013; **121**: 4439-4442
3. D'Andrea E., Marzuillo C., Pelone F., De Vito C., Villari P. Genetic testing and economic evaluations: a systematic review of the literature. *Epidemiol Prev*. 2015; **39**: 45-50
4. Plumpton C.O., Roberts D., Pirmohamed M., Hughes D.A. A systematic review of economic evaluations of pharmacogenetic testing for prevention of adverse drug reactions. *Pharmacoeconomics*. 2016; **34**: 771-793
5. Berm E.J.J., Looff M.d.e., Wilffert B., et al. Economic evaluations of pharmacogenetic and pharmacogenomic screening tests: a systematic review. Second update of the literature. *PLoS One*. 2016; **11**e0146262
6. Faruque F., Noh H., Hussain A., Neuberger E., Onukwugha E. Economic value of pharmacogenetic testing for cancer drugs with clinically relevant drug-gene associations: a systematic literature review. *J Manag Care Spec Pharm*. 2019; **25**: 260-271
7. Hatz M.H.M., Schremser K., Rogowski W.H. Is individualized medicine more cost-effective? A systematic review. *Pharmacoeconomics*. 2014; **32**: 443-455
8. Beaulieu M., de Denus S., Lachaine J. Systematic review of pharmacoeconomic studies of pharmacogenomic tests. *Pharmacogenomics*. 2010; **11**: 1573-1590
9. Liberati A., Altman D.G., Tetzlaff J., et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009; **339**: b2700
10. Paulden M. Calculating and interpreting ICERs and net benefit. *Pharmacoeconomics*. 2020; **38**: 785-807
11. Global economic monitor. The World Bank. https://datacatalog.worldbank.org/dataset/global-economic-monitor. Date accessed: December 3, 2020
12. Vallejo-Torres L., García-Lorenzo B., Castilla I., et al. On the estimation of the cost-effectiveness threshold: why, what, how?. *Value Health*. 2016; **19**: 558-566
13. Cameron D., Ubels J., Norström F. On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. *Glob Health Action*. 2018; **11**1447828
14. 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**: 929-935
15. Nakagawa S., Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol Evol*. 2013; **4**: 133-142
16. Lundh A., Lexchin J., Mintzes B., Schroll J.B., Bero L. Industry sponsorship and research outcome: systematic review with meta-analysis. *Intensive Care Med*. 2018; **44**: 1603-1612
17. Kim D.D., Silver M.C., Kunst N., Cohen J.T., Ollendorf D.A., Neumann P.J. Perspective and costing in cost-effectiveness analysis, 1974-2018. *Pharmacoeconomics*. 2020; **38**: 1135-1145
18. The world by income. The World Bank.

https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Date accessed: August 12, 2021

19. Whittington M.D., McQueen R.B., Ollendorf D.A., et al. Long-term survival and value of chimeric antigen receptor T-cell therapy for pediatric patients with relapsed or refractory leukemia. *JAMA Pediatr*. 2018; **172**: 1161-1168
20. Hornberger J., Degtiar I., Gutierrez H., et al. Cost-effectiveness of gene-expression profiling for tumor-site origin. *Value Health*. 2013; **16**: 46-56
21. Lin J.K., Lerman B.J., Barnes J.I., et al. Cost effectiveness of chimeric antigen receptor T-cell therapy in relapsed or refractory pediatric B-cell acute lymphoblastic leukemia. *J Clin Oncol*. 2018; **36**: 3192-3202

22. Das A., Callenberg K.M., Styn M.A., Jackson S.A. Endoscopic ablation is a cost-effective cancer preventative therapy in patients with Barrett's esophagus who have elevated genomic instability. ***Endosc Int Open***. 2016; **4**: E549-E559
23. Retel V.P., Joore M.A., Knauer M., Linn S.C., Hauptmann M., van Harten W.H. Cost-effectiveness of the 70-gene signature versus St. Gallen guidelines and Adjuvant Online for early breast cancer. ***Eur J Cancer***. 2010; **46**: 1382-1391
24. Greeley S.A.W., John P.M., Winn A.N., et al. The cost-effectiveness of personalized genetic medicine: the case of genetic testing in neonatal diabetes. ***Diabetes Care***. 2011; **34**: 622-627
25. Perez M.V., Kumarasamy N.A., Owens D.K., Wang P.J., Hlatky M.A. Cost-effectiveness of genetic testing in family members of patients with long-qt syndrome. ***Circ Cardiovasc Qual Outcomes***. 2011; **4**: 76-84
26. Hornberger J., Li Q., Quinn B. Cost-effectiveness of combinatorial pharmacogenomic testing for treatment-resistant major depressive disorder patients. ***Am J Manag Care***. 2015; **21**: e357-e365
27. Liu S., Cipriano L.E., Holodniy M., Owens D.K., Goldhaber-Fiebert J.D. New protease inhibitors for the treatment of chronic hepatitis C: a cost-effectiveness analysis. ***Ann Intern Med***. 2012; **156**: 279-290
28. Lala A., Berger J.S., Sharma G., Hochman J.S., Scott Braithwaite R., Ladapo J.A. Genetic testing in patients with acute coronary syndrome undergoing percutaneous coronary intervention: a cost-effectiveness analysis. ***J Thromb Haemost***. 2013; **11**: 81-91
29. Cromwell I., Regier D.A., Peacock S.J., Poh C.F. Cost-effectiveness analysis of using loss of heterozygosity to manage premalignant oral dysplasia in British Columbia, Canada. ***Oncologist***. 2016; **21**: 1099-1106
30. Groessl E.J., Tally S.R., Hillery N., Maciel A., Garces J.A. Cost-effectiveness of a pharmacogenetic test to guide treatment for major depressive disorder. ***J Manag Care Spec Pharm***. 2018; **24**: 726-734
31. Wisløff T., Hagen G., Hamidi V., Movik E., Klemp M., Olsen J.A. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. ***Pharmacoeconomics***. 2014; **32**: 367-375
32. Reed S.D., Yang J.C., Gonzalez J.M., Johnson F.R. Quantifying value of hope. ***Value Health***. 2021; **24**: 1511-1519
33. Chilcott EM, Muiruri EW, Hirst TC, Yáñez-Muñoz RJ. Systematic review and meta-analysis determining the benefits of in vivo genetic therapy in spinal muscular atrophy rodent models [published online October 6, 2021]. ***Gene Ther***. https://doi.org/10.1038/s41434-021-00292-4.
34. Schleidgen S., Klingler C., Bertram T., Rogowski W.H., Marckmann G. What is personalized medicine: sharpening a vague term based on a systematic literature review. ***BMC Med Ethics***. 2013; **14**: 55
35. Pokorska-Bocci A., Stewart A., Sagoo G.S., Hall A., Kroese M., Burton H. "Personalized medicine": what's in a name?. ***Per Med***. 2014; **11**: 197-210
36. Vellekoop H., Huygens S., Versteegh M., et al. Guidance for the harmonisation and improvement of economic evaluations of personalised medicine. ***Pharmacoeconomics***. 2021; **39**: 771-788

Supplementary materials

Supplementary materials can be found at https://doi.org/10.1016/j.jval.2022.01.006

Contents

- Appendix 1: Search strategy for the systematic literature review
- Appendix 2: Overview of *k* threshold per country
- Appendix 3: Flowchart systematic literature review
- Appendix 4: References of studies included for data analysis
- Appendix 5: Regression results under different assumptions

CHAPTER 3

Guidance for the harmonisation and improvement of economic evaluations of personalised medicine

Vellekoop H, Huygens S, Versteegh M, Szilberhorn L, Zelei T, Nagy B, Koleva-Kolarova R, Tsiachristas A, Wordsworth S, Rutten-van Mölken M, HEcoPerMed Consortium.

Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine. PharmacoEconomics. 2021;39(7):771-88.

Abstract

Objective

The objective of this study was to develop guidance contributing to improved consistency and quality in economic evaluations of personalised medicine (PM), given current ambiguity about how to measure the value of PM as well as considerable variation in the methodology and reporting in economic evaluations of PM.

Methods

A targeted literature review of methodological papers was performed for an overview of modelling challenges in PM. Expert interviews were held to discuss best modelling practice. A systematic literature review of economic evaluations of PM was conducted to gain insight into current modelling practice. The findings were synthesised and used to develop a set of draft recommendations. The draft recommendations were discussed at a stakeholder workshop and subsequently finalised.

Results

Twenty-two methodological papers were identified. Some argued that the challenges in modelling PM can be addressed within existing methodological frameworks, others disagreed. Eighteen experts were interviewed. They believed large uncertainty to be a key concern. Out of 195 economic evaluations of PM identified, 56% addressed none of the identified modelling challenges. A set of 23 recommendations was developed. Eight recommendations focus on the modelling of test-treatment pathways. The use of non-randomised controlled trial data is discouraged but several recommendations are provided in case randomised controlled trial data are unavailable. The parameterisation of structural uncertainty is recommended. Other recommendations consider perspective and discounting; premature survival data; additional value elements; patient and clinician compliance; and managed entry agreements.

Conclusions

This study provides a comprehensive list of recommendations to modellers of PM and to evaluators and reviewers of PM models.

Introduction

Personalised medicine (PM) aims to better stratify patients to enable more targeted healthcare. Personalised medicine, often used interchangeably with related terms such as precision medicine, stratified medicine and individualised medicine, has the potential to offer cost savings (e.g. due to therapies being prescribed only to those likely to benefit) and improved health outcomes (e.g. due to dose adjustment in those at high risk of adverse events). Personalised medicine has made great strides especially in oncology, where an increasing number of therapies is used to target specific genetic alterations [1,2,3,4].

However, high prices are often charged for PM interventions [5, 6]. Manufacturers of PM have argued that their price setting is justified, as PM has benefits that are not captured in conventional health economic frameworks. Reimbursement authorities, however, have been hesitant to accept these claims. Additionally, they have in several cases rejected PM interventions for a lack of convincing evidence on treatment effectiveness [7]. Manufacturers have argued that the small patient populations that are inherent to PM (due to high levels of stratification) hamper the collection of data that meets current standards for health technology assessment (HTA) and have suggested that the solution lies in updating HTA approaches. These (and other) issues have caused ambiguity about how to measure the value of PM, as reflected in the lack of national and international guidance on the evaluation of PM and considerable variation in the methodology and reporting in existing economic evaluations of PM [8, 9].

This study aimed to develop recommendations to health economic modellers in the field of PM and to those evaluating or reviewing PM models, to improve the consistency and quality across different health economic models of PM. The study was conducted within the context of the European Commission-funded Health Economics for Personalised Medicine (HEcoPerMed) Coordination and Support Action, which aims to identify optimal health economic modelling and payment strategies for evaluating and financing PM.

Working Definition of PM

The European Commission has defined PM as “a medical model using characterisation of individuals’ phenotypes and genotypes (e.g. molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention” [10]. While this definition is comprehensive, it was deemed too broad for the purpose of this study, as many existing interventions—for which no new modelling challenges arise—could be argued to fall under this definition. Indeed, it has been argued that a new term such as PM should be used to describe new innovations that

are distinctive from well-established practices [11]. Additionally, while it is acknowledged that some interpret PM as improved stratification based on personal preferences and/or behaviour, it is mostly understood to be informed by biological information [11]. In line with this reasoning, the following definition of PM was adopted for this paper: "A medical model that bases therapeutic choice on the result of gene profiling or aims to correct pathogenic gene mutations".

Note that the decision *not* to treat is a therapeutic choice as much as the decision to treat is, and thus gene profiling that results in 'watchful monitoring' or 'no further medical treatment' is included in this definition. Furthermore, the profiling of gene mutations does not always require sequencing of the genes themselves. Profiling of gene mutations may be done at the functional level, for example, using protein expression tests.

Methods

Several research methods were used to develop the final recommendations.

Targeted Literature Review of Methodological Papers

First, a targeted review was conducted to identify methodological studies discussing challenges in the health economic modelling of PM. Methodological studies were identified through targeted searches on PubMed and Google Scholar (using search terms related to "methodology", "economic evaluation" and "personalised medicine"), through the scientific network of the authors, and by snowballing through the reference lists of identified studies. Only studies in English were considered for inclusion, with no limits on the publication year. The issues that were reported in the studies were extracted and combined into a list of methodological challenges. The list was used to develop questions for expert interviews.

Expert Interviews

The subsequent expert interviews aimed to gain a more current and in-depth understanding of the modelling challenges and to receive expert opinion on what constitutes good modelling practice. Eighteen experts were interviewed between November 2019 and February 2020, with all interviews lasting between 1 and 1.5 h. Interview candidates were identified through screening the authors of the studies identified in the targeted literature review. Interviewees were selected only if they had experience in developing, assessing and/or using economic models of PM. Experts were interviewed separately (though on one occasion two experts were interviewed together) and interviews were conducted in a semi-structured manner. Interview responses were summarised per topic and reviewed for accuracy by the relevant interviewee. A data extraction table for the systematic literature review described below was developed

based on the challenges identified in the targeted literature review and refined using the expert responses.

Systematic Literature Review of Economic Evaluations

A systematic literature review of economic evaluations of PM was conducted to identify methods that have been used to address the identified modelling challenges. The review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. On 13 March, 2019, the following databases were searched: Embase, Google Scholar, MELDINE Ovid (similar to PubMed), and Web of Science. On 16 May, 2019, an additional search was performed in the Centre for Reviews and Dissemination and EconLit databases, as well as in the reimbursement dossier sections on the websites of the National Institute for Health and Care Excellence and the Institute for Clinical and Economic Review. The search strategy consisted of three groups of search terms, combined with the Boolean operator AND: "economic evaluation"; "modelling"; "personalised medicine". See the Electronic Supplementary Material (ESM) for the full search strategy. The search was limited to studies published in English from January 2009 onward. Given the rapid pace of innovation in the field of PM as well as improvements in HTA methodology over time, studies from before the 2009 cut-off were expected to be less relevant. Indeed, previous reviews have found increases in the quality and quantity of economic evaluations of PM since 2009 [4, 12]. Studies were only included if they presented a cost-effectiveness model incorporating final outcomes (i.e. life-years [LYs] or quality-adjusted life-years [QALYs]); met our working definition of PM; extrapolated outcomes beyond available clinical trial data; and described an existing (i.e. non-hypothetical) intervention. During the data extraction phase, the modelling methods used in the included studies were extracted if they addressed any of the challenges listed in the data extraction table.

Development of the Guidance

Finally, a detailed description of the challenges that may occur when modelling PM, as well as accompanying recommendations, were written based on the list of challenges identified in the targeted literature review, the summarised expert responses and the findings from the systematic literature review. The guidance was discussed at a stakeholder workshop and subsequently finalised. The stakeholder workshop took place in September 2020 and comprised around 30 participants, including health economists, representatives of national healthcare payers, and representatives from the pharmaceutical and diagnostics industry.

Results

Targeted Literature Review of Methodological Papers

Twenty-two methodological papers were identified through the targeted literature review (see the ESM for the full list). The papers discussed the challenges in the modelling and/or evaluation of interventions in a range of PM-related fields, such as genomic technologies [13], advanced therapeutic medicinal products [14] and gene therapies [15]. Several papers stated that the evaluation of PM is possible within existing HTA frameworks [15,16,17,18], though these are potentially not optimal [17]. One paper argued that changes to existing HTA methods and processes will likely be necessary as the field of PM develops [19], another proposed methodological adjustment [20]. The interview template that was developed based on the challenges reported in the methodological papers can be found in the ESM.

Expert Interviews

Several experts voiced similar opinions. Among these were the opinions that PM should be subject to the same methodological framework as other interventions to ensure consistency, and that additional value elements should generally not be included in economic evaluations until more research has been conducted and the consequences of including them are better understood. Many experts stated that a high degree of uncertainty is a key issue in PM. Uncertainty regarding treatment effectiveness (due to small study sample sizes) was considered especially prominent, as well as uncertainty whether the often-complex clinical pathways in PM are accurately reflected in economic models. Opinions differed regarding the desirability and necessity of using observational data to estimate treatment effectiveness. While some experts reasoned that economic evaluations based on observational data are unavoidable and/or acceptable, many experts argued that observational data are generally insufficient to inform pricing and reimbursement decisions and trial data should be demanded from manufacturers. Summarised responses, organised by topic, can be found in the ESM.

Systematic Literature Review of Economic Evaluations

The systematic literature review identified 7,787 studies through database searching and seven additional studies by searching the reimbursement dossiers sections of the websites of the National Institute for Health and Care Excellence and Institute for Clinical and Economic Review. A total of 4,774 individual studies remained after de-duplication [21], of which 195 were included after full-text screening (Fig. 1). The annual number of published economic evaluations of PM gradually increased over the search period, from ten studies in 2009 to 30 studies in 2018. More than half of the included studies (56%, n = 109) did not address any of the identified modelling challenges, while 33%

(*n* = 64) of studies incorporated one or two aspects, and 11% (*n* = 22) incorporated three or more aspects. See the ESM for a list of identified studies, including an overview of the aspects incorporated in each study.

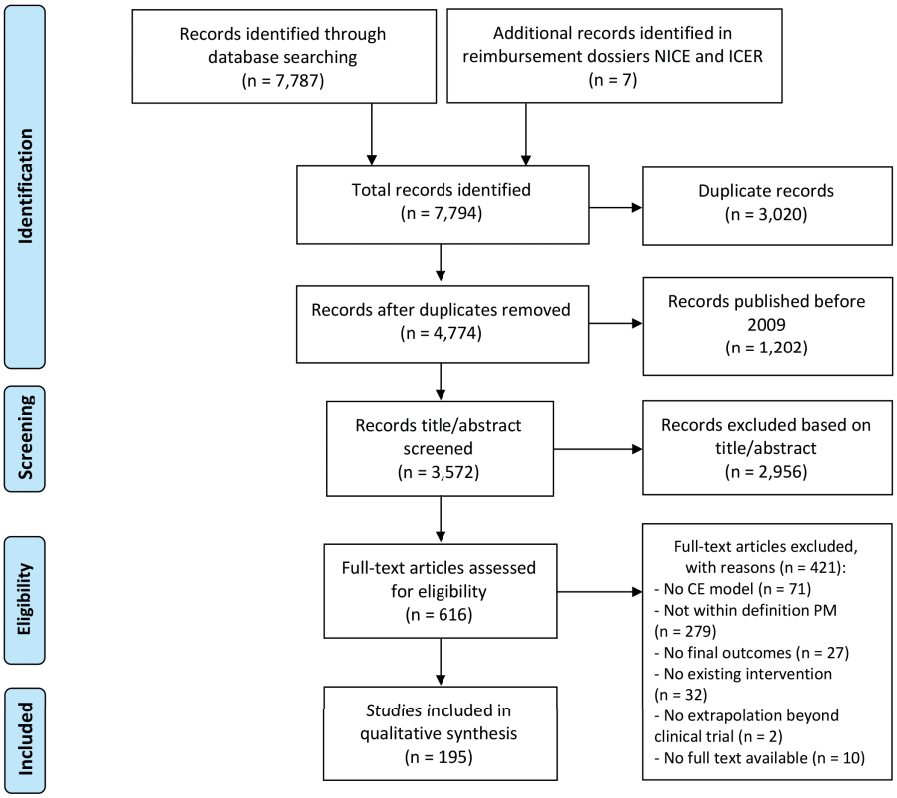


Figure 1 Flow chart of the systematic literature review.

CE cost-effectiveness, *ICER* Institute for Clinical and Economic Review, *NICE* National Institute for Health and Care Excellence, *PM* personalised medicine

As shown in Table 1, the most addressed aspects were patients' treatment compliance (19%, *n* = 38) and uptake of testing (11%, *n* = 22), outcomes for relatives of index patients (14%, *n* = 27), and the conditionality of test sequences and results (9%, *n* = 18). The studies incorporating patient compliance and outcomes for relatives of index patients mostly concerned genetic testing for disease risk factors and subsequent preventive treatment. Uptake and compliance were generally incorporated as the proportions of patients undergoing testing and adhering to treatment. Outcomes for relatives of index patients were mostly combined with outcomes for index patients into single incremental cost and QALY figures, but sometimes reported separately. Studies incorporating conditionality of

test sequences and results used a variety of assumptions, including varying sensitivity and specificity of tests across age groups, varying predictive values across ethnic groups, and conditional probabilities for test results (i.e. the probability of a positive/negative test result was dependent on whether a previous test had been positive/negative).

Table 1 Results data extraction table

Topic	Aspect	Number (%) of studies addressing each aspect *n = 195*
Discounting	Was there a deviation from standard discount rates for reasons particularly relevant to PM?	3 (2%)
Test-treatment combinations	Was the conditionality of test sequences and results incorporated?	18 (9%)
	Were waiting times incorporated?	10 (5%)
	Were outcomes for relatives of index patients incorporated?	27 (14%)
Effectiveness data	Were methods used to account for potential bias in non-RCT data?	5 (3%)
Extrapolating outcomes for interventions with a portion of long-term survivors	Were methods used to account for potential bias in the extrapolation of outcomes for interventions with a portion of long-term survivors?	11 (6%)
Additional elements of value	Were value elements beyond the QALY incorporated?	12 (6%)
Incorporating compliance	Was patients' uptake of testing incorporated?	22 (11%)
	Was patients' treatment compliance incorporated?	38 (19%)
	Was clinicians' compliance to protocols and guidelines incorporated?	10 (5%)
Uncertainty analysis	Were methods used to combine experts' judgements into a point estimate plus probability distribution?	0 (0%)
	Was uncertainty analysis particularly relevant to PM performed?	7 (4%)
Managed entry agreements	Were the conditions of a managed entry agreement incorporated?	8 (4%)

Studies incorporating additional value elements (6%, *n* = 12) mostly considered the psychological impact individuals experience when finding out about increased cancer risk through a genetic test (e.g. reduced uncertainty and/or increased anxiety), though two studies focused on psychological effects related to preventive surgery that may be performed based on the results of such a genetic test (e.g. reduced anxiety or worsened body image). These psychological effects were incorporated by applying a utility increase or decrease. Studies that applied methods to account for potential bias in the extrapolation of outcomes for interventions with a proportion of long-term survivors (6%, *n* = 11) were mostly economic evaluations of chimeric antigen receptor (CAR) T-cell therapies. The methods included mixture cure models [22] and the use of hazard ratios to adjust general population mortality for long-term survivors of the condition in question.

Clinician compliance to protocols and guidelines (*n* = 10) and waiting times for patients (*n* = 10) were incorporated in 5% of studies. Relevant uncertainty analysis (4%, *n* = 7) evaluated, for example, the uncertainty around future cost reductions of genetic testing and the possible consequences of inconclusive genetic test results. The conditions of a managed entry agreement were incorporated in 4% (*n* = 8) of the studies, and methods to adjust for potential bias in non-RCT data in 3% (*n* = 5) of the studies. One identified method to adjust for potential bias in non-RCT data was to exclude patients with certain characteristics from the comparator cohort to increase comparability between the comparator and intervention cohorts. Another method consisted of the estimation of survival curves per treatment given and the subsequent weighting of the survival curves based on the expected distribution of treatments in the target population. In 2% (*n* = 3) of studies, lowered discount rates were applied for both costs and benefits. These studies evaluated gene therapies and cited the National Institute for Health and Care Excellence Methods Guide stating that a discount rate of 1.5% (instead of 3.5%) may be considered when a treatment restores individuals experiencing conditions with high mortality and/or morbidity to sustained (near-)full health [23]. No studies parameterised expert judgement into a probability distribution.

Recommendations

The main outcome of this study is the following set of recommendations, informed by the results presented in the previous three sections. The numbered recommendations are introduced by an explanation. See the ESM to view the list of recommendations.

Perspective and Discounting

Some have argued that countries' standard HTA approaches may not always capture the full value of PM [20, 24, 25]. For example, for evaluating gene therapies that cure children from conditions with high mortality and morbidity, a societal perspective might be more

appropriate than a payer perspective. That is, in addition to measuring patients' QALY gains, a societal perspective could capture the lifetime reduction in the use of informal care that might result from a child receiving a cure, as well as the potentially increased quality of life of carers.

Similarly, some have suggested a lower discount rate for health outcomes in PM with high upfront costs and long-term benefits, such as curative gene therapies and large-scale genetic screening programmes, to place a higher value on future benefits [14]. (A similar effect can be achieved with hyperbolic discounting, in which the discount rate is gradually reduced over time [26].)

However, allowing the assumptions regarding perspective and discounting to be different for (some) PM would hamper comparability of cost-effectiveness results across interventions and implicitly favour PM over other interventions, especially given the fact that many non-PM interventions also have wider societal and/or long-term benefits.

1. For economic evaluations of PM, use the standard perspective as recommended by national HTA guidelines in the base case.
2. For economic evaluations of PM, use the standard discount rates as recommended by national HTA guidelines in the base case.

Test-Treatment Pathways

Given that stratification is a key tenet in PM, testing plays an important role in the clinical pathway. A number of topics to consider when modelling tests are discussed below [27].

A single patient may be subject to a range of tests. There may be various options for combining the different tests. The tests may be performed in parallel or sequentially; when they are performed in sequence, choices may have to be made regarding the order in which the tests are performed. The diagnostic strategies chosen may vary across subgroups of the patient population. Additionally, some tests can be applied at different points in the treatment pathway (e.g. genome sequencing). As a result of these factors, modellers may be faced with many possible pathways. When prioritising which options to include in the model, a key consideration should be the extent to which they are relevant given the decision-making context.

3. Identify all relevant test-treatment pathways and justify why the pathways included in the model were selected.

When a test is used to stratify patients into subgroups that are eligible and non-eligible for a specific treatment, the consequences of using the test may affect the cost effectiveness

of the test-treatment combination and should be explicitly considered. First, the costs of testing should be included in the economic evaluation of the treatment. When a new treatment requires the introduction of a new test (or the provision of an existing test to a wider target population), allocating 100% of the additional testing costs to the treatment under evaluation may be appropriate. This might seem unfair to "first movers", i.e. the first pharmaceutical or medical devices companies that require a specific test to be able to identify the right patients for their products, while the same test may later be used for other medical products. However, it is an accurate reflection of the decision problem at hand: the new treatment cannot be implemented in clinical practice without also implementing said test and thus their combined cost effectiveness should be assessed. When the stratification for the new treatment can be done with a test that is already part of current practice, none or only a proportion of the testing costs may be allocated to the new treatment. The specific assumptions made regarding cost allocation may vary according to the budgeting and/or reimbursement arrangements in the decision-making context at hand. Note that total testing cost for all tested patients should be incorporated in the model, as opposed to the testing cost only for patients with a positive test result.

Furthermore, adverse events due to the testing procedure may reduce quality of life and increase mortality rates (e.g. a collapsed lung due to a lung biopsy). Additionally, the test results may stimulate further testing and treatment (e.g. because of secondary findings), affecting final health outcomes and costs. Additionally, there will be false-positive and false-negative patients among those tested, who may be facing poorer health outcomes, potentially leading to additional costs.

4. When a treatment requires the use of a test to stratify patients, include in the model the (downstream) costs and health outcomes of testing for both individuals who test (false-)positive and individuals who test (false-) negative.

The rates of false-positives and false-negatives are largely determined by the diagnostic accuracy of the testing technology used. The diagnostic accuracy of the technology is likely to vary according to the (subgroups of) the patient population in which the technology is applied and may change over time.

5. Ensure that the data used to estimate the diagnostic accuracy of a testing technology are appropriate to the patient population in the model.

Tests may have a continuous outcome, thus cut-off values must be set to determine the result (e.g. low, medium or high risk; positive or negative). Different cut-off values

may be in use for the same test. For example, in the USA, pembrolizumab is indicated for patients with non-small cell lung cancer with high programmed death-ligand 1 (PD-L1) expression, which was first defined as PD-L1 expression in > 50% of tumour cells but later as PD-L1 expression in > 1% of tumour cells [28].

6. When different cut-off values are in use to determine test results, clearly define the cut-off value assumed in the base case. Investigate the effect of alternative cut-off values on cost-effectiveness results using a sensitivity analysis.

When various tests are modelled in sequence, their results may be correlated. See Box 1, for an example of the use of conditional probabilities to model test results.

7. When multiple tests are modelled in sequence, consider the interdependence between test results.

Box 1 Considering interdependence between test results [29]

The cost effectiveness of "expanded reflex" testing versus standard testing for HER2 mutations was estimated for breast cancer patients in the United States. Patients with HER2-positive breast cancer have a worse prognosis than HER2-negative patients but can be treated with trastuzumab, a monoclonal antibody specifically targeting HER2. HER2 status can be determined using either immunohistochemistry (IHC), which measures HER2 overexpression, or fluorescence in situ hybridisation (FISH), which measures HER2 gene amplification. In the standard approach, HER2 status is assessed using either IHC (80%) or FISH (20%). With expanded reflex testing, patients testing negative on either test subsequently receive the alternative test for confirmation. Interdependence between the two test results was incorporated in the model by using conditional probabilities obtained from a published systematic review. In the standard testing approach, only the outcome probabilities for either test in isolation are used. In the expanded reflex testing approach, the same outcome probabilities are used for the first test, while for the second test the outcome probabilities conditional on a negative outcome on the first test are assumed. Expanded reflex testing is estimated to render a QALY gain of 0.037 per patient treated compared to the standard approach, which is driven by a lower rate of (untreated) false negatives. Costs increase by $1455, resulting in an incremental cost-effectiveness ratio (ICER) of $39,721 per QALY gained. In the base case, it is estimated that 2.27% of patients are reclassified as HER2-positive after the second test. This percentage was varied between 1-8% in scenario analysis, rendering a maximum ICER of $47,110 and minimum ICER of $35,575.

Patients presenting with symptoms generally do not receive treatment instantaneously. They may be faced with periods of waiting between presenting with symptoms and the decision to get tested, between the decision to get tested and testing, between testing and getting results, and between test results and the start of treatment. These waiting periods may impact outcomes, especially in conditions with high short-term morbidity/mortality (see Box 2 for an example).

8. If there is a notable risk of increased morbidity or mortality as a result of waiting periods, incorporate in the model the costs and health outcomes due to the waiting periods.

Box 2 Incorporating waiting times [30]

The cost effectiveness of tisagenlecleucel, a CAR T-cell therapy, was estimated for paediatric patients with relapsed/refractory B-cell acute lymphoblastic leukaemia in the United States. Patients undergo leukapheresis so that their T-cells can be harvested and used to prepare the tisagenlecleucel infusion. During the waiting period between leukapheresis and receiving the infusion, patients are at risk of mortality. This is captured in the model by a short-term decision tree, in which the first node reflects the probability that patients receive the tisagenlecleucel infusion. The chance node has the branches "continue with the infusion" and "die before receiving the infusion". An additional branch captures the probability that patients "discontinue before infusion because of adverse events or manufacturing failure". In a scenario analysis where the model did not incorporate waiting time by starting at infusion instead of at leukapheresis, the incremental costs increased from $329,498 to $454,892 and incremental QALYs increased from 7.2 years to 9.1 years. The increase in incremental costs and QALYs is explained by the exclusion of patients who do not receive the tisagenlecleucel infusion and therefore have both lower treatment cost and shorter quality-adjusted life expectancy. Excluding waiting time from the model slightly increased the ICER, from $46,000 per QALY to $50,000 per QALY.

While some testing technologies are produced by a single provider and standardised, other tests are performed using local laboratory resources. There may be variation in testing costs between commercially developed test kits and local laboratory tests, as well as across laboratories.

9. Confirm that the assumed testing costs are accurate in the setting of interest and consider possible variation in costs across laboratories.

When an inheritable pathogenic mutation is identified in a patient, relatives are also at risk and may be offered genetic counselling and testing (e.g. in familial hypercholesterolaemia, BRCA-positive breast cancer). Focussing the economic evaluation of a test-treatment combination for inheritable mutations only on the index patients may offer an incomplete reflection of the clinical reality.

10. If relatives of index patients become eligible for genetic testing when the index patients test positive for a specific genetic marker, include the costs and health outcomes of testing relatives in the economic evaluation of the index patients.

Effectiveness Data

More stratification of patients leads to increasingly small patient (sub)groups, complicating the generation of (sufficiently) statistically powered data on treatment effectiveness through traditional randomised controlled trials (RCTs). Several alternative trial designs have been developed, including basket trials [31, 32], umbrella trials [31], n-of-1 trials [33] and adaptive trials [34]. While some of these alternative designs still allow for controlled studies, many are single-arm designs.

Doubts have been raised whether foregoing an RCT was justified in all of the non-RCT dossiers that have so far been submitted to regulators such as the European Medicines Agency and the US Food and Drug Administration [35]. Nonetheless, there appears to be increasing acceptance of non-RCT evidence among regulatory agencies [36].

The lack of RCT evidence poses challenges to health economic modelling as it complicates evidence synthesis (e.g. incomplete networks in a network meta-analysis) and increases the uncertainty around cost-effectiveness results. “Conditional reimbursement” or “coverage with evidence development” programmes, in which additional data are collected after market approval, have raised concerns. First, it has been questioned whether they are able to provide unbiased estimates of relative effectiveness when relying on observational data [37]. Concerns also exist about the feasibility of withdrawing medicines that were granted reimbursement once further evidence does not demonstrate their (cost-)effectiveness [38, 39].

11. Where possible, use effectiveness data from trials with two (or more) alternative treatment strategies.

In tandem with the increasing acceptance of non-RCT evidence, there has been a rise in the use of evidence from early trials [40]. However, the relationship between final outcomes and the surrogate outcome measures often used in early trials is not always

well established. For example, out of 93 cancer drug indications for which accelerated approval was granted by the US Food and Drug Administration based on surrogate outcomes (such as response rate or progression-free survival), confirmatory trials reported improved overall survival for only 19 (20%) [41].

12. When surrogate outcomes are used to estimate final outcomes, specify which data sources were used to estimate the relationship between surrogate and final outcomes and justify any assumptions made about the relationship.

When only data from single-arm studies are available, external data could be used to construct a control arm. However, treatment effectiveness may improve over time, owing to treatment-related factors such as dosing optimisation, improvements in the standard of care or external factors such as improvements in general population health. Historical data may, therefore, underestimate effectiveness in the control group, leading to the overestimation of the new treatment's effectiveness.

13. When effectiveness of the comparator is estimated using external data, account for a possible time trend in effectiveness.

Increasingly many treatments that target specific genetic markers are coming onto the market. If the genetic marker affects disease prognosis, combining data sources with a different prevalence of the genetic marker may be inappropriate for estimating comparative effectiveness. Nonetheless, the effectiveness estimate can potentially be adjusted when the prognostic value of the marker and the prevalence of the genetic marker across the different data sources are known.

For example, two TRK inhibitors (larotrectinib and entrectinib) have recently come onto the market for tumours with NTRK gene fusions. The effectiveness of both treatments was assessed using single-arm trials, meaning that external data are necessary to be able to construct a comparator arm reflecting the standard of care. However, while the trial population for the TRK inhibitors consisted of only NTRK-positive patients, the available data on comparator effectiveness stem from trials in which most patients were NTRK negative, owing to the low prevalence of NTRK fusions in many types of cancer. Preliminary evidence suggests that NTRK-positive patients have a worse prognosis [42]. The treatment effectiveness estimated on populations with a large share of NTRK-negative patients may therefore provide a biased estimate of the treatment effectiveness for NTRK-positive patients and may have to be adjusted.

14. When effectiveness of the comparator for patients with a specific genetic marker is estimated using external data, account for the prognostic value of the genetic marker and differences in its prevalence across the different data sources.

When new gene tests are developed to allocate patients to existing treatments, it is unlikely new RCTs will be performed for the existing treatments in each of the subgroups introduced by the new test. Instead, data on treatment effectiveness in the subgroups may come from genotype-phenotype studies, in which associations between genetic markers and clinical outcomes are investigated. Estimating a causal relationship between genotype and phenotype tends to be highly complicated, owing to, among other things, the potential for gene–gene and gene–environment interactions, heterogeneity in genetic markers (e.g. hundreds of different BRCA mutations have been found in patients with breast cancer [43]) and heterogeneity in clinical symptoms [44]. Additionally, it may be difficult to identify appropriate controls, to link genetic data to data on clinical outcomes [45] and to get a large enough sample size to meet statistical significance requirements. As a result, details about the relationship between a genetic marker and clinical outcomes are often uncertain or unknown. For example, while it has been shown that patients with acute coronary syndrome who carry a loss-of-function polymorphism of cytochrome P450 2C19 (CYP2C19) experience more thrombotic events when treated with clopidogrel, uncertainty remains regarding the degree of association between carrier status and thrombotic events [46]. Similarly, the relationship between the level of HER2 expression in patients with breast cancer and the extent to which progression-free survival is reduced is not fully known [47].

15. Specify which data sources were used to estimate the association between the genetic marker(s) of interest and clinical outcomes and justify any assumptions made about the association.

Extrapolating Survival

So far, innovations in PM have mostly been in disease areas with high mortality, such as oncology and rare severe genetic disorders. An accurate estimation of the effect that these interventions have on patient mortality is key to assessing their cost effectiveness. Trials generally provide only short-term data, bringing the need for modelling to estimate survival beyond the trial period. While long-term survival can sometimes be estimated using the surrogate outcomes measured in the trial (see recommendation 12), it is more common to extrapolate from the short-term mortality captured in the trial. The choice of survival model is often informed by assessing the statistical fit to the data. However, models with a good fit to short-term trial data do not always provide plausible

predictions regarding long-term outcomes [48]. Expert judgement may be used to evaluate the plausibility of the estimated survival.

16. When extrapolating survival data beyond the study period, use expert opinion alongside statistical fit to choose the survival model.

In several PM interventions, patients might be considered "cured" when they experience long-term survival. This may apply to, for instance, oncology patients experiencing sustained complete remission after receiving targeted therapy, or to patients with early-onset diseases with high mortality who respond to a gene therapy. However, even if patients experience long-term survival, they may be faced with poorer long-term health outcomes than the general population. For example, Janssen-Heijnen et al. showed that for several cancer types (stomach cancer, non-small cell lung cancer, stage II or III breast cancer, prostate cancer, Hodgkin lymphoma), patients who had been "cured" and survived between 10 and 20 years after diagnosis still had poorer survival than the general population [49]. They hypothesised that this could be due to late recurrences, secondary tumours or comorbidities associated with cancer risk factors [49]. Assuming (age- and sex-specific) general population mortality for "cured" patients may therefore be inappropriate (see Box 3 for an example).

17. When extrapolating survival data beyond the study period, account for any excess mortality and morbidity among long-term survivors.

Box 3 Accounting for excess mortality among long-term survivors [50]

In a report by the Institute for Clinical and Economic Review on the cost effectiveness of CAR T-cell therapies in B-cell malignancies, tisagenlecleucel and axicabtagene ciloleucel were evaluated, respectively for paediatric patients with relapsed/refractory B-cell acute lymphoblastic leukaemia and adults with relapsed/refractory aggressive B-cell lymphoma. Survival was extrapolated from the available trial data, owing to the limited observation period in the clinical trials (15.4 months in the ZUMA-1 trial, 3.7 months in the JULIET trial). It was presumed that patients who survived to the point at which the survival curve flattened (i.e. the slope equalled zero) were among the long-term survivors. A knot was introduced at this point in the survival curve (at five years), after which patients were assumed to be subject to general population mortality, adjusted for excess mortality observed among long-term survivors of B-cell malignancies. Based on published evidence, a standardised mortality ratio of 9.1 was used to reflect excess mortality among long-term survivors in the paediatric leukaemia cohort, while a standardised mortality ratio of 1 was used for the adult B-cell lymphoma cohort. The latter proved influential on model outcomes in one-way sensitivity analysis. The base case assumption of a standardised mortality ratio of 1 rendered an ICER of $136,078 for axicabtagene ciloleucel versus chemotherapy, which increased to $176,491 when a maximum standardised mortality ratio of 3.4 was assumed. An additional scenario analysis was conducted, in which it was assumed that only 80% of those alive and responding to treatment at five years would be long-term survivors in the paediatric leukaemia cohort and 95% of those alive and responding to treatment at five years would be long-term survivors in the adults B-cell lymphoma cohort. The remaining patients (20% and 5%, respectively) were assumed to die at five years. This increased the ICER for tisagenlecleucel versus clofarabine from $45,871 to $53,195 and the ICER for axicabtagene ciloleucel versus chemotherapy from $136,078 to $140,443.

Additional Elements of Value

It has been argued that the QALY insufficiently captures the full value interventions may have. The ISPOR Value Assessment Framework Special Task Force identified a list of additional value elements to be included in a cost-effectiveness analysis, including scientific spill-overs, equity, real option value, value of hope, severity of disease, insurance value, fear of contagion and reduction in uncertainty [20]. A related concept that has been suggested is “personal utility”, which is generally used either to describe the value of knowledge (e.g. knowledge of a test outcome) or as an umbrella term for the non-health outcomes that individuals might value [51, 52]. Patients may indeed value outcomes of healthcare beyond increased health. Diagnostic information, for example, may allow patients to make better life decisions or cause psychological effects such as alleviated (or increased) anxiety [53]. However, the suggested additional value elements raise several concerns.

First, it remains unclear how to define, measure and value many of the identified elements, partly owing to their conceptual ambiguity. There appears to be a risk of double counting, both within the set of elements (e.g. severity of disease can be argued to be part of equity; insurance value is likely strongly correlated with severity of disease, given that the value of being insured against the consequences of falling ill is higher when the diseases covered are more severe) and between the elements and the QALY (some of the "additional" elements may already be captured in preference-based quality-of-life assessments, such as the reduction in uncertainty).

Additionally, there appears to be a focus on positive value elements, while negative value elements may be equally relevant. For instance, the value of hope that patients might experience prior to treatment (e.g. the hope that they are among the long-term responders to treatment) may be (partly) offset by the disutility due to dashed hope once treatment outcomes are known.

It is important to be aware that including additional value elements may alter decision making, at the expense of length and quality of life. For example, once additional value elements are included, intervention A with high health benefits might be deemed less cost effective than intervention B with medium health benefits but many additional elements of value. When choosing to adopt intervention B instead of intervention A, we are implicitly trading off length and quality of life against other value elements (see Table 2 for a stylised example).

Table 2 Stylised example of the consequence of including a value of hope

Treatment	Cost-effectiveness threshold (λ)	ΔQALY	Δcost	Value of hope	Incremental net monetary benefit*
Treatment 1 (standard approach)	$50,000	2	$80,000	-	$20,000
Treatment 2	$50,000	2.5	$80,000	0	$45,000
Treatment 1 (including value of hope)	$50,000	2	$80,000	$30,000	$50,000**

* Incremental net monetary benefit = (λ * ΔQALY) – Δcost

** Incremental QALYs are higher for treatment 2. However, treatment 1 offers "hope", while treatment 2 does not. In this example, placing a monetary value on "hope" increases the incremental net monetary benefit of treatment 1 from $20,000 to $50,000. This may lead to the prioritisation of treatment 1 over treatment 2, despite treatment 2 offering higher QALY gains.

Individuals may indeed be willing to trade off certain elements of value against length and quality of their own life [54], conveying consumer value. However, this does not necessarily mean that people are willing to make such trade-offs between, for example, hope and health in others, casting doubt on whether such value elements should be prioritised at the national level. Indeed, it may be debated whether healthcare payers such as national governments should pay for all elements that bring value to individual patients, especially given that there likely is significant variation across patients in their valuations of specific elements, as well as within patients depending on the time they are asked. For example, risk-averse patients might not experience any value of hope and risk-loving patients may only experience a value of hope for a short time, after which dashed hope might be experienced.

If additional value elements were to be included in economic evaluations, this should be done for all interventions, not just PM, to ensure consistency and comparability across studies. Indeed, the suggested value elements may be relevant outside of PM. For example, while a patient with a family history of breast cancer likely experiences relief (i.e. gains personal utility) when they find out they are BRCA negative (PM), personal utility may equally be gained by a couple wanting to get pregnant when they find out they have no fertility issues (non-PM).

Note that the threshold value against which the cost effectiveness of the intervention is judged may have to be adjusted to account for the additional value elements. For example, if additional value elements are included in a sensitivity analysis, the resulting cost-effectiveness outcomes may have to be judged against a different threshold than the outcomes in the base-case analysis. The rationale for this depends on whether the threshold is viewed as a supply-side estimate (i.e. the opportunity cost of healthcare spending, or the marginal productivity of the healthcare system) or a demand-side estimate (i.e. the societal willingness to pay for improvements in health) [55]. In the former, the threshold changes when additional value elements are included because the opportunity cost of spending now includes not only health forgone but other benefits forgone as well. In the latter, the threshold changes because the societal willingness to pay for only health outcomes may be different from the willingness to pay for health and non-health outcomes combined.

18. Only include elements of value recommended by national HTA guidelines in the base case. If additional elements of value are included in a sensitivity analysis, ensure possible elements of negative value are equally considered and included for both the intervention and the comparator.

Incorporating Compliance

While health economic models are used to simulate the clinical reality, clinical reality is not always optimal. Depending on the decision context, modellers may choose to model a healthcare intervention at its optimal implementation level or at a level of implementation that is closer to reality (or both). It is important to be transparent about the extent to which the model reflects optimal implementation.

In PM, a significant cause of suboptimal implementation may be imperfect compliance owing to the perceptions and preferences of patients and clinicians regarding PM. For example, unwillingness to find out about risk-increasing gene mutations may hamper patients' uptake of genetic testing. Similarly, limited understanding of risk/probabilities may lessen patients' compliance to therapeutic plans based on genetic testing. In addition, clinicians may not be fully compliant to protocols and guidelines because of a limited knowledge of genetics, or they may already initiate treatment in rapidly deteriorating patients if the waiting time for test results is too long.

Compliance is likely affected by the perceived probability of disease (this applies to testing only), the severity of disease and/or the type of treatment (e.g. preventive or curative). For example, in the study described in Box 4, the uptake of genetic testing for the risk of breast and ovarian cancer is markedly lower for people aged under 30 years than for people aged 30 years and above. Similarly, compliance to genetic testing for cardiovascular disease risk and subsequent preventive measures may be lower than compliance to genetic testing of tumours and subsequent cancer treatment.

When incorporating compliance, note that an adjustment of the effectiveness estimate might not be necessary for pragmatic trials, where data reflect real-world compliance. Furthermore, note that reduced compliance does not automatically mean that intervention costs are lower (e.g. medicines may have been dispensed but not taken). Finally, patient compliance may vary considerably between for example socioeconomic, geographic and age groups. Clinician compliance might also vary according to the societal group their patients belong to.

19. Include parameters reflecting patient and clinician compliance in economic evaluations for decision makers who require cost-effectiveness results under realistic circumstances.

20. When including patient and clinician compliance in economic evaluations, confirm that the assumed compliance is accurate in the setting of interest and consider possible variation in compliance across societal groups.

Box 4 Incorporating compliance [56]

The cost effectiveness of genetic testing was evaluated for patients with early-onset breast cancer and their female relatives in Norway. BRCA-testing is currently routine care, with identified carriers being at higher risk of breast and ovarian cancer and being offered risk-reducing surgery (bilateral mastectomy with reconstructive surgery and bilateral salpingo-oophorectomy). Alternative 7- and 14-gene panels, which include BRCA1/2 as well as additional genes associated with increased risk of breast and ovarian cancer, were compared against BRCA testing only. Both imperfect uptake of testing and imperfect uptake of risk-reducing surgery were incorporated in the model. The input values for the uptake parameters were obtained from observational data and assumed to have a beta distribution. Uptake of testing as well as risk-reducing surgery among relatives of index patients were assumed to vary with age. Mean uptake of genetic testing among relatives of carriers was assumed to be 0.30 for ages 18-29; 0.82 for ages 30-49; and 0.80 for ages ≥ 50. Mean uptake of prophylactic mastectomy was assumed to be 0.12 for ages 25-34 and 0.11 for ages 35-60. Mean uptake of prophylactic salpingo-oophorectomy was assumed to be 0.10 for ages 25-34; 0.28 for ages 35-39; and 0.35 for ages 40-60. Uptake of the risk-reducing surgeries was assumed to be higher among index patients than their relatives (mean uptake of 0.39 for mastectomy among index patients, 0.36 for salpingo-oophorectomy). The ICER of the 7-gene panel versus BRCA only was estimated to be $53,310, while the ICER of the 14-gene panel versus the 7-gene panel was estimated to be $127,071.

Uncertainty Analysis

To enable optimal reimbursement decisions, it is important to present uncertainty in the cost-effectiveness estimates. Personalised medicine tends to be rife with parameter and structural uncertainty. While PM is often associated with the tailoring of treatments to individual patients, in practice, PM generally divides patients into groups, albeit small groups, and provides the same treatment within each group. The small sample sizes in RCTs and the use of observational data tend to increase the uncertainty of the treatment effect. Other input parameters in economic models of PM may also be uncertain, such as the prevalence of the genetic marker in the target population, testing costs, and the sensitivity and specificity of the testing technology used.

Given limited data availability, expert judgement may be used to provide estimated values for the input parameters. However, expert judgement, too, carries uncertainty, which should be reflected. Several methods have been developed to synthesise the estimated values by multiple experts for a single parameter into a probability distribution that can be included in a sensitivity analysis [57,58,59,60,61].

21. When expert judgement is used to estimate values for the input parameters in the model, synthesise the elicited values into a probability distribution to be included in a sensitivity analysis.

Considerable structural uncertainty arises in PM owing to, for example, the myriad assumptions and decisions that must be made about how to reflect complex testing and treatment pathways, about the expected duration of treatment effect or regarding the methods used to obtain effectiveness estimates when RCT data are not available. Structural uncertainty may have a significant impact on cost-effectiveness results. A failure to assess structural uncertainty provides an incomplete depiction of the decision problem to decision makers. Structural uncertainty may be assessed through a sensitivity analysis in which the effect of plausible alternative assumptions and decisions is investigated. However, performing a sensitivity analysis does not allow for the assessment of the decision uncertainty and the value of information [62]. Alternative options are (i) model averaging, which can provide an assessment of decision uncertainty and (ii) the parameterisation of structural uncertainties, which allows for the assessment of both decision uncertainty and the value of information [62].

22. Identify uncertainties in structural assumptions and decisions and investigate their impact on cost-effectiveness results through a sensitivity analysis. Parameterise structural aspects where possible.

Managed Entry Agreements

When uncertainty regarding an intervention's comparative effectiveness, cost effectiveness and/or budget impact precludes a reimbursement decision, as may commonly be the case for PM, managed entry agreements (MEAs) between a healthcare payer and a manufacturer can be used to offer patients access to the intervention. The two main types of MEA are financial (e.g. discounts, price-volume agreements) and outcomes based (e.g. payment based on individual patient response) and both can be constructed with different levels of risk sharing [63]. The conditions of a MEA can be incorporated into health economic models, as some of the costs may be shifted to a different point in time and should be discounted appropriately. Indeed, models may be used to optimise the conditions of a MEA, such as the period a pharmaceutical

and/or diagnostic test is provided at no cost, the price cap or the time point at which treatment effectiveness for individual patients is assessed (see Box 5 for an example). The optimisation criterion could be a combination of discounted cash flow (most relevant from the manufacturer's perspective) and incremental net monetary benefit and budget constraints (most relevant for payers).

23. If a managed entry agreement is being considered for an intervention, include its conditions in the model evaluating the intervention.

Box 5 Incorporating the conditions of a managed entry agreement [64]

The cost effectiveness of tisagenlecleucel was estimated for paediatric patients with relapsed/refractory B-cell acute lymphoblastic leukaemia in the United States. At the time of the study, the manufacturer of tisagenlecleucel was using an outcome-based payment scheme whereby the payer only bears the cost of treatment when the patient achieves initial remission (i.e. no cure, no pay). It was found by the authors that the cost effectiveness of tisagenlecleucel under the existing outcomes-based payment scheme was similar to the cost effectiveness without such a payment scheme, owing to the high rate of initial remission among patients treated with tisagenlecleucel. In a scenario analysis, the effect of changing the conditions of the payment scheme was evaluated. Under a payment scheme whereby the payer bears the cost of treatment only if the patient is still in remission at 6 months, tisagenlecleucel's ICER improved from $74,000 under the assumption of 40% relapse-free survival ($188,000 under the assumption of 20% relapse-free survival) to $47,000 ($115,000). When the payer was assumed to bear the cost only if the patient is still in remission at 12 months, the ICER improved to $28,000 ($58,000).

Discussion

Twenty-three recommendations for economic evaluations of PM are provided in this study, covering a broad range of topics. As mentioned in the Introduction, the recommendations were developed against a backdrop of calls for the review of HTA methodology, given developments in the field of PM. A systematic and multi-faceted effort was therefore undertaken to assess the extent to which existing HTA methods need to be adapted for PM. The consensus among interviewed experts was that existing methods are adequate and appropriate for assessing PM and non-PM alike. The experts also felt that, within jurisdictions, PM interventions should be subject to the same basic HTA framework as non-PM interventions, to ensure comparability between economic evaluations and consistency in decision making (as reflected in recommendations 1 and 2). Nonetheless, several challenges were identified that may be faced by those producing or evaluating economic evaluations of PM.

The guidance aims to serve as an overview of topics that should be considered for economic evaluations of PM. Some of the recommendations may remind modellers and evaluators of good practices that are often neglected (e.g. recommendations 3–5), others may provide direction when modellers and evaluators are uncertain how to proceed in the face of ongoing debate (e.g. recommendation 18). The guidance is intended to be used in addition to, rather than as a replacement of, existing, more general modelling guidance [65,66,67,68,69].

It is acknowledged that the recommendations are not relevant exclusively to PM as several challenges in the economic evaluations of PM are also encountered in non-PM. For example, the issue of large upfront costs with benefits stretching far into the future (mentioned in the "Recommendations" section) also appears in the modelling of vaccination programmes. Nonetheless, PM is unique in the range and extent of challenges it faces. Our recommendations are therefore particularly valuable in the modelling of PM.

Last, it should be noted that it is unlikely that all recommendations are relevant to each economic evaluation of PM, and some may be not feasible because of limited data and/or research resources. It is therefore left to the modeller to disregard recommendations when appropriate, though a justification for doing so should be provided.

Limitations

As mentioned in the Introduction, there is a range of interpretations of the term "personalised medicine" and the working definition of PM in this paper may not capture all of them. As a result, the developed guidance might not completely meet the needs of those with a different understanding of PM. Nonetheless, the definition of PM adhered to in this study focuses on "new innovations", which are those most likely to require additional modelling guidance. Those who hold a definition of PM that is more inclusive of well-established healthcare may find that existing guidance suffices for these interventions. Further work may be needed to meet the needs of those who understand PM to be informed by patient preferences [70].

Although machine learning-based technologies and digital health applications are sometimes classed under PM (e.g. [19]), they were excluded from our working definition of PM, for two main reasons. First, it may be debated to what extent these technologies are "personalised". Many machine learning approaches are rooted in statistics. There are numerous statistical tools that are widely used in medicine (such as the Simple Calculated Osteoporosis Risk Estimation prediction model) and not necessarily seen as "new innovations" or as PM. This begs the question as to where to draw the line between various statistical models in deciding whether they can be classed under PM.

Digital health applications are often used to complement or automate existing in-person healthcare services. They may for example be used to send automated appointment reminders to patients, to enable online consults with physicians, to automate some of the administrative tasks for healthcare professionals, to capture patients' health data or to monitor patients from a distance. While these developments mark a shift in the mode of healthcare delivery, they are not clearly more "personalised" than the existing healthcare practices that are usually not regarded as PM. Second, although search terms related to "machine learning" and "digital health" were included in the literature searches, it rendered little relevant hits. Most studies in the digital health category considered relatively simple devices for the monitoring of blood glucose in patients with diabetes mellitus, while no studies were identified for the machine learning category. It was therefore decided that insufficient literature on the economic evaluation of machine learning-based technologies and digital health applications was available to allow for their inclusion in this study.

A certain degree of interpretation and subjective prioritisation of the research findings was inevitable in developing the recommendations, given the normative nature of guidance on what constitutes "good practice". This issue was inherent to the research goal and was mitigated by the fact that voices of many different perspectives, backgrounds, countries and types of expertise were heard, both during the expert interviews (18 experts from different backgrounds and with different specialisations were interviewed), the drafting of the guidance (the recommendations were based on consensus opinion within the sizeable HEcoPerMed consortium) and the stakeholder workshop (around 30 participants from various fields were present).

Implications

A substantial amount of literature on the health economic modelling of PM already exists. Several studies discuss challenges in the modelling of PM and suggest potential solutions but do not provide clear guidance to health economic modellers and/or the evaluators of health economic models [14, 71,72,73]. Other relevant studies do provide guidance, mostly in the form of a quality checklist, but on topics more specific than "personalised medicine". Among these are checklists by Kip et al. [27] and Yang et al. [74] to assess the quality of economic evaluations of diagnostic tests, a checklist to assess economic evaluations of gene therapies [15], and a checklist of PM models focusing on the need for patient-level modelling [75]. A study by Christensen et al. [76] provides some guidance specifically on how to measure the costs of integrating whole-genome sequencing (WGS) into clinical practice.

Note that no recommendation on the specific modelling technique to be used in PM was included in this study as the appropriate modelling technique was deemed to depend

on the decision context. Nonetheless, it is acknowledged that the rise of PM has been argued to call for a more widespread use of patient-level modelling (as opposed to cohort-level modelling, which is currently most prevalent), owing to the former's ability to simulate a greater variety of clinical pathways and easily include patient history into the analysis [75, 77]. In patient-level modelling, in addition to considering the parameter and structural uncertainty that are discussed in the guidance, special attention should be paid to addressing patient heterogeneity and stochastic uncertainty [78].

The results of this systematic review add to the existing literature by providing comprehensive and specific guidance to all modellers of PM and evaluators/reviewers of PM models. This may increase the consistency, comparability and quality of economic evaluations of PM, and therefore improve the evidence about the added value of PM. The guidance could also be used for developing and/or evaluating early health economic models, though several recommendations may be more difficult to implement because of limited data (e.g. recommendation 11).

Even for standard health economic models, various recommendations encourage the use of data that may not be available in practice, among which: effectiveness data obtained through trials with two (or more) treatment strategies (recommendation 11); data on the relationship between surrogate outcomes and final outcomes (recommendation 12); and data on the prognostic value of the genetic marker of interest (recommendation 14). The absence of these data items can hamper an accurate assessment of the cost effectiveness of PM, as it may introduce a high level of uncertainty regarding the cost-effectiveness results. Although marketing authorisation may be granted to pharmaceuticals based on relatively limited data, the above-mentioned recommendations highlight that data needs are different for health economic modelling. This is in line with recent examples of PM receiving approval from the European Medicines Agency but subsequently being rejected for reimbursement by national HTA bodies because of inconclusive evidence [7]. While modelling can be used to address some of the data limitations, the main solution to insufficient data may be increased communication between regulators and national HTA agencies about what type of data is needed.

Recommendation 4 urges modellers to include downstream costs and health outcomes of testing. While estimating downstream costs and health outcomes is likely to be feasible for targeted gene panels, which are currently most widely used in healthcare, it may become an increasingly unwieldy task as whole-exome sequencing and WGS become more ingrained in standard clinical practice. Whole-exome sequencing and WGS can find genetic variants associated with (increased risk of) a wide range of conditions. Estimating the effect of the identification of these variants and subsequent (preventive) actions therefore requires knowledge of and data from many disease areas. Existing studies

on the topic have tended to simplify their analyses, for example, by only considering short-term downstream consequences [76] or a subset of possible disease areas [79]. More research on how to best capture downstream costs and health outcomes may be valuable. Nonetheless, estimates of the downstream costs and health outcomes of whole-exome sequencing and WGS are bound to be shrouded in uncertainty for the foreseeable future as much is still unknown about the relationships between genotype and phenotype. That is, whole-exome sequencing and WGS may identify many variants of *unknown* significance to the person's health, severely hindering a reliable estimation of the downstream costs and health outcomes of applying these technologies. The solution to this may lie to a larger extent in continued biomedical research rather than in increasingly sophisticated HTA methods.

In the "Additional Elements of Value" section, it was argued that it is unclear how to measure many of the suggested value elements, partly owing to their conceptual ambiguity. It was also noted that there appears to be an unduly focus on positive value, with limited attention for plausible elements of negative value. Recommendation 18 therefore discourages the incorporation of additional value elements in base-case analyses. Nonetheless, it is acknowledged that value elements beyond the traditional QALY may be relevant in decision making. Indeed, some elements, such as equity, are routinely considered in some countries. Currently, additional value elements tend to be incorporated qualitatively (though numerical values are sometimes used in a multiple-criteria decision analysis [80]), which does not enable the explicit assessment of the trade-offs between length and quality of life on the one hand and other value elements on the other hand. Quantifying additional value elements (in combination with estimating the change in the cost-effectiveness threshold if additional value elements are included) would provide insight into these trade-offs. Further research may be conducted to identify all relevant elements of value, clearly define them and determine how to measure them. Nonetheless, researchers are encouraged to stay mindful of the difference between value to individuals and value to society at large, given that healthcare payers often make decisions at the societal level and may not be willing to pay for all elements that bring value to individual patients.

Conclusions

This study provides a comprehensive list of recommendations to modellers of PM and evaluators/reviewers of such models. The recommendations provide valuable guidance, given the ongoing discussions about the value of PM and the many modelling complexities brought about by PM, and aim to contribute to improved consistency and quality across different health economic models of PM.

References

1. D'Andrea E, Marzuillo C, Pelone F, et al. Genetic testing and economic evaluations: a systematic review of the literature. Epidemiol Prev. 2015;39:45–50.
2. Wright SJ, Newman WG, Payne K. Accounting for capacity constraints in economic evaluations of precision medicine: a systematic review. Pharmacoeconomics. 2019;37:1011–27.
3. Hatz MH, Schremser K, Rogowski WH. Is individualized medicine more cost-effective? A systematic review. Pharmacoeconomics. 2014;32:443–55.
4. Plothner M, Ribbentrop D, Hartman JP, et al. Cost-effectiveness of pharmacogenomic and pharmacogenetic test-guided personalized therapies: a systematic review of the approved active substances for personalized medicine in Germany. Adv Ther. 2016;33:1461–80.
5. Ferkol T, Quinton P. Precision medicine: at what price? Am J Respir Crit Care Med. 2015;192:658–9.
6. O'Sullivan BP, Orenstein DM, Milla CE. Pricing for orphan drugs: will the market bear what society cannot? JAMA. 2013;310:1343–4.
7. Touchot N, Flume M. Early insights from commercialization of gene therapies in Europe. Genes (Basel). 2017;8:78.
8. Vegter S, Boersma C, Rozenbaum M, et al. Pharmacoeconomic evaluations of pharmacogenetic and genomic screening programmes: a systematic review on content and adherence to guidelines. Pharmacoeconomics. 2008;26:569–87.
9. Beaulieu M, de Denus S, Lachaine J. Systematic review of pharmacoeconomic studies of pharmacogenomic tests. Pharmacogenomics. 2010;11:1573–90.
10. European Commission. Commission Staff Working Document. Brussels, Belgium. https://ec.europa.eu/research/health/pdf/2013-10_personalised_medicine_en.pdf. Accessed 27 June 2019
11. Schleidgen S, Klingler C, Bertram T, et al. What is personalized medicine: sharpening a vague term based on a systematic literature review. BMC Med Ethics. 2013;14:55.
12. Simeonidis S, Koutsilieri S, Vozikis A, et al. Application of economic evaluation to assess feasibility for reimbursement of genomic testing as part of personalized medicine interventions. Front Pharmacol. 2019;10:830.
13. Buchanan J, Wordsworth S, Schuh A. Issues surrounding the health economic evaluation of genomic technologies. Pharmacogenomics. 2013;14:1833–47.
14. Jonsson B, Hampson G, Michaels J, et al. Advanced therapy medicinal products and health technology assessment principles and practices for value-based and sustainable healthcare. Eur J Health Econ. 2019;20:427–38.
15. Drummond MF, Neumann PJ, Sullivan SD, et al. Analytic considerations in applying a general economic evaluation reference case to gene therapy. Value Health. 2019;22:661–8.
16. European Network for Health Technology Assessment. Diemen, The Netherlands. https://eunethta.eu/wp-content/uploads/2018/01/2016-03-07_reflection_paper_pm_2nd_draft.pdf. Accessed 27 June 2019.
17. Office of Health Economics. London, UK. https://www.ohe.org/publications/exploring-assessment-and-appraisal-regenerative-medicines-and-celltherapy-products. Accessed 22 Mar 2019.
18. Rogowski W, Payne K, Schnell-Inderst P, et al. Concepts of 'personalization' in personalized medicine: implications for economic evaluation. Pharmacoeconomics. 2015;33:49–59.
19. Love-Koh J, Peel A, Rejon-Parrilla JC, et al. The future of precision medicine: potential impacts for health technology assessment. Pharmacoeconomics. 2018;36:1439–51.
20. Lakdawalla DN, Doshi JA, Garrison LP Jr, et al. Defining elements of value in health care: a health economics approach: an ISPOR Special Task Force report [3]. Value Health. 2018;21:131–9.
21. Bramer WM, Giustini D, de Jonge GB, et al. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc. 2016;104:240–3.
22. Othus M, Barlogie B, Leblanc ML, et al. Cure models as a useful statistical tool for analyzing survival. Clin Cancer Res. 2012;18:3731–6.

23. National Institute for Health and Care Excellence. London, UK. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technologyappraisal-2013-pdf-2007975843781. Accessed 26 Mar 2020.
24. Garrison LP, Towse A. Value-based pricing and reimbursement in personalised healthcare: introduction to the basic health economics. J Pers Med. 2017;7:10.
25. Garrison LP Jr, Zamora B, Li M, et al. Augmenting cost-effectiveness analysis for uncertainty: the implications for value assessment: rationale and empirical support. J Manag Care Spec Pharm. 2020;26:400–6.
26. Attema AE, Brouwer WB, Claxton K. Discounting in economic evaluations. Pharmacoeconomics. 2018;36:745–58.
27. Kip MM, IJzerman MJ, Henriksson M, et al. Toward alignment in the reporting of economic evaluations of diagnostic tests and biomarkers: the AGREEDT checklist. Med Decis Mak. 2018;38:778–88.
28. US Food and Drugs Administration. Silver Spring, Maryland, US. https://www.fda.gov/drugs/fda-expands-pembrolizumab-indication-first-linetreatment-nsclc-tps-1. Accessed 14 Sept 2020.
29. Garrison LP Jr, Lalla D, Brammer M, et al. Assessing the potential cost-effectiveness of retesting IHC0, IHC1+, or FISH-negative early stage breast cancer patients for HER2 status. Cancer. 2013;119:3113–22.
30. Whittington MD, McQueen RB, Ollendorf DA, et al. Long-term survival and value of chimeric antigen receptor T-cell therapy for pediatric patients with relapsed or refractory leukemia. JAMA Pediatr. 2018;172:1161–8.
31. Park JJH, Hsu G, Siden EG, et al. An overview of precision oncology basket and umbrella trials for clinicians. CA Cancer J Clin. 2020;70:125–37.
32. Cunanan KM, Iasonos A, Shen R, et al. An efficient basket trial design. Stat Med. 2017;36:1568–79.
33. Percha B, Baskerville EB, Johnson M, et al. Designing robust n-of-1 studies for precision medicine: simulation study and design recommendations. J Med Internet Res. 2019;21:e12641.
34. Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med. 2018;16:29.
35. Hatswell AJ, Baio G, Berlin JA, et al. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. BMJ Open. 2016;6:e011666.
36. Goring S, Taylor A, Muller K, et al. Characteristics of non-randomised studies using comparisons with external controls submitted for regulatory approval in the USA and Europe: a systematic review. BMJ Open. 2019;9:e024895.
37. Grieve R, Abrams K, Claxton K, et al. Cancer Drugs Fund requires further reform. BMJ. 2016;354:i5090.
38. Van De Wetering E, van Exel J, Brouwer WB. The challenge of conditional reimbursement: stopping reimbursement can be more difficult than not starting in the first place! Value Health. 2017;20:118–25.
39. McCabe C, Chilcott J, Claxton K, et al. Continuing the multiple sclerosis risk sharing scheme is unjustified. BMJ. 2010;340:c1786.
40. Garralda E, Dienstmann R, Piris-Gimenez A, et al. New clinical trial designs in the era of precision medicine. Mol Oncol. 2019;13:549–57.
41. Gyawali B, Hey SP, Kesselheim AS. Assessment of the clinical benefit of cancer drugs receiving accelerated approval. JAMA Intern Med. 2019;179:906–13.
42. Bazhenova L, Jiao X, Lokker A, et al. Cancers with NTRK gene fusions: molecular characteristics and prognosis. American Association for Cancer Research Annual Meeting; 22–24 June, 2020; virtual.
43. Heramb C, Wangensteen T, Grindedal EM, et al. BRCA1 and BRCA2 mutation spectrum: an update on mutation distribution in a large cancer genetics clinic in Norway. Hered Cancer Clin Pract. 2018;16:3.
44. Wiesch DG, Meyers DA. Strategies for analyzing genotype-phenotype relationships in asthma. J Allergy Clin Immunol. 2000;105:S482–6.
45. Wordsworth S, Doble B, Payne K, et al. Using “big data” in the cost-effectiveness analysis of next-generation sequencing technologies: challenges and potential solutions. Value Health. 2018;21:1048–53.
46. Kazi DS, Garber AM, Shah RU, et al. Cost-effectiveness of genotype-guided and dual antiplatelet therapies in acute coronary syndrome. Ann Intern Med. 2014;160:221–32.
47. Eggemann H, Ignatov T, Burger E, et al. Moderate HER2 expression as a prognostic factor in hormone receptor positive breast cancer. Endocr Relat Cancer. 2015;22:725–33.

48. Ouwens MJ, Mukhopadhyay P, Zhang Y, et al. Estimating lifetime benefits associated with immuno-oncology therapies: challenges and approaches for overall survival extrapolations. Pharmacoeconomics. 2019;37:1129–38.
49. Janssen-Heijnen M, Houterman S, Lemmens V, et al. Prognosis for long-term survivors of cancer. Ann Oncol. 2007;18:1408–13.
50. Institute for Clinical and Economic Review. Boston, Massachusetts, US. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101744954-pdf. Accessed 22 Mar 2019.
51. Foster MW, Mulvihill JJ, Sharp RR. Evaluating the utility of personal genomic information. Genet Med. 2009;11:570–4.
52. Regier DA, Weymann D, Buchanan J, et al. Valuation of health and nonhealth outcomes from next-generation sequencing: approaches, challenges, and solutions. Value Health. 2018;21:1043–7.
53. Lee DW, Neumann PJ, Rizzo JA. Understanding the medical and nonmedical value of diagnostic testing. Value Health. 2010;13:310–4.
54. Lakdawalla DN, Romley JA, Sanchez Y, et al. How cancer patients value hope and the implications for cost-effectiveness assessments of high-cost cancer therapies. Health Aff (Millwood). 2012;31:676–82.
55. Thokala P, Ochalek J, Leech AA, et al. Cost-effectiveness thresholds: the past, the present and the future. Pharmacoeconomics. 2018;36:509–22.
56. Asphaug L, Melberg HO. The cost-effectiveness of multigene panel testing for hereditary breast and ovarian cancer in Norway. MDM Policy Pract. 2019;4:2381468318821103.
57. Dias LC, Morton A, Quigley J. Elicitation, vol. 1. New York: Springer International Publishing; 2018. p. 3.
58. Bojke L, Claxton K, Sculpher M, et al. Characterizing structural uncertainty in decision analytic models: a review and application of methods. Value Health. 2009;12:739–49.
59. Grigore B, Peters J, Hyde C, et al. EXPLICIT: a feasibility study of remote expert elicitation in health technology assessment. BMC Med Inform Decis Mak. 2017;17:131.
60. Knol AB, Slottje P, van der Sluijs JP, et al. The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environ Health. 2010;9:19.
61. Bojke L, Claxton K, Bravo-Vergel Y, et al. Eliciting distributions to populate decision analytic models. Value Health. 2010;13:557–64.
62. Bilcke J, Beutels P, Brisson M, et al. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Med Decis Mak. 2011;31:675–92.
63. Antonanzas F, Juárez-Castelló C, Lorente R, Rodríguez-Ibeas R. The use of risk-sharing contracts in healthcare: theoretical and empirical assessments. Pharmacoeconomics. 2019;37:1469–83.
64. Lin JK, Lerman BJ, Barnes JI, et al. Cost effectiveness of chimeric antigen receptor T-cell therapy in relapsed or refractory pediatric B-cell acute lymphoblastic leukemia. J Clin Oncol. 2018;36:3192–202.
65. Philips Z, Bojke L, Sculpher M, et al. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. Pharmacoeconomics. 2006;24:355–71.
66. Evers S, Goossens M, de Vet H, et al. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. Int J Technol Assess Health Care. 2005;21:240–5.
67. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS): explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health. 2013;16:231–50.
68. Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care. 2003;41:32–44.
69. Vemer P, Corro Ramos I, van Voorn GA, et al. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. Pharmacoeconomics. 2016;34:349–61.
70. Huls SPI, Whichello CL, van Exel J, et al. What is next for patient preferences in health technology assessment? A systematic review of the challenges. Value Health. 2019;22:1318–28.
71. Annemans L, Redekop K, Payne K. Current methodological issues in the economic assessment of personalized medicine. Value Health. 2013;16(6 Suppl.):S20–6.
72. Payne K, Gavan SP, Wright SJ, et al. Cost-effectiveness analyses of genetic and genomic diagnostic tests. Nat Rev Genet. 2018;19:235–46.

73. Phillips KA, Deverka PA, Marshall DA, et al. Methodological issues in assessing the economic value of next-generation sequencing tests: many challenges and not enough solutions. Value Health. 2018;21:1033–42.
74. Yang Y, Abel L, Buchanan J, et al. Use of decision modelling in economic evaluations of diagnostic tests: an appraisal and review of health technology assessments in the UK. Pharmacoecon Open. 2019;3:281–91.
75. Degeling K, Koffijberg H, IJzerman MJ. A systematic review and checklist presenting the main challenges for health economic modeling in personalized medicine: towards implementing patient-level models. Expert Rev Pharmacoecon Outcomes Res. 2017;17:17–25.
76. Christensen KD, Phillips KA, Green RC, et al. Cost analyses of genomic sequencing: lessons learned from the MedSeq Project. Value Health. 2018;21:1054–61.
77. Marshall DA, Grazziotin LR, Regier DA, et al. Addressing challenges of economic evaluation in precision medicine using dynamic simulation modeling. Value Health. 2020;23:566–73.
78. Corro Ramos I, Hoogendoorn M, Rutten-van Mölken M. How to address uncertainty in health economic discrete-event simulation models: an illustration for chronic obstructive pulmonary disease. Med Decis Mak. 2020;40:619–32.
79. Bennette CS, Gallego CJ, Burke W, et al. The cost-effectiveness of returning incidental findings from next-generation genomic sequencing. Gen Med. 2015;17:587–95.
80. Thokala P, Devlin N, Marsh K, et al. Multiple criteria decision analysis for health care decision making: an introduction. Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health. 2016;19:1–13.

Supplementary materials

Supplementary materials can be found at https://doi.org/10.1007/s40273-021-01010-z

Contents

- Search strategy for the systematic literature review
- References of studies identified during targeted literature review
- Interview templates
- Summarised interview responses
- References of full-text inclusions, including aspects in the data extraction table (Table 1) addressed

CHAPTER 4

Prognostic value of the NTRK fusion biomarker in the Netherlands

Santi I, Vellekoop H, Versteegh M, Huygens S, Dinjens W, Rutten-van Mölken M.

Estimating the Prognostic Value of the NTRK Fusion Biomarker for Comparative Effectiveness Research in The Netherlands. Molecular Diagnosis & Therapy. 2024;28:319-328.

Abstract

Objectives

We evaluated the prognostic value of the neurotrophic tyrosine receptor kinase (NTRK) gene fusions by comparing the survival of patients with NTRK+ tumours with patients without NTRK+ tumours.

Methods

We used genomic and clinical registry data from the Center for Personalized Cancer Treatment (CPCT-02) study containing a cohort of cancer patients who were treated in Dutch clinical practice between 2012 and 2020. We performed a propensity score matching analysis, where NTRK+ patients were matched to NTRK− patients in a 1:4 ratio. We subsequently analysed the survival of the matched sample of NTRK+ and NTRK− patients using the Kaplan–Meier method and Cox regression, and performed an analysis of credibility to evaluate the plausibility of our result.

Results

Among 3,556 patients from the CPCT-02 study with known tumour location, 24 NTRK+ patients were identified. NTRK+ patients were distributed across nine different tumour types: bone/soft tissue, breast, colorectal, head and neck, lung, pancreas, prostate, skin and urinary tract. NTRK fusions involving the NTRK3 gene (46%) and NTRK1 gene (33%) were most common. The survival analysis rendered a hazard ratio (HR) of 1.44 (95% CI 0.81–2.55) for NTRK+ patients. Using the point estimates of three prior studies on the prognostic value of NTRK fusions, our finding that the HR is > 1 was deemed plausible.

Conclusions

NTRK+ patients may have an increased risk of death compared with NTRK− patients. When using historic control data to assess the comparative effectiveness of TRK inhibitors, the prognostic value of the NTRK fusion biomarker should therefore be accounted for.

Introduction

In line with an increased focus on genetic markers to better target cancer care, larotrectinib and entrectinib came to the market as the first two molecularly targeted therapies with a histology-independent (also called tissue or tumour agnostic) label. Larotrectinib received marketing authorisation from the US Food and Drug Administration (FDA) in November 2018 and was conditionally approved by the European Medicines Agency (EMA) in September 2019. Entrectinib followed in August 2019 and July 2020, respectively. Both pharmaceuticals are indicated for patients who have locally advanced or metastatic solid tumours with neurotrophic tyrosine receptor kinase (NTRK) gene fusions. They can be prescribed regardless of the tissue of origin of the tumour, and therefore are classed as histology-independent therapies [1]. While the prevalence of NTRK gene fusions varies significantly across different cancers, NTRK gene fusions are generally rare. A retrospective study conducted in the Netherlands found that NTRK gene fusions were identified in only 0.93% of patients referred for NTRK testing [2]. Indeed, the estimated prevalence across all cancer patients is only 0.30% [3]. NTRK fusions result from chromosomal rearrangements that cause the 3′ region of the NTRK gene to join the 5′ region of a fusion partner gene. Such fusions may result in TRK fusion proteins with constitutively active tyrosine kinases, which can lead to tumour growth [4]. In this paper, we will refer to patients who have cancer tumours with (without) NTRK gene fusions as NTRK+ (NTRK–) patients. Larotrectinib and entrectinib are inhibitors of the three most common types of TRK protein: TRKA, TRKB and TRKC (encoded by the NTRK1, NTRK2 and NTRK3 genes, respectively). Trial results for both larotrectinib and entrectinib appear promising. A pooled analysis for three phase I/II clinical trials for larotrectinib, which included 244 patients, found a 67% objective response rate (ORR) [95% confidence interval (CI) 63–75] and a median duration of response (DoR) of 32.9 months (95% CI 27.3–41.7) [5]. A pooled analysis for three phase I/II trials for entrectinib, including 121 patients with a median follow-up of 25.8 months, found an ORR of 63% (95% CI 31–89) and median DoR of 22.1 months (95% CI 7.4 to not estimable) [6]. However, larotrectinib and entrectinib were evaluated on a subset of tumour types, therefore there is uncertainty on the efficacy of these treatments for other tumour types. Additionally, the trials for larotrectinib and entrectinib were single-arm trials. Due to the lack of a control arm, the comparative effectiveness of the two TRK inhibitors cannot be established from the trial data alone. Additional data on the effectiveness of standard of care (SoC) for NTRK+ patients are needed. Briggs et al. outlined three methods for estimating the counterfactual in the absence of direct comparative data, using larotrectinib as a case study [7]. Two of these methods require access to the patient-level trial data. The first method uses the progression-free survival (PFS) that trial patients experienced during the previous line of therapy (i.e. before receiving the TRK inhibitor) as a proxy for the

comparator arm and assumes the relationship between PFS and overall survival (OS) to be the same for both the TRK inhibitor and the comparator. The second method uses the PFS and OS for non-responders in the trial (i.e. those with stable or progressive disease after receiving the TRK inhibitor) as a proxy for the comparator arm. When patient-level data are not available, a third method can be used, which involves the use of a historical cohort to estimate survival in the control arm. In their study, Briggs et al. performed a systematic literature review for SoC treatment outcomes for each tumour type included in the larotrectinib trial and subsequently weighted the obtained data according to the distribution of tumour types in the trial. However, estimates of the effectiveness of SoC are generally not available for NTRK+ patients specifically, given that cancer treatments have mostly been prescribed based on the tissue of origin (e.g. breast, pancreas), without identifying patients' NTRK status. If patients with NTRK fusions have a different disease prognosis from patients without NTRK fusions, historical data combining NTRK+ and NTRK− patients may provide biased estimates of SoC effectiveness for NTRK+ patients. To establish whether historical data are appropriate, it is important to identify the prognostic value of NTRK fusions. If needed, the estimated prognostic value can subsequently also be used to adjust historical estimates of SoC effectiveness to better reflect the NTRK+ population. That is, when (extrapolated) survival data are available for a sample of NTRK− patients, the estimated hazard ratio can be applied to the survival times to get an estimate of the survival had the population been NTRK+. In this study, we estimated a hazard ratio (HR) for the survival of NTRK+ patients relative to NTRK− patients. We performed a retrospective matching analysis on the Hartwig Medical Foundation (HMF) database, which comprises genomic and clinical data for metastatic cancer patients. We also used a Bayesian framework alongside the frequentist method to evaluate how plausible it is that there is indeed a difference in survival prognosis between NTRK+ and NTRK− patients (i.e. the effect of carrying an NTRK fusion on survival is non-zero, or $HR \neq 1$) through an analysis of credibility [8, 9].

Methods

Data

The HMF database encompasses de-identified genomic and clinical registry data for cancer patients who were treated in Dutch clinical practice. We used data from the Center for Personalized Cancer Treatment (CPCT-02) study (NCT01855477), which is a subcohort in the HMF database. The study was approved by medical ethical committees of the University Medical Center Utrecht and the Netherlands Cancer Institute and was conducted in concordance with the Declaration of Helsinki, Dutch law and Good Clinical Practice. In the CPCT-02 study, whole-genome sequencing of tumour DNA was performed for thousands of patients from 44 academic, teaching and general hospitals

in the Netherlands, over the period from 2012 until 2020. Patients were eligible for enrolment in the CPCT-02 study if (1) their age was ≥ 18 years, (2) they had a locally advanced or metastatic solid tumour, (3) they had an indication for a new line of systemic treatment with registered anti-cancer agents, (4) performing a biopsy on tumour tissue was safe according to the treating physician and (5) frozen blood and tissue samples were available and sufficient for whole-genome sequencing (WGS). All included patients gave explicit consent for the use of their genomic and clinical data for research purposes. From the HMF database we obtained various genetic markers that were identified in patients' tumour DNA, including NTRK gene fusions and other markers that are known as actionable targets for treatment. Detailed information on sample collection and the WGS procedure can be found elsewhere [10–12]. We also extracted data on several clinical variables, including the age and sex of the patient, the tumour type (i.e. tissue of origin), the year(s) in which tumour biopsies were performed, the starting date of the first post-biopsy treatment, the number of previous lines of therapy, a binary variable indicating whether the patient had died during the period of the study and, for patients remaining alive, the last known date at which they were still alive.

Matching Analysis

Patients were classified into two cohorts: NTRK+ patients and NTRK– patients. Given that the CPCT-02 study provides sequencing data of the tumour DNA, it cannot be known with certainty whether identified NTRK gene fusions are functional, i.e. whether they lead to the expression of fusion TRK proteins that have constitutive tyrosine kinase activity. Nonetheless, two necessary conditions for an NTRK gene fusion to be functional can be determined from the tumour DNA, namely an NTRK1, NTRK2 or NTRK3 gene with a complete tyrosine kinase domain is present on the 3′ end of the (postulated) transcript and the fusion gene (likely) encodes for an in-frame protein. Only patients who had NTRK gene fusions meeting the conditions were included in the NTKR+ cohort, while patients with NTRK gene fusions that did not meet the conditions were included in the NTRK– cohort.

To increase comparability between the NTRK+ and NTRK– patient cohorts, we only included NTRK– patients who had one of the tumour types appearing in the NTRK+ cohort. In both cohorts, patients who had received experimental treatments were excluded, given that our aim was to estimate the effectiveness of standard care. Patients for whom survival time could not be estimated because of missing dates on their appointment logs were also excluded.

We subsequently performed a propensity score matching analysis to identify a subgroup of NTRK– patients similar to the group of NTRK+ patients. Within each tumour type, patients were matched on the available demographic and clinical variables in the

HMF database, i.e. age, sex, year of biopsy and number of previous lines of therapy. Age and sex are well-reported factors influencing expected disease outcomes, hence were included. The 'year of biopsy' variable was included to address possible changes in treatment patterns and treatment effectiveness over the included time period (2012–2020). The number of previous lines of treatment was included as a binary variable (≤ 2 or > 2 previous lines) and served as a proxy reflecting patients' severity of disease, given that patients who have had many treatments already may be in a more advanced stage of disease. We used the optimal matching method [12] (see Online resource S1 for more details) without replacement, with a ratio of 1:4 (NTRK+: NTRK–) and a caliper width of 0.5 times the pooled estimate of the common standard deviation of the logits of the propensity scores. With smaller calipers, it was not possible to find a feasible optimal fixed ratio matching. Given the small sample size, no interaction terms or higher orders of the covariates were used. To assess whether the NTRK+ cohort and the matched NTRK– cohort were sufficiently similar to enable reliable estimation of the prognostic value of NTRK gene fusions, we used the three conditions outlined by Rubin [13]. First, the difference in the means of the propensity scores in the NTRK+ and NTRK– groups must be small, with the standardised measure Rubin's B < 0.25. Also, the ratio of the variances of the propensity scores in the two groups (Rubin's R), as well as the ratio of the variances of the residuals of the covariates after adjusting for the propensity score, must be between 0.5 and 2.

Although NTRK gene fusions are generally seen to be driver gene alterations (i.e. the alteration causing the onset and progression of tumour growth), they might in some cases not be the (only) oncogenic driver. We therefore performed a sensitivity analysis where we excluded NTRK+ patients whose tumour DNA contained other (known) oncogenic biomarkers. The remaining NTRK+ patients were matched to NTRK– patients using the same method as in the main analysis. Based on current knowledge about actionable biomarkers, we included mutations in the ALK, BRAF, EGFR, ERBB2, KRAS or ROS1 genes, as well as high tumour mutational burden (TMB) and microsatellite instability (MSI).

Survival Analysis

We analysed the survival of patients with and without NTRK gene fusions using the Kaplan–Meier method and Cox regression. To calculate patients' overall survival (OS), we estimated the period between the start of the first post-biopsy treatment and the time of death or censor. Patients who were not recorded as dead were censored at their last known date of being alive, which was the date of their last appointment to assess response to treatment. The survival analysis was also performed on the sensitivity analysis dataset described above.

Analysis of Credibility

Because of their small sample sizes, studies on the prognostic value of rare mutations such as NTRK gene fusions suffer from a lack of power in frequentist inference. This may lead to statistically insignificant study results. Also, p-values have been argued to be poor indicators of whether an effect is truly present (or absent) [8]. Instead, we used the analysis of credibility (AnCred) method, which originates from Bayesian methods, and is seen as a more nuanced alternative for evaluating the plausibility of study findings than the 'pass/fail' dichotomy posed by the p-value threshold of 0.05 [15]. In AnCred, the study finding (expressed as a point estimate and 95% CI) is used to calculate a critical prior interval (CPI) (see Online Resource S2 for more details) [9, 15]. The CPI indicates the level of support needed from prior studies to have credible evidence for a non-zero effect. For example, when the study finding of interest is non-significant, previous studies will make the finding plausible of a non-zero effect size if their point estimates fall within the CPI. This process is an inversion of Bayes' Theorem, as the study finding is used to deduce the range of prior effect sizes—the CPI—leading to a posterior interval that excludes 'no effect'.

Results

Patient Characteristics

Among 3,556 patients from the CPCT-02 study with known tumour location, 24 had tumours harbouring a likely functional NTRK gene fusion (Fig. 1). NTRK+ patients were spread across nine different tumour types: bone/soft tissue, breast, colorectal, head and neck, lung, pancreas, prostate, skin and urinary tract. The distribution of the different NTRK genes (NTRK1/NTRK2/NTRK3) across the tumour types is shown in Fig. 2. Among the remaining 2,719 patients without an NTRK gene fusion, 2,069 had one of the tumour types occurring in the NTRK+ cohort hence were included in the NTRK− cohort (Fig. 1).

In the NTRK+ cohort, the median age was 59 years (range 55–67 years), and 13 patients (54%) were female (Table 1). A minority of patients (33%) had received more than two lines of prior therapy. Most NTRK fusions involved the NTRK3 gene (11 patients, 46%) or NTRK1 gene (8 patients, 33%) (Fig. 2). Of the 24 different fusion partners identified, 20 were novel fusions that have not yet been reported in the Quiver database, a curated database of known oncogenic gene fusions. Other biomarkers found among NTRK+ patients were mutations in the BRAF, EGFR, ERBB2 and KRAS genes, as well as high TMB and MSI (Table 2) [16].

In the (non-matched) NTRK− cohort, the median age was higher (63 years, range 55–70 years), as was the percentage of patients with more than two lines of prior therapy (47%)

(Table 1). The tumour distributions also differed between the non-matched NTRK– cohort and the NTRK+ cohort.

Table 1 Patient characteristics

	NTRK+ (n = 24)	NTRK– (n = 2,165)			
		Non-matched (n=2,069)	SD	Matched (n=96)	SD
Age in years, median (range)	59.0 (55.5;67.5)	63.0 (55.0;70.0)	-0.150	59.0 (55.0;67.0)	0.081
Gender, n (%)			0.043		0.020
Female	13 (54.2)	1,077 (52.1)		55 (57.3)	
Male	11 (45.8)	992 (47.9)		41 (42.7)	
Primary tumour location, n (%)			0.657		0.000
Bone/Soft tissue	1 (4.2)	126 (6.1)		4 (4.2)	
Breast	5 (20.8)	560 (27.1)		20 (20.8)	
Colon/Rectum	1 (4.2)	362 (17.5)		4 (4.2)	
Head and neck	1 (4.2)	34 (1.6)		4 (4.2)	
Lung	4 (16.7)	242 (11.7)		16 (16.7)	
Pancreas	4 (16.7)	106 (5.1)		16 (16.7)	
Prostate	3 (12.5)	267 (12.9)		12 (12.5)	
Skin	4 (16.7)	249 (12.0)		16 (16.7)	
Urinary tract	1 (4.2)	123 (5.9)		4 (4.2)	
Number of previous lines of therapy (categories), n (%)			0.282		0.040
<=2	16 (66.7)	1,096 (53.0)		62 (64.6)	
>2	8 (33.3)	973 (47.0)		34 (35.4)	
Year of biopsy, n (%)			-0.151		0.003
2015	1 (4.2)	48 (2.3)		3 (3.1)	
2016	6 (25.0)	303 (14.6)		23 (24.0)	
2017	8 (33.3)	788 (38.1)		35 (36.5)	
2018	5 (20.8)	694 (33.5)		25 (26.0)	
2019	2 (8.3)	200 (9.7)		3 (3.1)	
2020	1 (4.2)	27 (1.3)		3 (3.1)	
2021	1 (4.2)	9 (0.4)		4 (4.2)	

Table 2 Identified NTRK gene fusions and concurrent biomarkers

Tumour location	Gene fusion
Bone/soft tissue (n=1)	TPM3_NTRK1
Breast (n=5)	CYP11A1_NTRK3
	EFTUD1P1_NTRK3
	GCNT1_NTRK2
	RP11-315D16.2_NTRK3
	SEMA4B_NTRK3
Colorectal (n=1)	SFPQ_NTRK1
Head and neck (n=1)	PRCC_NTRK1
Lung (n=4)	ITLN2_NTRK1
	PIGR_NTRK1
	SLC25A21_NTRK1
	TGM6_NTRK2
Pancreas (n=4)	CAMK2A_NTRK3
	EML4_NTRK3
	SH2D2A_NTRK1
	TPR_NTRK1
Prostate (n=3)	AC005772.2_NTRK3
	ATAD2_NTRK3
	RBPJ_NTRK3
Skin (n=4)	CTD-2034I4.1_NTRK3
	PTGFRN_NTRK2
	SH3GL3_NTRK3
	TNKS_NTRK2
Urinary tract (n=1)	MAPKAP1_NTRK2

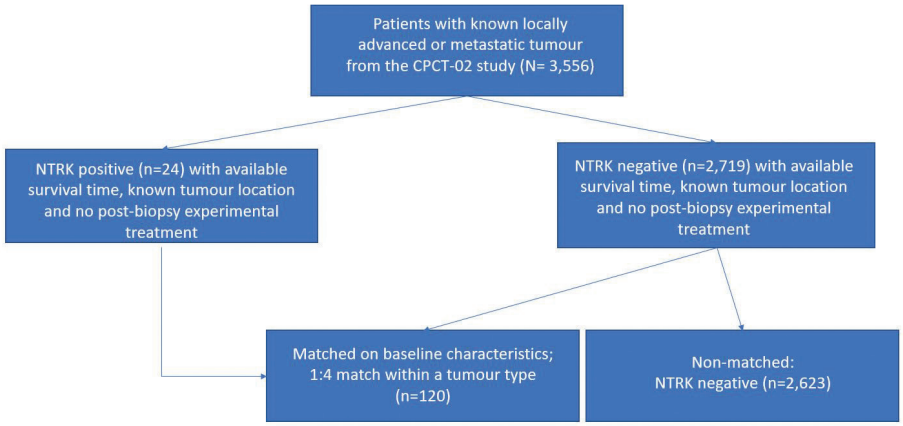


Figure 1 Study schema. CPCT-02, Center for Personalized Cancer Treatment, NTRK, neurotrophic tyrosine receptor kinase

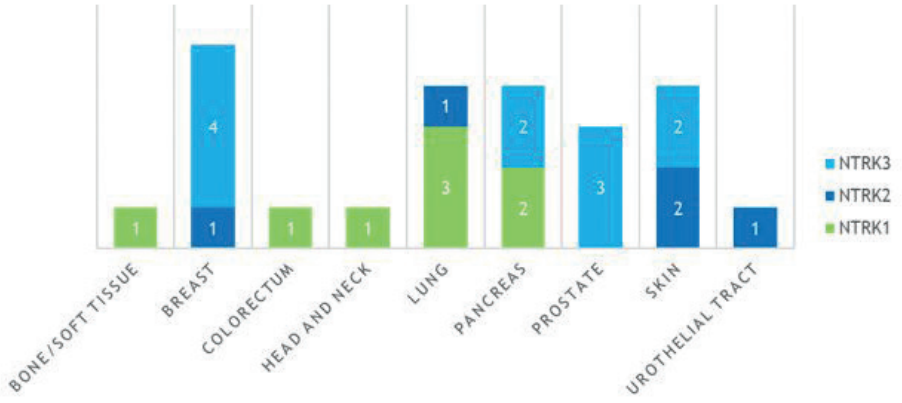


Figure 2 Distribution of tumour types in the NTRK+ cohort

Matching and Survival Analysis

In the propensity score matching analysis, the 24 patients in the NTRK+ cohort were matched with 96 NTRK– patients. Standardised mean difference between groups were reduced for all covariates after propensity score matching (Table 1). Rubin's B was 0.02 after matching, well below the recommended upper limit of 0.25. The variance ratios (Rubin's R) of the propensity score and the covariates were also within the recommended range of 0.5–2 (Online Resource S3). Moreover, the box plot of the distribution of the logit of the propensity score shows an optimal overlap for the matched observations (Online Resource S4). Similarly, balance was obtained in the propensity score matching

sensitivity analysis (Online Resources S5 and S6). Median OS of 12.7 months (95% CI 6.3–17.4) and 11.6 months (95% CI 7.8–17.9) were observed in the NTRK+ cohort and the matched NTRK− cohort, respectively. Despite the longer median OS for NTRK+ patients, the survival analysis rendered an HR of 1.44 (95% CI 0.81–2.55) (Fig. 3), meaning that NTRK+ patients are at higher risk of dying than NTRK− patients. The adjusted Cox regression provided an HR very close to the unadjusted, i.e. HR of 1.41 (95% CI 0.79–2.52). This result is in line with the reduction in the standardised mean difference between the covariates used for the propensity score, which is below 0.10, implying that performing a double adjustment is not necessary [17]. Additionally, a restricted mean survival time (RMST) analysis was conducted up to 40 months, representing the minimum of the largest observed event time within the NTRK− cohort. A 16.3 month RMST (95% CI 13.0–19.7) was estimated for NTRK+ patients, compared to 12.5 months RMST (95% CI 9.3–16.3) for NTRK− patients, supporting the results of the survival analysis. In the sensitivity analysis, where NTRK+ patients with concurrent oncogenic biomarkers were excluded, we found a lower HR than in the main analysis (1.20, 95% CI 0.61–2.36) (Fig. 4).

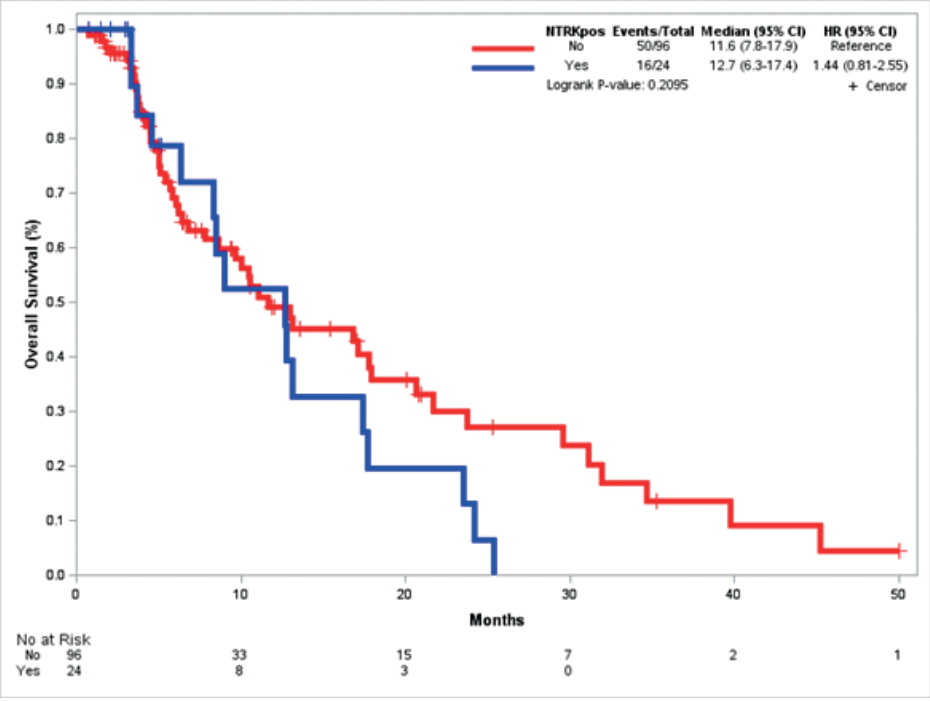


Figure 3 Kaplan-Meier plot for OS analysis

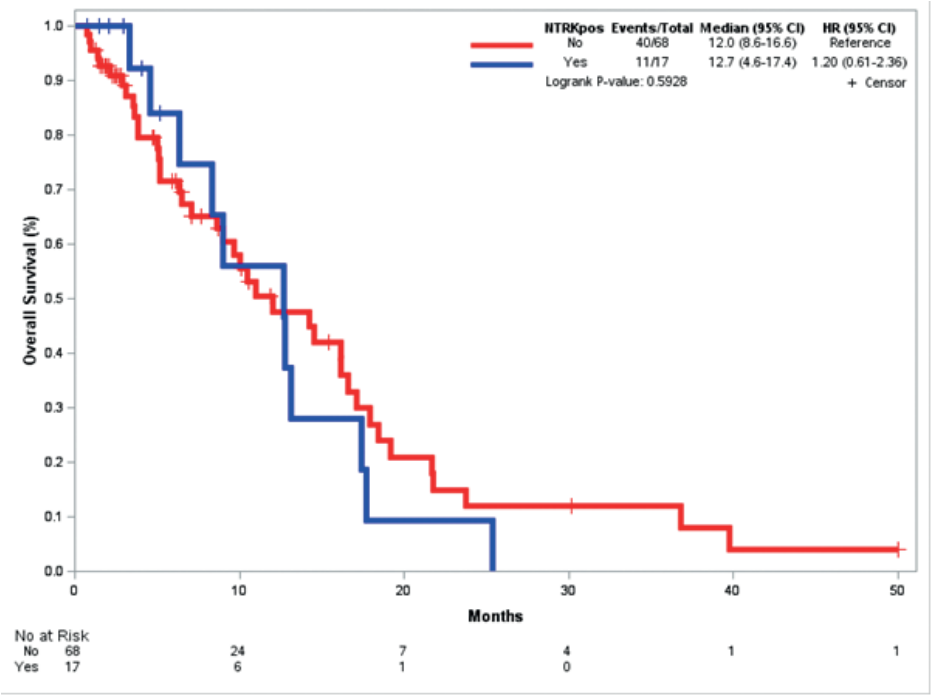


Figure 4 Kaplan-Meier plot for OS sensitivity analysis

Analysis of Credibility

The point estimate (1.44) and 95% CI (0.81–2.55) in our main analysis show that the central effect is in the direction of a positive effect (i.e. HR > 1). However, HR values smaller than 1 are included in the 95% CI, and so the estimated point value is statistically non-significant. The CPI associated with our results was calculated to be 1.0–11.2 (see Online Resource S2 for details), meaning that prior studies with estimates falling within this range make it more plausible that the HR for the survival of patients with NTRK+ tumours is larger than 1. To our knowledge, only three other studies have estimated the prognostic value of NTRK fusions. Two used the Flatiron Health-Foundation Medicine clinic-genomic database and one used the Genomic England database. Hibar et al. found an HR of 1.6 (95% CI 1.0–2.5) on survival analysis of 28 NTRK+ patients and 280 matched NTRK− patients. Bazhenova et al. found an HR of 1.44 (95% CI 0.61–3.37) in an analysis of 27 NTRK+ and 107 matched NTRK− patients [18]. Bridgewater et al. analysed 18 NTRK+ and 72 matched NTRK− patients and found a similar HR value of 1.47 (95% CI 0.39–5.57) [19]. Given that the point estimates of all three studies fall within the CPI, the studies support the plausibility of our finding that the survival HR for NTRK+ patients is > 1. That is, it is plausible that NTRK+ patients have a worse prognosis than NTRK− patients.

Discussion

Our study describes the clinical characteristics and survival of NTRK+ patients with advanced or metastatic disease who have previously been treated in Dutch clinical practice with SoC therapies other than targeted TRK inhibitors. NTRK+ patients appeared to have worse survival compared with NTRK– patients.

As the focus on better targeted, or 'personalised', cancer care continues, NTRK+ patients may be the first of many small patient groups with a specific genetic marker for whom treatment effectiveness must be evaluated. It has been argued that randomised controlled trials (RCTs), the preferred option to reliably estimate effectiveness [20, 21], may be difficult and time consuming to conduct for such small patient groups [22]. Although adapted, more flexible versions of the RCT design have been suggested [23, 24], pharmaceutical companies have so far mostly resorted to single-arm trials [23, 24]. Single-arm trials are poorly equipped to provide estimates of relative treatment effectiveness, due to the absence of a control arm reflecting the effectiveness of standard care. Briggs et al. [7] outlined possible ways to construct a control arm when faced with single-arm trial data for tumour-agnostic (i.e. genetic marker-focussed) treatments, including the use of historical data. We have expanded their work by arguing that, when evaluating the effectiveness of a treatment targeting a specific genetic marker, historical data may have to be adjusted for the prognostic value of said genetic marker. In this study we focussed on estimating the prognostic value of NTRK gene fusions. How the results can subsequently be used in an economic model evaluating treatment effectiveness can be found elsewhere [25]. While the results indicate that NTRK+ is a prognostic factor for earlier death relative to NTRK–, when using the HR on extrapolated survival estimated on NTRK– patients, the proportional hazards assumption is adopted for the entire forecasted period. Looking at the Kaplan–Meier survival plots, it is uncertain whether this assumption holds true.

As mentioned in the "Results" section, the prognostic value of NTRK gene fusions has been estimated in three prior studies focussing on the UK and the USA, using the Genomic England and FlatIron Health-Foundation Medicine databases. All studies to date, including ours, have been retrospective. A number of key differences among the studies can be noted. Firstly, in our study the median age of the patients was around 60 years and no patients under 18 were included, while Bridgewater et al. included paediatric patients. Our study excluded patients treated with either TRK inhibitor or unlabelled therapy. Even though Bazhenova et al. conducted their study prior to the approval of larotrectinib and entrectinib in the USA, one patient with NTRK+ disease had received an unknown investigational agent in a clinical trial. Our cohort included some patients with tumour types not found in other studies (e.g. prostate and urothelial

cancer). Also, the subtype of tumour was missing for some patients in the HMF database. This may be the reason why our study includes head and neck cancer as a broad tumour type, which potentially includes salivary gland tumours. Furthermore, the index date from which OS was measured varied between studies; Bazhenova et al. used the date of gene sequencing report in their primary analysis, Hibar et al. and Bridgewater et al. used the date of diagnosis. Hibar et al. used the start of last available treatment line before the NGS report in a sensitivity analysis. In our study, we used the date of first post-biopsy treatment to avoid potential immortal bias between the date of biopsy and the start of the treatment [26].

Despite these differences, all studies reported the same direction of effect, i.e. NTRK+ status increases the risk of mortality, with varying degrees of uncertainty.

The AnCred methodology has typically been used to interpret study results in the light of prior studies that have demonstrated an effect. Our application of AnCred is slightly different as there was no previous conclusive evidence, but rather previous uncertain evidence due to the sample size restrictions. We interpret our results in light of these previous studies to reflect a credible direction of effect. As EMA increasingly approves drugs based on evidence from single-arm studies, the challenge of dealing with uncertainty in HTA and reimbursement decision making is increasing. Against this background, it is important to use different means of managing uncertainty, one of which is the comparison of the previous results with the critical prior interval of AnCred.

We add to the literature by presenting findings obtained in a different country setting and using a different clinicogenomic database. Our sample distribution over age and primary tumour type broadly aligns with figures on solid cancer incidence in Western Europe, suggesting our Netherlands-focused research results may be applicable to Western Europe more broadly [27, 28].

Limitations

The aim of the CPCT-02 study was to identify patients eligible for clinical trials of targeted therapies (NCT01855477). That is, most patients enrolled in CPCT-02 had little to no SoC alternatives remaining. This is in line with the therapeutic indications for TRK inhibitors entrectinib and larotrectinib, both of which are for patients 'who have no satisfactory treatment options' according to the EMA [29, 30]. Nonetheless, there appear to be differences in patient characteristics between our study and a recent study focussing on NTRK testing in Dutch routine care [2], suggesting that the population included in the CPCT-02 study may not be fully representative of the population subject to NTRK testing (and treatment) in clinical practice. It is unknown to what extent such differences might affect our estimated HR for overall survival.

Because of limited availability of clinical data in the HMF database, we may not have included all relevant covariates in the matching process. Residual confounding can therefore not be ruled out. For example, known predictors of mortality [31] such as disease stage, severity of disease (e.g. measured by Eastern Cooperative Oncology Group [ECOG] performance status), serum albumin and platelet count, were not available in the dataset. For lack of explicit data on patients' severity of disease, we used 'the number of previous lines of therapy' as a proxy. We theorise that patients who have had many treatments already are likely to be in a more advanced stage of disease, but this might not always be true as severely ill patients may be too weak to receive many lines of treatment.

In this study, we estimated a single HR value for all NTRK+ patients. However, evidence suggests heterogeneity in the prognostic value of NTRK gene fusions across tumour types [32]. We deemed our sample of 24 patients with NTRK+ tumours too small to obtain meaningful results from a subgroup analysis. Nonetheless, we encourage further research into methods that might be used to perform subgroup analyses on small patient samples [32].

When excluding NTRK+ patients with concurrent oncogenic biomarkers, we found a lower HR (1.20, 95% CI 0.61–2.36). When concurrent biomarkers are oncogenic drivers, there may be an interplay between said oncogenic drivers and the NTRK fusion gene, whereby collaborating oncogenic pathways are activated and tumour growth may be increased [33]. Thus, including patients with concurrent biomarkers in the NTRK+ cohort, as we did in the main analysis, may lead to an overestimation of the prognostic value of NTRK gene fusions per se. Nonetheless, the HR value estimated in the sensitivity analysis is larger than 1, suggesting that even if the HR value was overestimated in the main analysis, NTRK+ patients are still faced with worse survival than NTRK− patients.

Research and Policy Considerations

The advent of tumour-agnostic cancer care expands treatment opportunities and possibly enables better targeting of care. However, pooling patients in a tumour-agnostic manner when estimating treatment effectiveness may be inappropriate. There is likely heterogeneity in treatment effectiveness across tumours with different histologies and tissues of origin, for example because of differences in survival between tumour types and differences in the prognostic value of oncogenic drivers (e.g. it has been found that the tumour-promoting activity of oncogenic drivers may depend on the tissue of origin) [33, 34]. We therefore recommend that treatment effectiveness is estimated not only for the whole patient population with a specific genetic marker but for relevant subgroups as well. We acknowledge that doing so would reduce the sample sizes per disease indication even further. Solutions may be found during the running of the trial (e.g. stopping rules in an adaptive trial design framework) and in applying statistical

methods that do not assume identical treatment effect between tumour types (e.g. exchangeability assumption in the Bayesian approach), as well as in more extensive collection of (real-world) data [35–37].

Beyond heterogeneity in treatment effect, there may also be heterogeneity in comparative effectiveness and cost-effectiveness, due to differences in the effectiveness and costs of comparative therapies across tumour types, as well as differences in existing testing protocols [e.g. broad genetic testing is already commonplace for non-small cell lung cancer (NSCLC) in the Netherlands, making the additional cost of testing for NTRK gene fusions negligible]. Reimbursement decisions for tumour-agnostic treatments may therefore also have to be specified for relevant subgroups instead of the whole population with the genetic marker.

Our research on the prognostic value of NTRK fusions and, relatedly, the treatment effectiveness of larotrectinib and entrectinib [25], was hampered by limited data. With a larger database and data on more clinical variables, we might have been able to provide further insights. Given that genetic marker-based pharmaceuticals (and single-arm trials) are likely to become more frequent, we encourage policymakers to consider more widespread collection of clinicogenomic data, and better linking of existing databases. As pharmaceutical trials have been notoriously Caucasian- and male focused [38, 39], we would like to stress the importance of ensuring that the populations included in clinic-genomic databases reflect real-life populations.

In conclusion, our findings suggest that patients with tumours harbouring an NTRK fusion gene may have an increased, or at least similar, risk of death compared with matched patients with tumours harbouring NTRK wildtype genes. This emphasises the relevance of NTRK gene fusions as actionable drug targets and provides support for the potential clinical benefit of TRK inhibitor therapy. By showing that survival may differ between NTRK+ and NTRK– patients, our study underscores the need to correct historic control data for the prognostic value of biomarkers when assessing comparative effectiveness.

References

1. Dunn DB. Larotrectinib and entrectinib: TRK inhibitors for the treatment of pediatric and adult patients with NTRK gene fusion. J Adv Pract Oncol. 2020. https://doi.org/10.6004/jadpro.2020. 11.4.9.
2. Koopman B, et al. Detection of NTRK fusions and TRK expression and performance of pan-trk immunohistochemistry in routine diagnostics: results from a nationwide community-based cohort. Diagnostics. 2022;12(3):668. https://doi.org/10.3390/diagnostic s12030668.
3. Westphalen CB, et al. Genomic context of NTRK1/2/3 fusion-positive tumours from a large real-world population. NPJ Precis Oncol. 2021;5(1):69. https://doi.org/10.1038/ s41698-021-00206-y.
4. Amatu A, Sartore-Bianchi A, Bencardino K, Pizzutilo EG, Tosi F, Siena S. Tropomyosin receptor kinase (TRK) biology and the role of NTRK gene fusions in cancer. Ann Oncol. 2019;30:viii4–15. https://doi.org/10.1093/annonc/mdz383.
5. Drilon AE, et al. Long-term efficacy and safety of larotrectinib in a pooled analysis of patients with tropomyosin receptor kinase (TRK) fusion cancer. J Clin Oncol. 2022;40(16_suppl):3100. https://doi.org/10.1200/JCO.2022.40.16_suppl.3100.
6. Demetri GD, et al. Updated integrated analysis of the efficacy and safety of entrectinib in patients with NTRK fusion-positive solid tumors. Clin Cancer Res. 2022;28(7):1302–12. https://doi.org/10. 1158/1078-0432.CCR-21-3597.
7. Briggs A, et al. Comparison of alternative methods to assess the cost-effectiveness of tumor-agnostic therapies: a triangulation approach using larotrectinib as a case study. Value Health J Int Soc Pharmacoecon Outcomes Res. 2022;25(6):1002–9. https:// doi.org/10.1016/j.jval.2021.11.1354.
8. Matthews RAJ. Moving towards the post p < 0.05 era via the analysis of credibility. Am Stat. 2019;73(sup1):202–12. https:// doi.org/10.1080/00031305.2018.1543136.
9. Matthews RAJ. Beyond "significance": principles and practice of the analysis of credibility. R Soc Open Sci. 2018;5(1): 171047. https://doi.org/10.1098/rsos.171047.
10. Priestley P, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575(7781):210–6. https://doi.org/10. 1038/s41586-019-1689-y.
11. Samsom KG, et al. Study protocol: whole genome sequencing implementation in standard diagnostics for every cancer patient (WIDE). BMC Med Genomics. 2020;13(1):169. https://doi.org/ 10.1186/s12920-020-00814-w.
12. Austin PC. A comparison of 12 algorithms for matching on the propensity score. Stat Med. 2014;33(6):1057–69. https://doi.org/ 10.1002/sim.6004. Epub 2013 Oct 7. PMID: 24123228; PMCID: PMC4285163.
13. Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. Health Serv Outcomes Res Methodol. 2001;2(3/4):169–88. https://doi.org/10. 1023/A:1020363010465.
14. 'Driver Catalog'. [Online]. https://github.com/hartwigmedical/ hmftools/blob/master/purple/DriverCatalog.md. Accessed Aug 2022.
15. Held L, Matthews R, Ott M, Pawel S. Reverse-Bayes methods for evidence assessment and research synthesis. Res Synth Methods. 2022;13(3):295–314. https://doi.org/10.1002/jrsm.1538.
16. Archer DX Inc. Quiver database, 2021 [Internet]. [cited March 12, 2021]. http://quiver.archerdx.com. Accessed Aug 2022.
17. Nguyen T-L, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. BMC Med Res Methodol. 2017;17(1):78. https://doi.org/10. 1186/s12874-017-0338-0.
18. Bazhenova L, et al. TRK fusion cancer: patient characteristics and survival analysis in the real-world setting. Target Oncol. 2021;16(3):389–99. https://doi.org/10.1007/s11523-021-00815-4.
19. Bridgewater J, et al. Prognosis and oncogenomic profiling of patients with tropomyosin receptor kinase fusion cancer in the 100,000 genomes project. Cancer Treat Res Commun. 2022;33: 100623. https://doi.org/10.1016/j.ctarc.2022.100623.

20. Levi M, McGovern DPB, Summerskill WSM, Valori RM. Key topics evidence-based medicine. Boca Raton: CRC Press; 2001.
21. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. J Clin Nurs. 2003;12(1):77– 84. https://doi.org/10.1046/j.1365-2702.2003.00662.x.
22. Parmar MKB, Sydes MR, Morris TP. How do you design randomised trials for smaller populations? A framework. BMC Med. 2016;14(1):183. https://doi.org/10.1186/s12916-016-0722-3.
23. Burnett T, Mozgunov P, Pallmann P, Villar SS, Wheeler GM, Jaki T. Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs. BMC Med. 2020;18(1):352. https://doi.org/10.1186/s12916-020-01808-2.
24. Superchi C, et al. Study designs for clinical trials applied to personalised medicine: a scoping review. BMJ Open. 2022;12(5): e052926. https://doi.org/10.1136/bmjopen-2021-052926.
25. Huygens S, et al. Cost-effectiveness analysis of treating patients with NTRK-positive cancer with the histology-independent therapy entrectinib. Value Health J Int Soc Pharmacoecon Outcomes Res. 2023;26(2):193–203. https://doi.org/10.1016/j.jval.2022.08. 006.
26. Yadav K, Lewis RJ. Immortal time bias in observational studies. JAMA. 2021;325(7):686–7. https://doi.org/10.1001/jama.2020. 9151.
27. Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49. https://doi.org/ 10.3322/caac.21660.
28. Ferlay J, et al. Cancer statistics for the year 2020: an overview. Int J Cancer. 2021. https://doi.org/10.1002/ijc.33588.
29. European Medicines Agency (EMA). Assessment report VITRAKVI International non-proprietary name: larotrectinib. July 25, 2019. Available https://www.ema.europa.eu/en/documents/ assessment-report/vitrakvi-epar-public-assessment-report_en.pdf.
30. European Medicines Agency (EMA). CHMP assessment report Rozlytrek International non-proprietary name: entrectinib. May 28, 2020. Available https://www.ema.europa.eu/en/documents/ assessment-report/rozlytrek-epar-public-assessment-report_en. pdf.
31. Owusuaa C, Dijkland SA, Nieboer D, van der Heide A, van der Rijt CCD. Predictors of mortality in patients with advanced cancer-a systematic review and meta-analysis. Cancers. 2022;14(2):328. https://doi.org/10.3390/cancers14020328.
32. Murphy P, et al. Modelling approaches for histology-independent cancer drugs to inform NICE appraisals: a systematic review and decision-framework. Health Technol Assess Winch Engl. 2021;25(76):1–228. https://doi.org/10.3310/hta25760.
33. El Tekle G, et al. Co-occurrence and mutual exclusivity: what cross-cancer mutation patterns can tell us. Trends Cancer. 2021;7(9):823–36. https://doi.org/10.1016/j.trecan.2021.04.009.
34. Hoadley KA, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell. 2018;173(2):291-304.e6. https://doi.org/10.1016/j.cell.2018.03. 022.
35. Liu R, Liu Z, Ghadessi M, Vonk R. Increasing the efficiency of oncology basket trials using a Bayesian approach. Contemp Clin Trials. 2017;63:67–72. https://doi.org/10.1016/j.cct.2017.06.009.
36. Ventz S, Barry WT, Parmigiani G, Trippa L. Bayesian response-adaptive designs for basket trials. Biometrics. 2017;73(3):905–15. https://doi.org/10.1111/biom.12668.
37. Bokemeyer C, et al. Overall survival (OS) of patients with TRK fusion–positive cancer receiving larotrectinib versus standard of care (SoC): a matching-adjusted indirect comparison (MAIC) using real-world data (RWD). J Clin Oncol. 2022;40(16_ suppl):6579–6579. https://doi.org/10.1200/JCO.2022.40.16_suppl.6579.
38. Camidge DR, et al. Race and ethnicity representation in clinical trials: findings from a literature review of phase I oncology trials. Future Oncol Lond Engl. 2021;17(24):3271–80. https://doi.org/ 10.2217/fon-2020-1262.
39. Woods-Burnham L, Johnson JR, Hooker SE, Bedell FW, Dorf TB, Kittles RA. The role of diverse populations in US clinical trials. Med N Y N. 2021;2(1):21–4. https://doi.org/10.1016/j.medj. 2020.12.009.

Supplementary materials

Supplementary materials can be found at https://doi.org/10.1007/s40291-024-00704-2

Contents

- S1: Details on the optimal pair matching method
- S2: Details Analysis of Credibility method
- S3: Balance measures for the propensity score matching analysis
- S4: Logit Propensity Score box plot
- S5: Balance measures for the propensity score matching sensitivity analysis
- S6: Logit Propensity Score box plot (sensitivity analysis)

CHAPTER 5

Cost-effectiveness analysis of treating patients with NTRK-positive cancer with the histology-independent therapy entrectinib

Huygens S, Vellekoop H, Versteegh M, Santi I, Szilberhorn L, Zelei T, Nagy B, Tsiachristas A, Koleva-Kolarova R, Wordsworth S, Rutten-van Mölken M.

Cost-Effectiveness Analysis of Treating Patients with NTRK-Positive Cancer With the Histology-Independent Therapy Entrectinib. Value in Health. 2023;26(2):193-203.

Abstract

Objectives

This study tackles several challenges of evaluating histology-independent treatments using entrectinib as an example. Histology-independent treatments are provided based on genetic marker(s) of tumors, regardless of the tumor type. We evaluated the lifetime cost-effectiveness of testing all patients for NTRK fusions and treating the positive cases with entrectinib compared with no testing and standard of care (SoC) for all patients.

Methods

The health economic model consisted of a decision tree reflecting the NTRK testing phase followed by a microsimulation model reflecting treatment with either entrectinib or SoC. Efficacy of entrectinib was based on data from basket trials, whereas historical data from NTRK-negative patients were corrected for the prognostic value of NTRK fusions to model SoC.

Results

"Testing" (testing for NTRK fusions, with subsequent entrectinib treatment in NTRK-positive patients and SoC in NTRK-negative patients) had higher per-patient quality-adjusted life-years (QALYs) and costs than "No testing" (SoC for all patients), with a difference of 0.0043 and €732, respectively. This corresponded to an incremental cost-effectiveness ratio (ICER) of €169,957/QALY and, using a cost-effectiveness threshold of €80,000/QALY, an incremental net monetary benefit of −€388. When excluding the costs of genetic testing for NTRK fusions, the ICER was reduced to €36,290/QALY and the incremental net monetary benefit increased to €188.

Conclusions

When treatment requires the identification of a genetic marker, the associated costs and effects need to be accounted for. Because of the low prevalence of NTRK fusions, the number needed-to-test to identify patients eligible for entrectinib is large. Excluding the testing phase reduces the ICER substantially.

Introduction

Recently, the European Medicines Agency (EMA) approved the first histology-independent treatments, entrectinib and larotrectinib, for tumors with NTRK gene fusions.[1,2] Histology-independent (also called "tumor-agnostic") treatments are prescribed based on a genetic marker of the tumor, whereas most other oncology treatments are prescribed based on the tumor type. Evaluating the efficacy and cost-effectiveness of existing histology-independent treatments has proven challenging for various reasons.[3]

First, clinical trials for entrectinib and larotrectinib were basket trials where patients with different tumor types were pooled together.[4,5,6] Because of the small number of patients per tumor type, tumor-specific effectiveness was not provided and marketing authorization for the pharmaceuticals was granted for all NTRK-positive (NTRK+) tumors, assuming similar efficacy across tumor types. Nevertheless, there may be heterogeneity in treatment effect across the tumor types.[7] Second, the trials were single-arm trials. The lack of randomized controlled trial (RCT) data for entrectinib and larotrectinib creates additional uncertainty around their effectiveness. Although historical data might be used to construct a synthetic control arm to the trial arm, it can be highly difficult to ascertain that the patient populations in control and intervention arm are sufficiently comparable.[8] A key issue for histology-independent treatments is that all patients in the trial have tumors with a specific genetic marker, whereas the available historical data is likely for a mixed patient population with tumors with and without the genetic marker. Given that genetic markers may affect disease prognosis, historical data from patients without the genetic marker should be corrected for the prognostic value of the genetic marker.[9]

Third, for most tumor types the standard of care (SoC) does not include testing for NTRK fusions, meaning that the introduction of TRK inhibitors would also require the introduction of NTRK testing. Evaluating the cost-effectiveness of a new treatment requires a comparison between the new situation, in which the intervention is implemented, and the current situation.[9] To accurately reflect the new situation, all changes to the care pathway that are needed to identify and treat eligible patients need to be accounted for. That is, the health and cost consequences associated with introducing NTRK testing should be included in the cost-effectiveness analysis of TRK inhibitors. Various topics are to be considered when modeling tests, including the expected testing procedures in clinical practice, test properties (eg, sensitivity and specificity), and mortality during the testing phase.[9]

In this article, we estimate the cost-effectiveness of entrectinib compared with SoC in cancer patients in The Netherlands. We compare a strategy in which patients are tested for NTRK fusions and receive subsequent treatment (entrectinib for NTRK+ patients and SoC for NTRK-negative [NTRK–] patients) to a strategy in which no additional testing

is used and all patients receive SoC. We illustrate how some of the challenges arising from single-arm trial data may be addressed (the second challenge mentioned above) and how testing pathways can be incorporated in cost-effectiveness analyses (third challenge).

Methods

Intervention and Comparator

The intervention comprised NTRK gene fusion testing for all patients with locally advanced or metastatic solid tumors followed by treatment with entrectinib in NTRK+ patients and SoC in NTRK– patients. Patients in the comparator arm were not tested for NTRK fusions and all patients received SoC. National tumor-specific treatment guidelines were used to identify the treatments provided in SoC for each tumor type (Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). We included treatments indicated for patients with locally advanced or metastatic solid tumors who had already received (at least) 1 systemic anticancer therapy.[2,10] Experimental treatments (ie, treatments outside current clinical guidelines) were excluded.

Larotrectinib also targets oncology patients with NTRK gene fusions and could be an alternative to entrectinib. Nevertheless, we were unable to include both pharmaceuticals in the model because of key differences in the trial populations, including the distribution of tumor types and the presence of pediatric patients in the larotrectinib trial but not the entrectinib trial. Publicly available data are insufficient to adjust the estimated effectiveness in the entrectinib and larotrectinib trials for differences in the trial populations. We opted to include entrectinib in our analysis because we had data available on the fitted distribution for overall survival (OS) and treatment discontinuation for entrectinib, whereas this data was not available for larotrectinib.

Study Population

The study population included adult patients with locally advanced or metastatic solid tumors who have received one or more lines of treatment and are willing to undergo further testing and treatment. Although entrectinib also received EMA approval for pediatric patients ≥ 12 years, there were no pediatric patients included in the ALKA-372-001, STARTRK-1, and STARTRK-2 trials for entrectinib (hereafter called entrectinib trials, N = 54), so we opted to focus on the adult population. Based on the patient characteristics in the trials, patients were assumed to be 58 years old upon entering the model, with 59% of patients being female.[6] The included cancers were breast, bile duct (ie, cholangiocarcinoma), colorectal, endometrial, ovarian, pancreatic, and thyroid

cancer, as well as neuroendocrine tumor, non-small cell lung cancer, sarcoma, secretory carcinoma of the breast, and secretory carcinoma of the salivary gland.

To model the testing phase of patients, the tumor types were categorized into 3 groups based on a consensus report of Dutch experts, which outlines the envisioned NTRK testing policy in Dutch clinical practice[11]: (1) tumor types with high NTRK fusion prevalence (> 90%), (2) tumor types with low NTRK fusion prevalence but wild-type TRK protein expression, and (3) tumor types with low NTRK fusion prevalence and no/very little wild-type TRK protein expression (Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006).

The assumed distribution of patients across tumor types in the model was based on the distribution in the pooled data set from the entrectinib trials. As the trials included only NTRK+ patients, the trial distribution of tumor types was combined with the tumor-specific NTRK prevalence to obtain the distribution of tumor types among the total group of patients eligible for NTRK fusion testing (ie, both NTRK+ and NTRK– patients; Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006 also shows the tumor distribution in Dutch clinical practice among our study population, as observed in data from the Hartwig Medical Foundation (HMF), which cover 44 hospitals in The Netherlands. Nonetheless, because we are using the pooled effectiveness estimate from the entrectinib trials, we used the tumor distribution based on the entrectinib trials in our model.

Model Structure

The model consists of a decision tree (reflecting the testing phase) and a microsimulation (reflecting treatment) as shown in Figure 1. We used a lifetime time horizon and a cycle length of 1 month in the microsimulation. Analyses were performed from a Dutch societal perspective and effects and costs were discounted at 1.5% and 4%, respectively.[12]

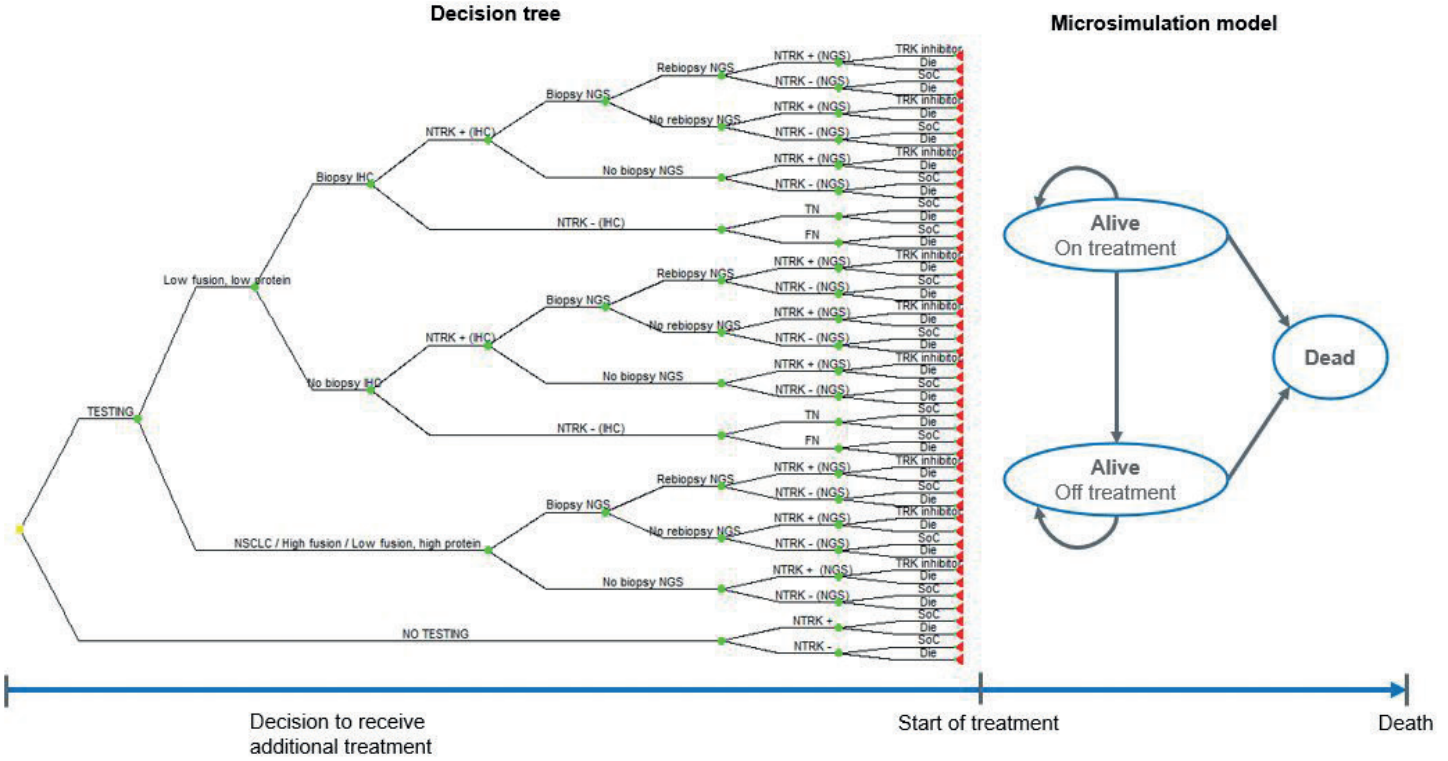


Figure 1 Model structure

FN indicates false negative; IHC, immunohistochemistry; NGS, next-generation sequencing; SoC, standard of care; TN, true negative.

Decision Tree

Patients enter the model in a decision tree that compares "NTRK fusion testing" with a "no testing" strategy.

For patients receiving testing, the decision tree reflects the period from the decision to test for potential eligibility for entrectinib until the start of treatment. All patients were tested using immunohistochemistry (IHC) and/or next-generation sequencing of RNA (RNA-NGS). Patients with tumor types in groups 1 (high NTRK fusion prevalence) and 2 (wild-type TRK protein expression) were tested only through RNA-NGS. Patients in group 3 (low NTRK fusion prevalence, low TRK protein expression) first received an IHC test to identify those with elevated levels of TRK proteins. Patients who tested positive on the IHC test subsequently underwent confirmatory RNA-NGS testing. The latter strategy may save costs, because IHC tests tend to be much cheaper than RNA-NGS tests. Nevertheless, using IHC as a first screening tool has little added value in groups 1 and 2 because most patients will test positive on the IHC test, either because of the high prevalence of NTRK gene fusions (group 1) or wild-type TRK protein expression (ie, TRK protein expression that is not due to an NTRK fusion) (group 2). See Appendix Figures 1 and 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006 for a graphical representation of the different testing strategies.

Various probabilities are captured in the decision tree, including the probability that a new tumor biopsy needs to be performed to enable NTRK testing. Because biopsies sometimes fail to include sufficient tumor tissue and laboratory processes can fail, we also included the probability that a rebiopsy is required. Given that IHC tests generally can be performed using small amounts of tissue, the probability of rebiopsies was included only for RNA-NGS tests. Because of a relatively high short-term mortality in our study population, we also incorporated the probability of death during the testing phase. We assumed that all patients received (tumor specific) SoC during the testing period.

Microsimulation Model

Patients who survived the testing period subsequently entered the microsimulation model, where they received entrectinib if the test results were positive and SoC if negative. We used an individual-level state-transition model with 3 health states: "alive and on treatment," "alive and off treatment," and "dead."

Note that in group 3, patients with a negative test result on the IHC test did not receive any further testing. Patients with a false negative result (ie, undetected NTRK+ patients) were assigned the mortality rates for NTRK+ patients receiving SoC. Similarly, although all patients receive SoC in the "no testing" strategy, the NTRK status of patients was

tracked in the model so that the appropriate probabilities for treatment discontinuation and survival could be applied.

Model Parameters

The model input parameters are summarized in Table 1[10,11,12,13,14,6,7,8,9] and described in more detail below

Table 1 Input parameters

Input parameter	Value	Source	OWSA	PSA
Testing phase				
NTRK prevalence (by tumour type)	Table S2		+/-20%	Beta
IHC sensitivity (by tumour type)	Table S3		+/-20%	Beta
IHC specificity (by tumour type)	Table S3		+/-20%	Beta
Probability biopsy NGS	0.098	[37]	+/-20%	Beta
Probability rebiopsy NGS	0.159	[37]	+/-20%	Beta
Probability biopsy IHC	equal to NGS	Assumption	+/-20%	Beta
Duration waiting time - Biopsy requested until biopsy done - Re-biopsy requested until re-biopsy done - (Re-)biopsy done until IHC test results - (Re-)biopsy done until RNA-NGS results - Final test results available until start of treatment	10.5 days 14 days 3.5 days 10 days 5.5 days	Expert judgement by 3 clinical geneticists and 1 oncologist (mean estimate was used)	+/-20%	Multivariate normal
Cost biopsy	Table S6		+/-20%	Gamma*
Cost RNA-NGS	Table S6		+/-20%	Gamma*
Costs IHC	Table S6		+/-20%	Gamma*
Treatment phase				
Starting age in years	58	[38]	+/-20%	Normal*
Proportion females	0.59	[38]	+/-20%	Beta
OS NTRK- (by tumour type)	Table S5	HMF	-	Asymptomatic normal[39]
TTD NTRK- (by tumour type)	Table S5	HMF	-	Asymptomatic normal[39]
OS entrectinib	exponential	Roche	-	Multivariate normal

Table 1 Input parameters (continued)

Input parameter	Value	Source	OWSA	PSA
TTDisc entrectinib	exponential	Roche	-	Multivariate normal
HR NTRK+ OS adjusted	1.44	HMF	95% CI	Lognormal
HR NTRK+ TTDisc adjusted	1.37	HMF	95% CI	Lognormal
Tumour distribution†	Table S2	[38]	-	-
Utilities - time to death	-	[23]	+/-20%	Multivariate normal
Informal care use	-	Santi et al. (submitted)	+/-20%	Multivariate normal
Treatment costs SoC (per month)	Table S7		+/-20%	Gamma*
Treatment costs entrectinib (per month)	€5,913	Roche	+/-20%	Gamma*
Adverse event costs (per event)	Table S8		+/-20%	Gamma*
Informal care costs (per hour)	€14.77	[40]	+/-20%	Gamma*
Hospital costs related and unrelated to cancer per year (not year before death) at age 58	€4,453	PAID	+/-20%	-
Hospital costs related and unrelated to cancer in the year before death at age 58	€58,064	PAID	+/-20%	-

CI indicates confidence interval; HMF, Hartwig Medical Foundation; HR, hazard ratio; IHC, immunohistochemistry; NGS, next-generation sequencing; NTRK–, NTRK negative; NTRK+, NTRK positive; OS, overall survival; OWSA, one-way sensitivity analysis; PAID, Practical Application to Include Disease Costs; PSA, probabilistic sensitivity analysis; SoC, standard of care; TTD, time-to-treatment-discontinuation.

* Assumption SD is 10% of the mean.

† Calculated using the following formula: tumor type distribution before test = tumor type distribution in entrectinib trials / NTRK prevalence, rescaled to proportions.

Transition Probabilities

Decision tree

Estimates from the literature were used to obtain input values for various parameters, including the probabilities of patients needing biopsies (9.8%) and rebiopsies (15.9%) to enable testing[13] and the test properties of IHC tests. Tumor-specific sensitivity and specificity of the IHC test were used and varied between 73% to 100% and 50% to 100%, respectively (Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). The sensitivity and specificity of RNA-NGS were assumed to be 100%.[15]

Waiting times per element of the testing pathway were estimated by 4 experts (3 clinical geneticists and 1 oncologist). The total waiting time in each arm was determined by the

number of tests, biopsies, and rebiopsies performed and varied between 1 and 8 weeks. The probability to die during the testing phase was based on the estimated waiting times combined with tumor- and NTRK status-specific estimates of weekly mortality rates. The latter were derived from the HMF database (Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006).

Microsimulation model

OS and time-to-treatment discontinuation (TTD) in NTRK+ patients receiving entrectinib were based on the entrectinib trials. Because of the small sample size of the entrectinib trials and the resulting lack of reliable tumor-specific estimates, we used single parametric survival function data for all tumor types. We used the best fitting distribution (exponential) and its coefficients from Roche's model for reimbursement submission.[16]

As mentioned above, the entrectinib trials were single-arm trials. To be able to assess the relative effectiveness of entrectinib compared with SoC, we created a synthetic comparator arm. We used the HMF database, containing data from cancer patients with locally advanced or metastatic tumors who had genomic profiling between 2012 and 2020 (CPCT-02 study – NCT01855477).[10,17] Patients were eligible for enrolment in the CPCT-02 study if (1) their age ≥ 18 years, (2) they had a locally advanced or metastatic solid tumor, (3) they had an indication for new line of systemic treatment with registered anticancer agents, and (4) performing a tumor biopsy was safe according to the treating physician.[10] There are no publicly available patient-level data from the entrectinib trials, meaning that statistical methods to match the study populations from the HMF database and the entrectinib trials (eg, propensity score matching) could not be applied.[18]

OS and TTD in NTRK− patients receiving SoC were based on data from 1,596 NTRK− patients who received SoC (Appendix Fig. 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). For each tumor type, parametric distributions were fitted to data on OS and TTD, using the Akaike Information Criterion to determine the parametric distribution with the best fit (Appendix Table 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). Subsequently, tumor-specific monthly transition probabilities to "death" and "discontinuation of treatment" were extracted to be used in the model. The transition probabilities to discontinuation of treatment (entrectinib or SoC) and death were estimated independently.

To estimate transition probabilities to "death" and "treatment discontinuation" in NTRK+ patients receiving SoC, we applied a hazard ratio (HR) reflecting the prognostic value of carrying an oncogenic NTRK gene fusion to the OS and TTD estimates for NTRK− patients.

Prognostic value of NTRK fusions

We used the HMF database to estimate the prognostic value of an NTRK fusion. Patients in the database were classified into 2 cohorts: NTRK+ and NTRK–. A subgroup of the NTRK– cohort was matched to the patients in the NTRK+ cohort, using the optimal matching method.[19] Within each tumor type, patients were matched based on baseline patient characteristics, including age, sex, number of previous lines of therapy, and year of biopsy, using a ratio of 1:4 (NTRK+:NTRK–). A total of 24 NTRK+ patients were matched with 96 NTRK– patients, with a successful covariate balance between the 2 groups. OS and TTD analyses were performed using the Kaplan-Meier method (Fig. 2) and Cox regressions, with the date of the first postbiopsy treatment as the index date and age, sex, and number of previous lines of treatment as covariates. If a subject was known to be alive before the cutoff date, the subject was censored at the last known alive date. A HR of 1.44 (95% confidence interval [CI] 0.81-2.56) was estimated for OS among NTRK+ patients and 1.37 (95% CI 0.86-2.18) for TTD (Appendix Fig. 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). This suggests that in SoC, NTRK+ patients have worse prognosis than NTRK– patients. We assumed that the HRs were constant across tumor types.

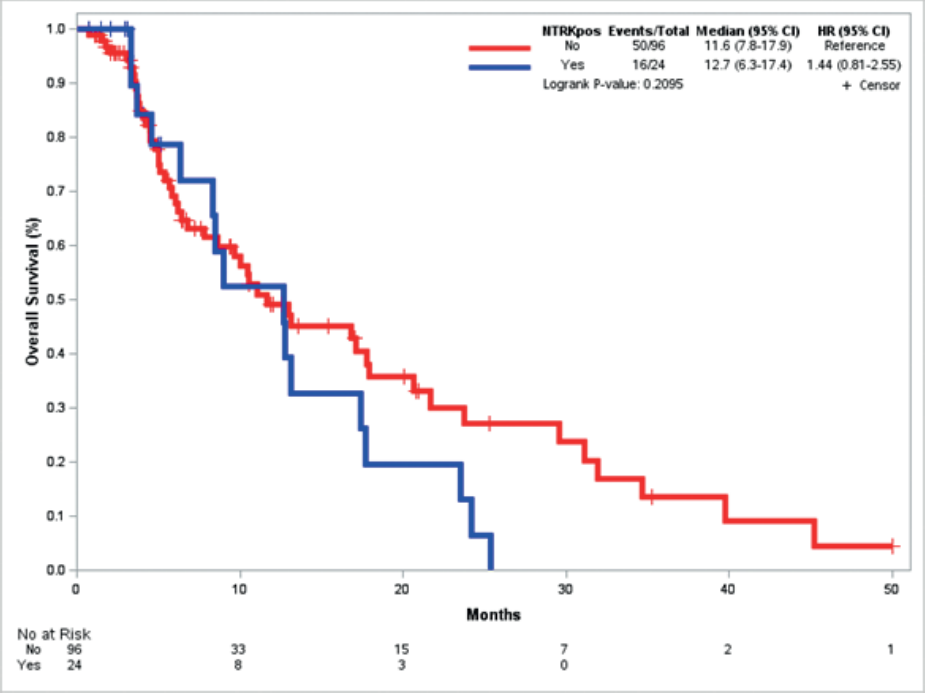


Figure 2 Prognostic value of NTRK fusion on OS

CI indicates confidence interval; HR, hazard ratio; OS, overall survival; pos, positive.

Costs

All costs are expressed in 2020 euros.

Decision tree

For each tumor type, we calculated the mean costs of biopsy, IHC testing, and RNA-NGS testing. Cost variations across tumor types were caused by differences in the type of biopsy needed (eg, more resources are needed for a lung biopsy than for a skin biopsy) and differences in price-setting among the main treatment centers for the various tumor types. The cost of an IHC test varied between €67 and €328 and the cost of an RNA-NGS test varied between €870 and €2,137 (Appendix Table 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006).

Microsimulation model

We included the costs of oncology drugs and their administration (including admissions at a day care unit if needed). For entrectinib, the monthly treatment cost was €5,913 for all tumor types. The treatment costs for the various SoC treatments were combined into an average cost per tumor type with a maximum of €5,501 per month for colorectal cancer (Appendix Table 7 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). In addition, costs for the treatment of adverse events (AEs) were included. Because of a lack of data on AEs with entrectinib, we assumed that the occurrence of AEs when receiving entrectinib was equal to AE occurrence in SoC. The tumor-specific prevalence of AEs was multiplied by the cost of treating the AEs, with input values for both variables based on estimates from the literature (Appendix Table 8 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). The costs of AEs were applied for patients in the "alive and on treatment" health state.

As per the Dutch health technology assessment (HTA) guideline, we also included the costs of non-hospital care related to cancer (eg, psychosocial care, home nursing, and general practitioner check-ups), as well as hospital and non-hospital care unrelated to cancer. We obtained age-dependent cost estimates through the "Practical Application to Include Disease Costs" tool.[20] At baseline, the total annual other healthcare costs derived from "Practical Application to Include Disease Costs" (ie, excluding drugs, administration of drugs, and AE costs) were €4,453. In the year before death, they increased to €58,064.

Regarding societal costs, we included informal care costs but no productivity costs. Because the patients in our study population are in an advanced stage of cancer, we assumed that they are already out of the workforce upon entering our model meaning that no additional productivity losses occur (ie, the friction period has already passed).[21] Informal care estimates were provided by a regression analysis that estimated the impact of proximity to death on the use of informal care, correcting for covariates, such as age

and sex, using the Survey of Health, Ageing, and Retirement in Europe data.[22] The analysis provided the probability of using informal care, as well as the amount of hours used. We valued informal care at the standard rate recommended by the national HTA guideline (€14.77 per hour).[12]

Utilities

Based on previous research findings that the quality of life decreases as patients approach death, we incorporated patient utility in our model as a function of time to death.[23] We used as model inputs the regression coefficients from a study that estimated the relationship between proximity to death and utility.[23] In the study, proximity-to-death values were obtained using OS functions and utility values based on the SF-6D method were used. Age and sex were also included as covariates, allowing us to include age- and sex-specific patient utility in our model.[23] We included the utility during the testing pathway, as well as during treatment.

Analyses

The decision tree and the microsimulation model were programmed in line with the Decision Analysis in R for Technologies in Health (DARTH) modeling framework in R 3.6.1 using RStudio 1.2.1335 (RStudio, Boston, MA).[24,25,26]

Base-Case Analysis

The base-case analysis reflects the cost-effectiveness of testing cancer patients with locally advanced or metastatic solid tumors for NTRK gene fusions and subsequently treating NTRK+ patients with entrectinib and NTRK– patients with SoC compared with no testing and treating all patients with SoC. The incremental net monetary benefit (INMB) was calculated using a cost-effectiveness threshold of €80,000 per quality-adjusted life-year (QALY), which is the recommended value by Dutch HTA guidelines, given the calculated disease burden.[27,28] We opted to simulate 5,000 patients after evaluating the stability of model outcomes at various cohort sizes (Appendix Fig. 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006).

Scenario Analyses

The first scenario analysis aimed to evaluate the impact of testing costs on the cost-effectiveness of implementing entrectinib by excluding the costs of NTRK testing. The second scenario excluded the costs, as well as the health effects resulting from the testing pathway (ie, the probability to die and the QALYs during the waiting time). This scenario reflects a setting in which the patients in our study population would not need any NTRK testing, for example, because RNA-NGS testing is part of standard practice and has already been done at an earlier stage. The third scenario takes the approach

that is common in economic evaluations of targeted treatments, which is to only include the patients who carry the targeted genetic marker and to disregard the costs and health effects of testing.[9] We also performed the base-case analysis from a healthcare perspective.

As mentioned in section "Study Population", a stratified testing protocol (first IHC, then RNA-NGS for patients who test positive) is used for patients in group 3 (tumor types with low NTRK fusion prevalence and no/little wild-type TRK protein expression), which is the largest group. Nevertheless, RNA-NGS has much better test sensitivity and specificity than IHC.[29] If costs were not considered, providing RNA-NGS to all patients would therefore likely be preferred. RNA-NGS is a relatively new technology and in many settings is not yet part of standard care. As RNA-NGS becomes more widespread and perhaps further technological improvements are made to achieve efficiency gains, its cost may decrease. We therefore investigated at what price of RNA-NGS the provision of RNA-NGS to all patients would become cost-effective (ie, renders 0 INMB).

Sensitivity Analyses

Parameter uncertainty was tested using univariate sensitivity analysis and probabilistic sensitivity analysis (PSA). In the univariate analysis, model parameters were independently varied over the extremities of the 95% CI and, when this was not available, by a 20% increase/decrease from the parameter value in the base case. In the PSA, all parameters were varied simultaneously according to predefined distributions (Table 1[6,12,13,14]). The model was run with 1,000 iterations while sampling 1,000 patients.

Budget Impact Analysis

A 5-year budget impact was estimated by multiplying the annual incremental healthcare costs per patient tested in the first 5 years with the expected number of patients tested. The number of patients tested each year in The Netherlands was determined by multiplying the number of expected NTRK+ patients (N = 90)[30,31] by the number of patients that need to be tested to identify 1 NTRK+ patient (as derived from our model results).

Results

Base-Case Analysis

The results of the base-case analysis are presented in Table 2 and Appendix Table 9 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006. They showed that testing for NTRK fusions and treatment with entrectinib in NTRK+ patients and SoC in NTRK− patients was associated with a QALY gain of 0.0043 at an increased cost of €732 per patient as compared with no testing and SoC for all patients. The incremental cost-effectiveness ratio (ICER) was €169,957/QALY and the INMB was −€388.

Scenario Analyses

Table 2 shows the results of scenario analyses with or without testing costs and consequences. Not including testing costs had a large impact, whereas not including mortality and QALYs during waiting time only had a minor impact on the cost-effectiveness results. In the third scenario analysis, in which only NTRK+ patients were considered and the testing phase was disregarded, the ICER of entrectinib versus SoC was €38,563/QALY. The results from the healthcare perspective were similar to the results in the base case (Appendix Table 10 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006). We estimated that the cost of RNA-NGS would have to be reduced by 90%, to €186, before testing all eligible patients with RNA-NGS would be cost-effective.

Sensitivity Analyses

The results of the univariate analysis are presented in Appendix Figure 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006. The specificity of the IHC test had the largest impact on the INMB, followed by the cost of IHC testing and the HR for the prognostic value of NTRK fusions on OS. Improvements in the specificity of IHC tests (ie, fewer false positives), decreases in the costs for IHC tests, or increases in the HR for the prognostic value of NTRK fusions on OS (ie, worse prognosis of NTRK+) led to increased INMB.

The results of the PSA are presented in the cost-effectiveness plane in Figure 3 and the cost-effectiveness acceptability curve in Appendix Figure 7 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.08.006. Mean incremental costs of all iterations were €696 and mean incremental effects were 0.0037, resulting in an INMB of −€399 and an ICER of €187,681/QALY. The probability of "Testing + entrectinib/SoC" being cost-effective compared with "No testing + SoC" was 0.2% at a threshold of €80,000 per QALY.

Table 2 Base-case analysis and scenario analyses with or without testing costs and consequences

Base case or scenario analysis	Comparator: Base case, scenario 1 and scenario 2	Intervention: Base case (including testing costs and consequences)	Intervention: Scenario 1 (excluding testing costs)	Intervention: Scenario 2 (excluding test costs and consequences)	Comparator: Scenario 3	Intervention: Scenario 3
Patient population	NTRK+ and NTRK- patients	NTRK+ and NTRK- patients	NTRK+ and NTRK- patients	NTRK+ and NTRK- patients	NTRK+ patients	NTRK+ patients
Treatment	No NTRK testing, SoC for all patients	NTRK testing, entrectinib for NTRK+ patients, SoC for NTRK- patients	NTRK testing, entrectinib for NTRK+ patients, SoC for NTRK- patients	NTRK testing, entrectinib for NTRK+ patients, SoC for NTRK- patients	SoC	Entrectinib
Total costs (in €)	74,071	74,803	74,227	74,244	71,131	129,240
Total effects (in QALYs)	0.985	0.989	0.989	0.989	0.679	2.186
Incremental costs (in €)		732	156	172		58,109
Incremental effects (in QALYs)		0.0043	0.0043	0.0045		1.507
INMB (in €)		-388	188	161		69,945
ICER (in €/QALY)		169,957	36,290	38,658		38,563
Proportion of patients with entrectinib treatment (%)	0	0.28	0.28	0.30	0	100

ICER indicates incremental cost-effectiveness ratio; INMB, incremental net monetary benefit; NTRK–, NTRK negative; NTRK+, NTRK positive; QALY, quality-adjusted life-year; SoC, standard of care.

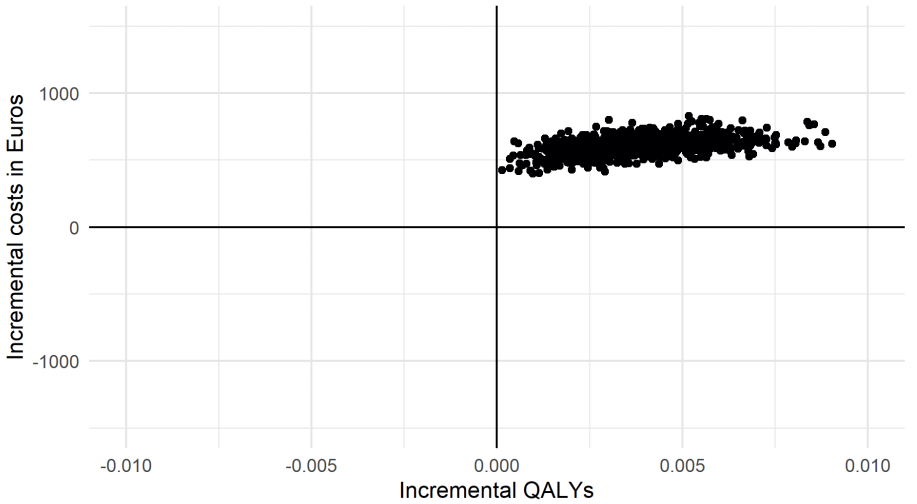


Figure 3 Cost-effectiveness plane of "Testing + entrectinib/SoC' versus 'No testing + SoC."
QALY indicates quality-adjusted life-year; SoC, standard of care.

Budget Impact Analysis

The proportion of patients who received entrectinib from all patients that were tested in our model was 0.28%. Combined with the expected number of patients treated with entrectinib (N = 90), this results in 31,630 patients being eligible for NTRK testing in The Netherlands annually. The 5-year incremental budget impact of testing and treatment with entrectinib in NTRK+ patients was €93 million, with testing costs making up 82% of the total budget impact (Table 3).

Table 3 Results of budget impact analysis (in €)

Entrectinib	Annual test costs	Annual healthcare costs	Annual total costs
year 1	15,249,931	810,033,052	825,282,982
year 2	15,249,931	1,317,036,538	1,332,286,468
year 3	15,249,931	1,661,785,334	1,677,035,265
year 4	15,249,931	1,800,859,346	1,816,109,277
year 5	15,249,931	1,827,257,744	1,842,507,675
SoC	**Annual test costs**	**Annual healthcare costs**	**Annual total costs**
year 1	-	807,719,086	807,719,086
year 2	-	1,314,096,035	1,314,096,035
year 3	-	1,658,446,610	1,658,446,610
year 4	-	1,796,990,043	1,796,990,043
year 5	-	1,822,922,722	1,822,922,722
Incremental	**Annual test costs**	**Annual healthcare costs**	**Annual total costs**
year 1	15,249,931	2,313,966	17,563,896
year 2	15,249,931	2,940,503	18,190,433
year 3	15,249,931	3,338,724	18,588,655
year 4	15,249,931	3,869,303	19,119,234
year 5	15,249,931	4,335,022	19,584,953

SoC indicates standard of care.

Discussion

Summary of Results

Our results showed that incorporating the consequences of NTRK testing has a large impact on the cost-effectiveness of implementing the histology-independent treatment entrectinib, to the extent that it would alter the reimbursement decision. Excluding the costs and health effects associated with NTRK testing reduced the ICER from €169,957/QALY to €38,563/QALY. The difference is primarily owing to the rarity of NTRK gene fusions, meaning that a large number of patients need to be tested to identify the

few patients that are eligible for entrectinib treatment. This means that few patients experience QALY gains from NTRK testing (only the NTRK+ patients who experience better health outcomes with entrectinib treatment than SoC), while additional costs (for the tests) are added for all patients.

The estimated five-year budget impact of testing (€93 million) was based on the testing capacity that is needed when all eligible patients are tested. Nevertheless, the testing practice for NTRK fusion-positive cancers in The Netherlands is still in development.[11] Hence, fewer tests may be conducted in the earlier years of the implementation of NTRK testing and subsequent treatment, which would make the actual budget impact lower than what we estimated.

Strengths

When the introduction of a new drug requires a new diagnostic test to identify the target population, HTA bodies should be fully informed about the cost-effectiveness of the test-treatment combination.[9] Therefore, in this study we considered the entire patient population that would be affected (ie, all patients eligible for NTRK testing) and calculated the (downstream) costs and benefits of NTRK testing and subsequent treatment for patients who tested NTRK+ and NTRK–.

Entrectinib was conditionally approved by the EMA and the Food and Drug Administration based on small single-arm basket trials showing a durable response and longer survival.[32] The single-arm nature of the trial data poses a great challenge to HTA bodies making decisions based on comparative effectiveness and cost-effectiveness. Briggs et al,[33] in a study about the estimation of the counterfactual for tumor-agnostic treatments with only single-arm trial data available, described 3 options: (1) historical controls reported in the literature, (2) previous line of therapy in intervention patients of basket trials, or (3) nonresponders in basket trials. As we did not have access to individual patient data from the entrectinib trials, we had to rely on historical data. Nevertheless, unlike Briggs et al,[33] we used genetic data to adjust the historical data for the prognostic value of carrying oncogenic NTRK gene fusions, therewith likely increasing the accuracy of the estimated comparative effectiveness. Indeed, 2 other studies that estimated an HR for the OS of NTRK+ patients found comparable values (1.44[34] and 1.6,[35] respectively) to the value we estimated (1.44). With our approach, we illustrated how a database with genomic and clinical data can be used to match patients with and without a specific genetic marker and to estimate HRs (for OS and TTD) for the patients with the marker. A similar approach could be used to estimate HRs for patients with different genetic markers. Although residual confounding cannot be ruled out in our analysis because of the small sample size, lack of matching patients on all relevant covariates (eg, Eastern Cooperative Oncology Group performance status), and other limitations, more robust

statistical estimation may be achieved with larger sample sizes (eg, for genetic markers that are less rare) and with data for more (clinical) variables. Nonetheless, RCT data are preferred over historical data for the assessment of comparative effectiveness, and the kind of matching exercise we conducted should only be done when no RCT data are available.

Most health economic models for cancer treatments include the health states "progression-free" and "progression," which require data on progression status in addition to OS. Because appropriate data on progression for patients receiving SoC were not available, we instead used a regression model that estimated the impact of proximity to death on the utility of cancer patients. Although this is not a conventional approach, we believe that linking utility to proximity to death is likely more appropriate (ie, closer to the lived experience of cancer patients) than using singular utility values for progression-free and progressed patients. We therefore consider this approach a strength of our study.

Limitations

Although we estimated tumor-specific effectiveness of SoC, we had to assume that the effectiveness of entrectinib was constant across tumor types, as the small number of observations per tumor type in the entrectinib trials precluded the estimation of reliable tumor-specific effectiveness. Nevertheless, Murphy et al[7] showed heterogeneity in clinical effects across histologies that should be accounted for, for example, by using Bayesian hierarchal models. We present a single ICER in this study, implying an all-or-nothing decision regarding the reimbursement of entrectinib for NTRK+ patients, yet it might be more appropriate to differentiate between tumor types. That is, even when histology-independent therapies receive marketing authorization for all histologies, reimbursement might be warranted only for a subset of histologies because of heterogeneity in the treatment effect. If indeed there is heterogeneity in effectiveness across tumor types, the single ICER estimate we estimated based on the patient population in the entrectinib trials may be biased because the proportional distribution of tumor types in the trial is not fully representative of the distribution of tumor types among eligible patients in clinical practice. Additionally, because of the small number of NTRK+ observations per tumor type in the HMF database, we had to assume that the prognostic value of NTRK fusions was constant across tumor types.

Note that although our analysis does not account for heterogeneity in the treatment effect of entrectinib and the HRs for NTRK gene fusions across histologies, it does account for differences in the treatment costs and effects of SoC (both TTD and OS) across tumor types and uses tumor-specific estimates of testing costs.

We attempted to construct an appropriate comparator arm by estimating outcomes on a similar target population and adjusting for the prognostic value of NTRK fusions. That is, we only considered patients with locally advanced or metastatic disease who had received at least one previous line of treatment, in line with the patient population in the entrectinib trials. We also only considered patients who had one of the tumor types included in the entrectinib trials. We attempted to create a subgroup of HMF patients with a similar average age and percent of females as those in the entrectinib trials, but the subgroup was too small to enable reliable statistical estimation. Despite our efforts, we cannot be sure that the patient populations in the respective arms are fully comparable because we did not have access to patient-level data from the entrectinib trials. Moreover, there may be unobserved differences between the patient populations.

Implications for Decision Making

Considering the impact of testing costs on the cost-effectiveness of entrectinib, payers should focus on policies supporting a reduction in the costs of testing. Furthermore, the uncertainty around the effectiveness of entrectinib leaves HTA bodies with 2 main choices: wait for more (tumor type-specific) evidence or provide coverage through a managed entry agreement between the pharmaceutical company and the healthcare payer. Considering the rarity of NTRK fusions, stronger evidence will likely not become available soon, meaning that waiting for more evidence would leave patients without access to entrectinib for a long time. Therefore, despite concerns about the feasibility of discontinuing reimbursement for medicines once further evidence does not demonstrate their (cost-)effectiveness,[36,37] a managed entry agreement, for example, in the form of coverage with evidence development, may be the best option.[38] Through coverage with evidence development agreements, treatments are temporarily reimbursed while further evidence is collected. After a specified period, the cost-effectiveness of the treatment is re-evaluated using the additional data. Ideally, data on final outcomes, such as survival and quality of life, are collected. When the follow-up time is limited, surrogate outcomes, such as progression-free survival and tumor response rates, are also used. Although, it is of note that surrogate outcomes are not necessarily predictive of final outcomes.[39]

In addition, we found that the uncertainty around the HR for NTRK gene fusions on OS has a large effect on cost-effectiveness outcomes. Because of the low prevalence of NTRK fusions, larger genomic databases (paired with clinical information) are needed to gather sufficiently large numbers of NTRK+ patients to obtain statistically significant results on the prognostic value of NTRK fusions. It may therefore be valuable, for patients with other rare genetic markers as well, if decision-makers invest in expanding genomic data collection, possibly including policies that encourage the linking of existing databases. Finally, although our analysis has focused on entrectinib, larotrectinib has a similar target

population, mechanism of action, and price-setting in The Netherlands. Our finding that the introduction of TRK inhibitor entrectinib appears to not be cost-effective when considering the consequences of introducing NTRK testing but might be cost-effective if the relevant tests would become standard practice, likely also applies for larotrectinib. A recent study that investigated the cost-effectiveness of larotrectinib in The Netherlands and focused on NTRK+ patients without considering the testing phase, comparable with our scenario 3, found an ICER of €41,424/QALY, which is similar to the ICER of €38,563/QALY we estimated for entrectinib in scenario 3.[40]

Conclusions

In conclusion, with the currently available evidence, it seems that entrectinib is not cost-effective compared with SoC. Nevertheless, if genetic testing of cancer patients (including RNA-NGS panels that can identify NTRK gene fusions) becomes standard practice, entrectinib may be cost-effective. Nonetheless, our study results are very uncertain because of data limitations.

References

1. Assessment report VITRAKVI international non-proprietary name: larotrectinib. European Medicines Agency. https://www.ema.europa.eu/en/ documents/assessment-report/vitrakvi-epar-public-assessment-report_en. pdf. Accessed January 20, 2022.
2. Rozlytrec (entrectinib). European Medicines Agency. https://www.ema. europa.eu/en/medicines/human/EPAR/rozlytrek. Accessed January 20, 2022.
3. Rodes Sanchez M, Henderson N, Steuten L. Bridging the gap: pathways for regulatory and health technology assessment of histology independent therapies. https://www.ohe.org/publications/bridging-gap-pathways-regulat ory-and-health-technology-assessment-histology-independent. Accessed January 20, 2022.
4. Drilon A, Laetsch TW, Kummar S, et al. Efficacy of larotrectinib in TRK fusion–positive cancers in adults and children. N Engl J Med. 2018;378(8):731–739.
5. Hong DS, DuBois SG, Kummar S, et al. Larotrectinib in patients with TRK fusion-positive solid tumors: a pooled analysis of three phase 1/2 clinical trials. Lancet Oncol. 2020;21(4):531–540.
6. Doebele RC, Drilon A, Paz-Ares L, et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumors: integrated analysis of three phase 1-2 trials. Lancet Oncol. 2020;21(2):271–282.
7. Murphy P, Claxton L, Hodgson R, et al. Exploring heterogeneity in histologyindependent technologies and the implications for cost-effectiveness. Med Decis Mak. 2021;41(2):165–178.
8. van Rosmalen J, Dejardin D, van Norden Y, Löwenberg B, Lesaffre E. Including historical data in the analysis of clinical trials: is it worth the effort? Stat Methods Med Res. 2018;27(10):3167–3182.
9. Vellekoop H, Huygens S, Versteegh M, et al. Guidance for the harmonisation and improvement of economic evaluations of personalised medicine. Pharmacoeconomics. 2021;39(7):771–788.
10. Priestley P, Baber J, Lolkema MP, et al. Pan-cancer whole-genome analyses of metastatic solid tumors. Nature. 2019;575(7781):210–216.
11. Aerts JGJV, Dinjens WNM, Evers MPJ, et al. Consensus diagnose en behandeling van ntrk-genfusie gerelateerde solide tumoren. https://nfk.nl/media/1/ Consensus_Rapport_Diagnostiek_en_Behandeling_van_NTRK-Genfusie_ Gerelateerde_Solide_Tumoren_14022020.pdf. Accessed January 20, 2022.
12. Guideline for Conducting Economic Evaluations in Healthcare [in Dutch: Richtlijn Voor Het Uitvoeren van Economische Evaluaties in de Gezondheidszorg]. Zorginstituut Nederland (ZIN). https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-hetuitvoeren-van-economische-evaluaties-in-de-gezondheidszorg. Accessed January 20, 2022.
13. Bins S, Cirkel GA, Gadellaa-Van Hooijdonk CG, et al. Implementation of a multicenter biobanking collaboration for next-generation sequencing-based biomarker discovery based on fresh frozen pretreatment tumor tissue biopsies. Oncologist. 2017;22(1):33–40.
14. Jackson C, Metcalfe P, Amdahl J, Warkentin MT, Kunzmann K. Package “Flexsurv”: Flexible Parametric Survival and Multi-state Models. CRAN. https://cran.r-project.org/web/packages/flexsurv/flexsurv.pdf. Accessed January 20, 2022.
15. Williams HL, Walsh K, Diamond A, Oniscu A, Deans ZC. Validation of the OncomineTM focus panel for next-generation sequencing of clinical tumor samples. Virchows Arch. 2018;473(4):489–503.
16. Entrectinib for treating NTRK fusion-positive solid tumors. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/guidance/ ta644/evidence/committee-papers-pdf-8831613997. Accessed January 20, 2022.
17. van Riet J, van de Werken HJG, Cuppen E, et al. The genomic landscape of 85 advanced neuroendocrine neoplasms reveals subtype-heterogeneity and potential therapeutic targets. Nat Commun. 2021;12(1):4612.
18. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. Clin Epidemiol. 2020;12:457–467.
19. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci. 2010;25(1):1–21.

20. van Baal P, Perry-Duxbury M, Bakx P, Versteegh M, van Doorslaer E, Brouwer W. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. Health Econ. 2019;28(1):87–100.
21. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ. 1995;14(2):171–189.
22. Hoogendoorn M, de Groot S, Corroramos I, et al. Deliverable 6.1 Report on Most and Least Cost-Effective SMI's. Appendix 6. – Estimating the Broader Societal Impact of SMIs. Deliverable to the European Commission for COMPAR-EU (Grant Nr: 754936); 2021. To be available at: https://cordis. europa.eu/project/id/754936.
23. Versteegh M, van der Helm I, Mokri H, Oerlemans S, Blommestein H, van Baal P. Estimating quality of life decrements in oncology using time to death [Published online July 6, 2022]. Value Health. https://doi.org/10.1016/j.jval.2 022.06.002
24. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An overview of R in health decision sciences. Med Decis Mak. 2017;37(7):735–746.
25. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MGM, Pechlivanoglou P. Microsimulation modeling for health decision sciences using R: a tutorial. Med Decis Making. 2018;38(3):400–422.
26. Krijkamp EM, Alarid-Escudero F, Enns EA, et al. A multidimensional array representation of state-transition model dynamics. Med Decis Making. 2020;40(2):242–248.
27. Kosteneffectiviteit in de Praktijk. Zorginstituut Nederland. https://www. zorginstituutnederland.nl/publicaties/rapport/2015/06/26/kosteneffectiviteitin-de-praktijk. Accessed January 20, 2022.
28. Versteegh MM, Ramos IC, Buyukkaramikli NC, Ansaripour A, ReckersDroog VT, Brouwer WBF. Severity-adjusted probability of being cost effective. Pharmacoeconomics. 2019;37(9):1155–1163.
29. Koopman B, Kuijpers CCHJ, Groen HJM, et al. Detection of NTRK fusions and TRK expression and performance of pan-TRK immunohistochemistry in routine diagnostics: results from a nationwide community-based cohort. Diagnostics. 2022;12(3):1–15.
30. Forsythe A, Zhang W, Phillip Strauss U, Fellous M, Korei M, Keating K. A systematic review and meta-analysis of neurotrophic tyrosine receptor kinase gene fusion frequencies in solid tumors. Ther Adv Med Oncol. 2020;12: 1758835920975613.
31. Advies potentiële kandidaat voor voorwaardelijke toelating van arotrectinib (Vitrakvi®) bij solide tumoren met een ntrk-genfusie (Procedure: Weesgeneesmiddelen, Conditionals En Exceptionals). Zorginstituut Nederland (ZIN). https://www.zorginstituutnederland.nl/publicaties/adviezen/2021/03/10/advies-over-kandidaat-voorwaardelijke-toelating-larotrectinib-vitrakvi. Accessed January 20, 2022.
32. Ardini E, Siena S. Entrectinib approval by EMA reinforces options for ROS1 and tumor agnostic NTRK targeted cancer therapies. ESMO Open. 2020;5(5): e000867.
33. Briggs A,Wehler B,Gaultney JG, et al. Comparison of alternativemethods to assess the cost-effectiveness of tumor-agnostic therapies: a triangulation approach using larotrectinib as a case study. Value Health. 2022;25(6):1002–1009.
34. Bazhenova L, Lokker A, Snider J, et al. TRK fusion cancer: patient characteristics and survival analysis in the real-world setting. Target Oncol. 2021;16(3):389–399.
35. Demetri GD, Peters S, Hibbar DP, et al. Characteristics and outcomes of patients (pts) with NTRK fusion-positive (NTRK1) metastatic/locally advanced (LA) solid tumors receiving non-TRK inhibitor (TRKi) standard of care (SoC), and prognostic value of NTRK fusions in clinical practice. Ann Oncol. 2021;32(suppl 5):S399.
36. van de Wetering EJ, van Exel J, Brouwer WB. The challenge of conditional reimbursement: stopping reimbursement can be more difficult than not starting in the first place. Value Health. 2017;20(1):118–125.
37. McCabe C, Chilcott J, Claxton K, et al. Continuing the multiple sclerosis risk sharing scheme is unjustified. BMJ. 2010;340:c1786.
38. Wenzl M, Chapman S. Performance-Based Managed Entry Agreements for New Medicines in OECD Countries and EU Member States: How They Work and Possible Improvements Going Forward. OECD. https://www.oecd.org/ health/health-systems/pharma-managed-entry-agreements.htm. Accessed January 20, 2022.

39. Gyawali B, Hey SP, Kesselheim AS. Assessment of the clinical benefit of cancer drugs receiving accelerated approval. JAMA Intern Med. 2019;179(7):906–913.
40. Michels RE, Arteaga CH, Peters ML, Kapiteijn E, Van Herpen CML, Krol M. Economic evaluation of a tumor-agnostic therapy: Dutch economic value of larotrectinib in TRK fusion-positive cancers. Appl Health Econ Health Policy. 2022;20(5):717–729

Supplementary materials

Supplementary materials can be found at https://doi.org/10.1016/j.jval.2022.08.006

Contents

CHAPTER 6

Cost-effectiveness of alternative NTRK-testing strategies in three countries

Vellekoop H, Huygens S, Versteegh M, Szilberhorn L, Zelei T, Nagy B, Koleva-Kolarova R, Wordsworth S, Rutten-van Mölken M, HEcoPerMed consortium.

Cost–effectiveness of Alternative NTRK Testing Strategies in Cancer Patients Followed by Histology-Independent Therapy with Entrectinib: an Analysis of Three European Countries. Personalized Medicine. 2023;20(4):321-38.

Abstract

Aim

To explore variations in the cost-effectiveness of entrectinib across different testing strategies and settings.

Methods

We evaluated four testing strategies where adult cancer patients received entrectinib if they tested positive for neurotrophic tyrosine receptor kinase (NTRK) gene fusions, compared to “no testing” and standard of care (SoC) for all patients.

Results

Immunohistochemistry (IHC) for all patients, followed by ribonucleic acid-based next-generation sequencing (RNA-NGS) after a positive result was the optimal strategy in all included countries. However, the incremental net monetary benefit (INMB) compared to SoC was negative in all countries, ranging between int€ –206 and –404. In a subgroup analysis with only NTRK-positive patients, INMB was int€ 8,405 in England, int€ –53,088 in Hungary, and int€ 54,372 in the Netherlands.

Conclusion

Using the cost-effectiveness thresholds recommended by national guidelines, none of the testing strategies were cost-effective compared to “no testing”. The implementation of entrectinib is unlikely to become cost-effective in Hungary, due to the large cost difference between the entrectinib and SoC arms, while there might be more potential in England and the Netherlands.

Introduction

In a move toward more precise cancer care, several therapies targeting specific genetic tumour markers have emerged onto the market. Among these are histology-independent (also called "tumour-agnostic") therapies, which are prescribed solely based on genetic markers of the tumour, without regard for its tissue of origin. The first histology-independent treatments to receive approval from the European Medicines Agency (EMA) and the Food and Drugs Administration (FDA) were larotrectinib and entrectinib.(1, 2) Both are prescribed for tumours with neurotrophic tyrosine receptor kinase (NTRK) gene fusions (which causes overexpression of TRK proteins) and both are inhibitors of TRK proteins.(3-6)

The advent of larotrectinib and entrectinib has challenged reimbursement authorities and health economists tasked with evaluating their cost-effectiveness. The main challenges include uncertainty about the pharmaceuticals' efficacy due to the use of single-arm (as opposed to randomised controlled) trials. Because of the lack of control arms in the TRK inhibitor trials, there is insufficient knowledge about the health outcomes under the standard of care (SoC) among patients with NTRK fusions. While historical data could be used to construct a synthetic control arm,(7, 8) historical data likely pool NTRK-positive (NTRK+) and NTRK-negative (NTRK-) patients, given that oncology patients were not tested for NTRK fusions in the past. To assess whether using such pooled historical data is appropriate and to be able to adjust historical data where necessary, knowledge on the prognostic value of carrying an NTRK gene fusion is imperative. Yet limited data on the prognostic value of NTRK fusions is available. In a previous study, where we performed an economic evaluation of entrectinib in the Netherlands, we proposed possible approaches to address these (and other) challenges.(9) In the study, we also tried to accurately model the health and cost consequences of testing eligible cancer patients to identify those with oncogenic NTRK fusions. Because the previous study concentrated on the Netherlands, we modelled the testing strategy that was suggested by a group of Dutch experts.(10) However, other testing strategies are possible for the identification of NTRK+ patients.

In this study, to investigate the potential impact of alternative NTRK-testing strategies on the cost-effectiveness and budget impact of entrectinib, we compare four different NTRK-testing strategies. Additionally, the optimal NTRK-testing strategy may differ between countries due to differences in healthcare systems. The original model reflecting the Dutch setting was therefore also adapted for the English and Hungarian healthcare systems. England and Hungary both have highly centralised single-payer healthcare systems, while healthcare is more decentralised in the Netherlands, with curative care being covered by competing private health insurers.(11) All three countries provide

universal healthcare to their residents but while England and the Netherlands achieve broad coverage, Hungary offers a more limited benefits package.(11-13) Other key differences between the three countries given the context of our study are the available healthcare resources, medical practice patterns, and the infrastructure and price-setting for genomic testing (and subsequent treatment).

Methods

Target population

The target population comprised adult patients with locally advanced or metastatic solid tumours who have received one or more lines of treatment and are willing to undergo further (testing and) treatment.(9) While entrectinib is indicated for all NTRK+ tumours, only the tumour types for which data were available from the entrectinib registration trials(14) were incorporated in our model. The included cancers were breast, bile duct (i.e. cholangiocarcinoma), colorectal, endometrial, ovarian, pancreatic, and thyroid cancer, as well as neuroendocrine tumour, non-small cell lung cancer (NSCLC), sarcoma, secretory carcinoma of the breast, and secretory carcinoma of the salivary gland.

Model structure

The model consists of a decision tree followed by a microsimulation model (see Figure S1). The decision tree reflects the testing phase and covers the period from the decision to test for potential eligibility for entrectinib until the start of treatment. The microsimulation model reflects the time from the start of treatment until death and has a cycle length of one month. A comprehensive description of the original model is published elsewhere.(9)

To align with national health technology assessment (HTA) guidelines, analyses were performed from a healthcare payer perspective for England, healthcare perspective for Hungary, and from a societal perspective for the Netherlands.(15-17) Costs and effects were discounted with the prescribed discount rates in each country: 3.5% for costs and effects in England, 3.7% for costs and effects in Hungary, and 4% for costs and 1.5% for effects in the Netherlands. The nationally recommended cost-effectiveness thresholds (after severity-weighting) were £36,000 (int€ 35,576) for England, 4,921,820 Ft (int€ 21,294) for Hungary, and €80,000 (int€ 69,666) for the Netherlands.

Decision tree to model testing phase

Four NTRK-testing strategies were explored in the decision tree. All strategies included next generation sequencing panels used to screen tumour ribonucleic acid (RNA-NGS) for NTRK fusions and/or immunohistochemistry (IHC) tests to assess TRK protein expression. (4, 18-20) RNA-NGS tests can detect oncogenic NTRK fusions with high sensitivity and

specificity. However, RNA-NGS tends to be expensive. IHC tests are generally more affordable than RNA-NGS tests, but also less accurate. The following testing strategies were included: (1) IHC test for all tumour types, (2) RNA-NGS test for all tumour types, (3) IHC test followed by RNA-NGS in patients with a positive IHC test result for all tumour types, and (4) Stratified test strategies depending on the NTRK fusion prevalence and TRK wild-type protein expression of the tumour types.(21)

In the first two strategies, all patients receive the same test: IHC in strategy 1 and RNA-NGS in strategy 2. Strategy 3 is a sequential strategy, in which patients first undergo an IHC test, followed by RNA-NGS for those with a positive IHC test result. In this strategy the false-positives of the IHC test are identified with the RNA-NGS test (though false-negative results are missed), while costs are saved by not having to perform the more expensive RNA-NGS test for all patients. However, in tumour types with high NTRK fusion prevalence (>90%) or wild-type TRK protein expression, adding an IHC test has little value. In the former the reason is that most patients will test positive (because of the high NTRK prevalence) and still need an additional, confirmatory RNA-NGS test, leading to increased costs. In the latter, there will be many false-positive test results after the IHC test, because IHC testing does not distinguish between wild-type and fusion protein expression. Therefore, strategy 4 uses RNA-NGS tests (without preceding IHC tests) for patients with tumour types with high NTRK fusion prevalence (>90%) or wild-type TRK protein expression (Table S1 shows which tumour types fall in these categories). All other patients in strategy 4 are subject to the sequential testing protocol used in strategy 3.

Microsimulation to model treatment

In the intervention group, patients first enter the decision tree and subsequently (those who survive beyond the duration of the testing phase) enter the microsimulation model. (9) The individual-level state transition model included the health states 'alive and on treatment', 'alive and off treatment', and 'dead'. Patients received entrectinib if they had tested NTRK-positive in the decision tree and were treated with SoC if they had tested NTRK-negative. In the comparator group, patients did not receive testing (i.e. they skipped the decision tree and went straight into the microsimulation model) and were all given SoC.

Model parameters

Table 1 presents the country-specific values for the input parameters. More detail on the data sources and assumptions for the input parameters is given below as well as in the Appendix.

Transition probabilities

Decision tree

The probabilities of patients needing biopsies and re-biopsies to enable testing, as well as the test properties of IHC and RNA-NGS tests (see Table S2) were based on estimates from the literature and assumed to be similar across countries.(9, 22, 23)

The waiting times for the various stages of the testing phase were based on publicly available sources in England and based on expert judgement in Hungary and the Netherlands. Patient mortality during the testing phase was based on tumour-specific estimates of weekly mortality.(9) Although weekly mortality was assumed to be similar across countries, the probability of dying during the testing phase differed between the three countries because of differences in waiting times.

Microsimulation model

Time-to-treatment discontinuation (TTD) and overall survival (OS) in NTRK+ patients receiving entrectinib were based on the entrectinib trials.(9, 14) For NTRK- patients receiving SoC, TTD and OS were estimated using registry data from 1,596 Dutch NTRK-cancer patients (Table S3). Patients from the registry data were included in the TTD and OS estimations only if they had one of the tumour types that were included in the entrectinib trials.(9, 14)

Previous studies suggest that patients with NTRK fusions face a worse disease prognosis than patients without NTRK fusions.(9, 24) To account for this, we applied a hazard ratio (HR) for TTD and OS in NTRK+ patients. That is, TTD and OS in NTRK+ patients receiving SoC were calculated by applying the HRs for NTRK+ patients to the TTD and OS that were estimated for NTRK- patients receiving SoC. See Huygens et al.(9) for more detail. We assumed the HRs were equal across the three countries.

Costs

Costs are reported in the national currency and in international euros (int€), calculated using purchasing power parities (PPP) from OECD.Stat.(25, 26) PPPs were used to account for price level differences between countries. The cost year was 2021.

We included country-specific costs for biopsies, tests, entrectinib, and SoC pharmaceuticals (Tables S4-S6). It was assumed that the treatments provided in England were similar to the treatments described in Dutch guidelines. In Hungary, several treatments provided in England and the Netherlands were not available, hence only the treatments that were available in Hungary were included in the SoC cost calculations. Costs for adverse events in England and Hungary were based on the Dutch estimates

reported by Huygens et al.(9) and converted to 2021 Hungarian Forints and British pounds using PPPs.(27)

Based on research where healthcare costs were found to increase steeply in people's last year of life,(28) separate cost figures for the last year of life were used in our model. The estimated costs in the last year of life were specific to cancer patients and came from Luta et al. (2020) for England and the Practical Application to Include Disease Costs (PAID) for the Netherlands.(28, 29) Because no country-specific data were available for Hungary, we calculated the last-year-of-life costs based on the assumption that the ratio between the last-year-of-life costs and costs in other years was the same in Hungary as in England (because both countries use a healthcare perspective, while a societal perspective is used in the Netherlands).

Unlike in England and Hungary, the health economic guidelines in the Netherlands prescribe the inclusion of healthcare costs unrelated to the disease of interest.(17) The Dutch healthcare costs unrelated to cancer were estimated using PAID.(29) Additionally, informal care costs were included for the Netherlands. Productivity costs were excluded, as patients in our model have an advanced stage of cancer and were assumed to have already been out of the work force, so that no additional productivity losses are faced during the period that the model covers.(30)

Utilities

As previous research has found that quality of life decreases as patients approach death, we incorporated patient utility in our model as a function of proximity to death.(31) We obtained age- and gender-specific utility at different proximities to death from a study that estimated the relationship between proximity to death and SF-6D utility using a linear regression model.(31) The study used quality of life and survival data from a Dutch cancer registry.(31) The estimated overall survival distributions from the microsimulation were used to estimate proximity to death for the patients in our model, after which the corresponding utility values were applied. Because of a lack of appropriate data for England and Hungary, we assumed the utilities to be similar in all three countries.

Table 1 Input parameters

Input parameter	England	Hungary	Netherlands	Source
Purchasing power parity (local currency in $)	0.667865	152.550219	0.757906	OECD.Stat(26)
Purchasing power parity (EU average in $)	0.66			OECD.Stat(26)
Testing phase				
Cost biopsy	£727 (int€ 694)	Ft 5,919 (int€ 27)	€166 (int€ 138)	Table S4
Cost RNA-NGS	£350 (int€ 334)	Ft 300,000 (int€ 1,347)	€1,857* (int€ 1,552)	Table S4
Costs IHC	£150 (int€ 143)	Ft 45,000 (int€ 202)	€426 (int€ 356)	Table S4
NTRK prevalence (by tumour type)	Similar across countries			Table S1
IHC sensitivity and specificity (by tumour type)	Similar across countries			Table S2
RNA-NGS sensitivity and specificity	100%	100%	100%	Assumption
Probability biopsy RNA-NGS	0.098	0.098	0.098	Bins et al.(22) Assumption: similar across countries
Probability rebiopsy RNA-NGS	0.159	0.159	0.159	Bins et al.(22) Assumption: similar across countries
Probability biopsy IHC	equal to RNA-NGS			Assumption
Duration waiting time				EN: NHS England HU: Expert judgement by oncologist NL: Expert judgement by 3 clinical geneticists and 1 oncologist (mean estimate was used)
- Biopsy requested until biopsy done	5 days	14 days	10.5 days	
- Re-biopsy requested until re-biopsy done	5 days	9 days	14 days	
- (Re-) biopsy done until IHC test results	5.5 days	5 days	3.5 days	
- (Re-) biopsy done until RNA-NGS results	21 days	14 days	10 days	
- Final test results available until start of treatment	7 days	17.5 days	5.5 days	

Table 1 Input parameters (continued)

Input parameter	England	Hungary	Netherlands	Source
Treatment phase				
Treatment costs SoC (per month, weighted average)	£3,105 (int€ 2,964)	Ft 393,688 (int€ 1,768)	€2,084 (int€ 1,741)	Table S6
Treatment costs entrectinib (per month)	£5,232 (int€ 4,994)	Ft 2,199,642 (int€ 9,851)	€5,912 (int€ 4,938)	Roche
Adverse event costs (per event, weighted average)	£912 (int€ 870)	Ft 193,732 (int€ 870)	€1,042 (int€ 870)	Table S6
Informal care costs (per hour)	NA	NA	€14.77 (int€ 12.34)	Zorginstituut Nederland(17)
Related non-hospital and unrelated (hospital + non-hospital) healthcare costs per year (except year before death)	NA	NA	€4,453 (at age 58, increases with age) (int€ 3,719)	PAID(29)
Healthcare costs in the last year-of-life	£8,994 (int€ 8,586)	Ft 1,140,365 (int€ 5,121)	€58,064 (at age 58, increases with age) (int€ 48,499)	EN: Luta et al.(28) HU: Assumption that ratio between last-year-of-life cost and cost in other years is same as in EN NL: PAID
Starting age in years	58			Doebele et al.(14)
Proportion females	0.59			Doebele et al.(14)
OS NTRK- (by tumour type)	Similar across countries			Table S3
TTDisc NTRK- (by tumour type)	Similar across countries			Table S3
OS entrectinib	Exponential, similar across countries			Roche
TTDisc entrectinib	Exponential, similar across countries			Roche
HR NTRK+ OS adjusted	1.44			HMF
HR NTRK+ TTDisc adjusted	1.37			HMF
Tumour distribution	Similar across countries			Table S1
Utilities - time to death	Similar across countries			Versteegh et al. (31)
Informal care use	NA	NA	Dependent on time to death	de Groot et al. (53)

*weighted mean because the price of RNA-NGS tests varies across included tumour type

EN= England, EU = European Union, HU = Hungary, IHC = immunohistochemistry, HMF = Hartwig Medical Foundation, HR = hazard ratio, NHS = National Health Service, NL = the Netherlands, NTRK+ = NTRK fusion-positive, OS = overall survival, PAID = Practical Application to Include Disease Costs, RNA-NGS = next-generation sequencing panel of tumour RNA, SoC = standard of care, TTDisc = time to treatment discontinuation

Analyses

Cost-effectiveness analysis

For all three countries, cost-effectiveness analyses were performed for the different testing strategies. An additional analysis was conducted in which only the subgroup of NTRK+ patients were considered and the health and cost consequences of NTRK testing were excluded. The latter analysis reflects a scenario in which NTRK testing is already part of the standard of care (for example as part of broader gene fusion panel testing), so that the introduction of NTRK testing does not need to be considered when evaluating the implementation of entrectinib. To investigate the possible benefits of future price drops for RNA-NGS (which might be expected as further technological improvements are made and the use of gene sequencing becomes more widespread), we also performed an analysis to identify the cost for RNA-NGS at which it becomes cost-effective (i.e. INMB is zero) to provide RNA-NGS to all patients eligible for NTRK testing.

Cost-effectiveness is presented as incremental cost-effectiveness ratios (ICERs) as well as incremental net monetary benefit (INMB). Fully incremental analysis was used for the ICERs, meaning the strategies were ranked by ascending cost and ICERs for each strategy were calculated by comparing it to the next best alternative.(32) The INMB for each of the testing strategies was calculated by comparing it to the "no testing" strategy.

Sensitivity analysis

The effect of uncertainty around our model parameters on our results was assessed using univariate sensitivity analysis and probabilistic sensitivity analysis (PSA). In the univariate analysis, parameter values were varied by a maximum of 20% deviation from the original input value (except for the HRs, for which the values from the estimated 95% confidence interval were used). In the PSA, parameter values were varied simultaneously according to predefined distributions for all three countries (see Table S7). After investigating the stability of the results for varying sample sizes / number of iterations, the univariate analysis was performed with 5,000 patient samples, and the PSA with 1,000 iterations of 1,000-patient samples.

Budget impact analysis

For all three countries, the budget impact of the four NTRK-testing strategies was estimated by multiplying the estimated annual incremental healthcare costs with the expected annual number of patients tested. The number of patients tested each year per country was determined by multiplying the *number of expected NTRK+ patients* with the *number of patients to be tested to identify one NTRK+ patient*.(33) In line with national HTA guidelines, we took a healthcare payer perspective for England, and a healthcare perspective for Hungary and the Netherlands. The time horizon of the budget

impact analyses was 5 years, though additional analyses were performed to align with national HTA guidelines, assuming a 3 and 4-year time horizon for England and Hungary, respectively.(16, 17, 34) We present the incremental budget impact of each of the NTRK-testing strategies compared to SoC in absolute values, and expressed as percentages of cancer care expenditure and cancer care expenditure.(35, 36)

Results

Cost-effectiveness analysis

The outcomes for the four NTRK testing strategies and for the comparator (no NTRK testing, provide SoC to all patients) are provided below for England, Hungary and the Netherlands.

Intermediate outcomes

Table 2 presents, for each testing strategy, the average waiting time between the decision to receive NTRK testing and the start of entrectinib treatment for those who test positive. Also presented are the number of patients receiving entrectinib treatment and the number of false positive and false negative test results (per 100,000 patients tested).

Table 2 Intermediate outcomes per strategy

Strategy	Average waiting time (in days)			Number of false positives (per 100,000 patients tested)	Number of false negatives (per 100,000 patients tested)	Number of patients treated with entrectinib (per 100,000 patients tested)		
	EN	***HU***	***NL***	***all***	***all***	***EN***	***HU***	***NL***
IHC for all	14.7	22.4	8.4	4,938	50	5,123	5,064	5,170
RNA-NGS for all	28.7	36.5	15.6	0	0	323	317	332
IHC then RNA-NGS	15.8	23.2	9.1	0	50	270	269	279
Stratified*	16.7	24.1	9.5	0	46	276	274	285

* In the *Stratified* strategy, direct RNA-NGS testing is used for tumour types with high NTRK prevalence or wild-type TRK protein expression, while immunohistochemistry followed by RNA-NGS after a positive result is used for other tumour types

EN = England, IHC = immunohistochemistry, HU = Hungary, NL = the Netherlands, RNA-NGS = next-generation sequencing panel of tumour RNA

In all three countries, average waiting times are lowest in the *IHC for all* strategy (14.7, 22.4 and 8.4 days for England, Hungary and the Netherlands, respectively) and highest for *RNA-NGS for all* (28.7, 36.5 and 15.6 days, respectively). For all strategies, Hungary has the highest average waiting time, while the Netherlands has the shortest waiting time. The metastatic cancer patients in our model tend to have a short life expectancy, and some patients are expected to die during the testing phase before receiving treatment. Because of this, longer average waiting times result in a smaller number of patients treated with entrectinib, as shown in Table 2.

Because test sensitivity and specificity were assumed to be equal across the three countries, the proportions of false positive and false negative results were also equal across the countries, as reflected in the single columns for these measures in Table 2. Table 2 shows that the *IHC for all* strategy renders a high number of false positives (4,938 per 100,000 patients tested compared to 0 for the other strategies). For all three countries, the 'Number of patients treated with entrectinib' is also much higher for *IHC for all* than for the other strategies. Given the high number of false positives in *IHC for all*, many of the patients identified as NTRK+ and receiving entrectinib do not actually carry oncogenic NTRK fusions and likely do not receive any benefit from entrectinib treatment. Although there are no false positive results in the *IHC then RNA-NGS* and *Stratified* scenarios, there are some false negative results (50 and 46 per 100,000 patients tested, respectively). False negative results mean that some NTRK+ patients are incorrectly identified as NTRK- and wrongfully receive SoC instead of entrectinib.

Cost-effectiveness results

Cost-effectiveness outcomes for the testing strategies are presented in Table 3. Costs and effects are shown, and incremental costs and effects and cost-effectiveness ratios compared to the next best alternative. For dominated strategies (strategies with higher cost yet equal or less QALYs than another strategy), no incremental costs, effects and cost-effectiveness ratio are presented. Nonetheless, incremental net monetary benefit (INMB) is calculated for each testing strategy, comparing the strategy to the *No testing* base case.

The ranking of the strategies is the same for all countries, both in terms of ICERs and INMB. The *No testing* strategy renders the lowest costs and lowest amount of QALYs, while *RNA-NGS for all* has the highest costs and QALYs. For all countries, the *IHC then RNA-NGS* option has the highest INMB. The strategy results in less QALYs than *Stratified* and *RNA-NGS for all*, caused by a higher number of unidentified NTRK+ patients (false negatives) not receiving entrectinib treatment (see Table 2). Nonetheless, the issue of false negatives appeared to be offset by the cost-savings from performing RNA-NGS for a smaller group of patients (i.e. only those who receive a positive IHC result first). However, all estimated ICERs (int€ 89,196 for England, int€ 138,135 for Hungary, and int€ 142,663 for the Netherlands) are above national cost-effectiveness thresholds and all estimated INMB values are negative. This implies that the implementation of NTRK fusion testing and subsequent treatment would cause a net loss to the healthcare system and is not cost-effective.

When considering only the subgroup of NTRK+ patients and focussing only on the treatment (i.e. excluding the cost and health effects associated with testing for NTRK fusions), the implementation of entrectinib is estimated to be cost-effective in England and in the Netherlands. INMB is int€ 8,405 in England, int€ –53,088 in Hungary, and int€ 54,372 in the Netherlands. The incremental cost of treating NTRK+ cancer patients with entrectinib instead of SoC is much higher in Hungary than in England and the Netherlands (int€ 97,525 in Hungary compared to int€ 41,439 in England and int€ 50,603 in the Netherlands).

Finally, we found that in England and Hungary the provision of RNA-NGS to all patients eligible for NTRK testing (and subsequently providing entrectinib to NTRK+ patients) would not be cost-effective even at zero cost for RNA-NGS. In the Netherlands, RNA-NGS would have to reduce by 90%, to int €162, before testing all patients with RNA-NGS would be cost-effective.

Table 3 Cost-effectiveness outcomes for all strategies

Strategy	Cost (int€)			Effect (QALY)			Incremental cost (int€)			Incremental effect (QALY)			ICER (int€)			INMB (vs. SoC)		
	EN	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*
No testing	26,695	17,505	64,503	0.95927	0.95685	0.98459												
IHC then RNA-NGS	27,038	18,031	65,108	0.96311	0.96065	0.98883	343	526	606	0.00385	0.00381	0.00425	89,196	138,135	142,663	-206	-404	-310
Stratified*	27,052	18,108	65,140	0.96317	0.96070	0.98890	14	-	32	0.00006	-	0.00006	242,668	Extendedly dominated	502,431	-218	-480	-338
RNA-NGS for all	27,248	19,113	66,298	0.96384	0.96132	0.98962	196	1,082	1,158	0.00067	0.00066	0.00072	293,640	1,629,295	1,834,617	-391	-1,465	-1,445
IHC for all	29,077	22,329	67,472	0.96353	0.96114	0.98907	-	-	-	-	-	-	Dominated	Dominated	Dominated	-2,231	-4,687	-2,657

* In the *Stratified* strategy, direct RNA-NGS testing is used for tumour types with high NTRK prevalence or wild-type TRK protein expression, while immunohistochemistry followed by RNA-NGS after a positive result is used for other tumour types

EN = England, ICER = incremental cost-effectiveness ratio, IHC = immunohistochemistry, int€ = international euros, HU = Hungary, INMB = incremental net monetary benefit, NL = the Netherlands, QALY = quality-adjusted life year, RNA-NGS = next-generation sequencing panel of tumour RNA, SoC = standard of care.

Table 4 Cost-effectiveness outcomes for the NTRK+ patient subgroup, testing excluded

Strategy	Cost (int€)			Effect (QALY)			Incremental cost (int€)			Incremental effect (QALY)			ICER (int€)			INMB (vs. SoC)		
	EN	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*	*EN*	*HU*	*NL*
SoC	23,621	8,661	60,978	0.7159	0.7146	0.7296												
Entrectinib	67,729	111,480	112,545	2.0685	2.0576	2.1862	41,439	97,525	50,603	1.4010	1.3912	1.5068	29,577	70,100	33,582	8,405	-53,088	54,372

EN = England, ICER = incremental cost-effectiveness ratio, int€ = international euros, HU = Hungary, INMB = incremental net monetary benefit, NL = the Netherlands, QALY = quality-adjusted life year, SoC = standard of care.

Sensitivity analysis

The outcomes of the univariate sensitivity analysis are shown in Figure S2. Outcomes are similar across the three countries. In all countries, the parameters that most impact cost-effectiveness outcomes include the cost of entrectinib treatment, the HR for OS in NTRK+ patients, and utility values. IHC test specificity is a key parameter in the *IHC for all*, *IHC then RNA-NGS*, and *Stratified* strategies, while the cost of RNA-NGS is highly influential in the *RNA-NGS for all* strategy. In England, NTRK prevalence is less influential than in Hungary and the Netherlands, while the cost of taking a biopsy is more influential.

The outcomes of the probabilistic sensitivity analysis are shown in cost-effectiveness planes in Figure 1, and cost-effectiveness acceptability curves (CEACs) are presented in Figure 2. For the latter, the limits on the x-axes were set at roughly fivefold the nationally recommended cost-effectiveness thresholds. The cost-effectiveness planes look similar across the three countries. For the *RNA-NGS*, *IHC then RNA-NGS*, and *Stratified* strategies, uncertainty in our model parameters mostly affects estimated QALY outcomes. For the *IHC* strategy, parameter uncertainty also affects estimated cost outcomes. As shown in the univariate analysis, IHC test specificity is an important parameter, as higher (lower) specificity causes fewer (more) false positive test results, hence fewer (more) patients unnecessarily treated with entrectinib and lower (higher) cost outcomes. The CEACs also appear similar across the countries. The *IHC then RNA-NGS* and *Stratified* strategies have similar curves, but *IHC then RNA-NGS* has a slightly higher probability of being cost-effective at all threshold values. For all countries, the *IHC* strategy has a low probability of being cost-effective even at very high threshold values.

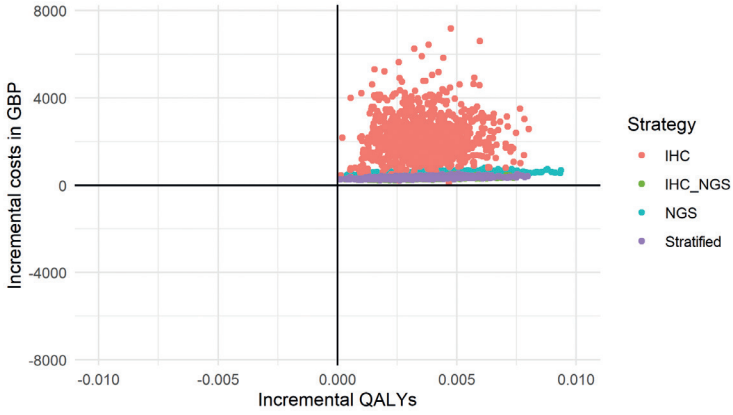


Figure 1A Cost-effectiveness planes: England

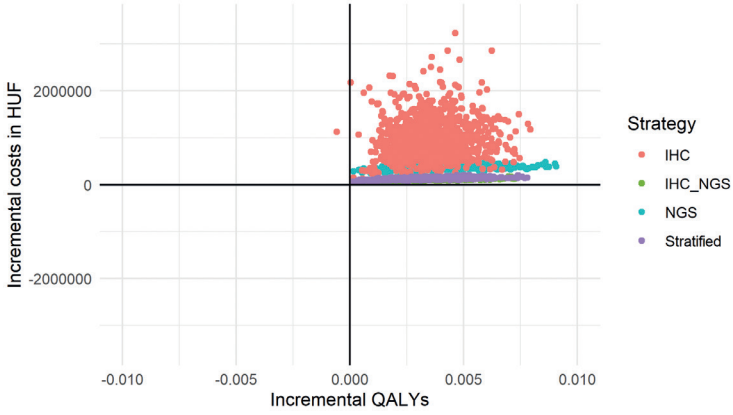


Figure 1B Cost-effectiveness planes: Hungary

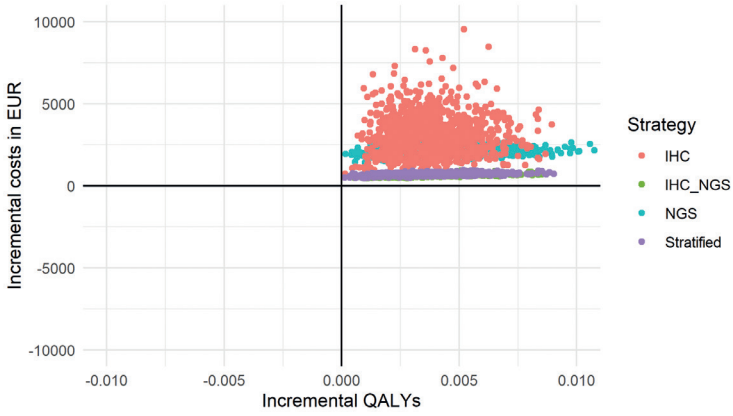


Figure 1C Cost-effectiveness planes: The Netherlands

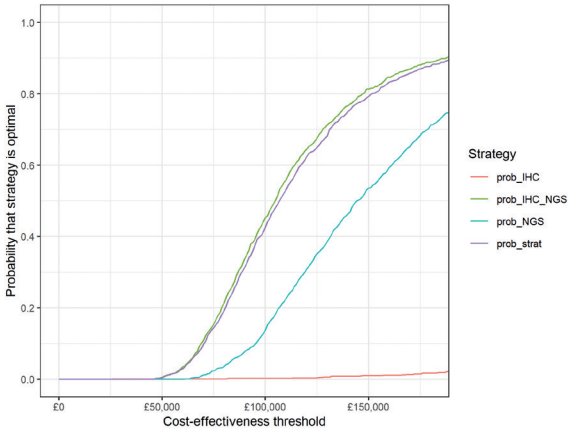


Figure 2A Cost-effectiveness acceptability curves: England

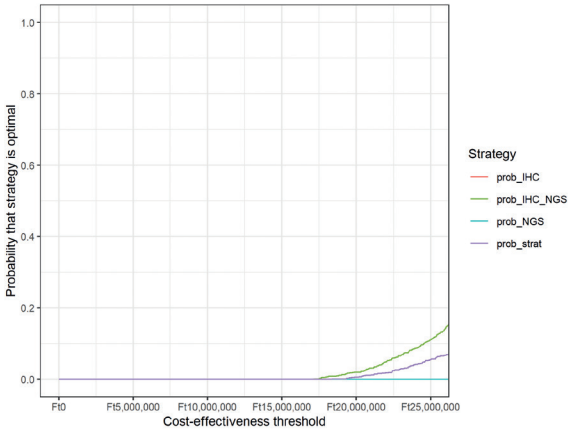


Figure 2B Cost-effectiveness acceptability curves: Hungary

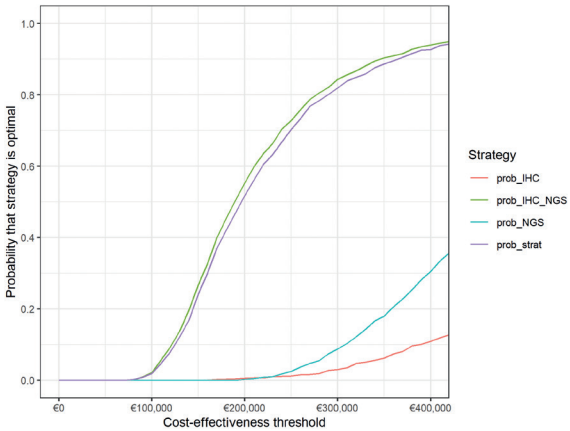


Figure 2C Cost-effectiveness acceptability curves: The Netherlands

Budget impact analysis

The total 5-year budget impact of the *No testing* scenario, in which the entire target population (i.e. adult patients with locally advanced or metastatic solid tumours who have received one or more lines of treatment) is treated with SoC, was estimated to be int€ 6,472,649,274 in England, int€ 1,136,924,382 in Hungary and int€ 6,444,222,855 in the Netherlands. For the testing scenarios, the annual number of patients receiving NTRK testing was estimated to be 102,040 in England, 14,801 in Hungary and 26,476 in the Netherlands.

The incremental budget impact figures for the four testing strategies compared to *No testing* are reported in Table 6. See Table S8 for the budget impact results with country-specific time horizons. In England, the incremental 5-year budget impact of implementing the *IHC then NGS* strategy, which was identified as the most optimal out of the four testing strategies, would be int€ 156,347,606, which is 0.02% of current healthcare expenditure and 0.27% of current cancer care expenditure.(35, 36) Testing costs make up 65.85% of the incremental budget impact of the *IHC then RNA-NGS* strategy.

Although the 5-year incremental budget impact of *IHC then RNA-NGS* is lowest in absolute figures in Hungary (int€ 37,874,049), the relative impact is higher here than in the other two countries, with the budget impact taking up 0.11% of total healthcare expenditure and 1.23% of cancer care expenditure.(35, 36) Just over half (52.19%) of the incremental budget consists of testing costs.

In the Netherlands, testing costs make up a larger part of the 5-year incremental budget impact (int€ 76,879,546), as 81.39% comprises of testing costs. The relative impact on care expenditure in the Netherlands is similar to that in England, with the incremental budget being 0.03% of healthcare expenditure and 0.29% of cancer care expenditure. (35, 36)

In all three countries, the budget impact of *IHC then RNA-NGS* is lowest and the budget impact of *IHC for all* highest. The percentage of the total budget going to testing is lowest in the *IHC for all* strategy (8.80%, 4.32% and 15.62% for England, Hungary and the Netherlands, respectively) because many more (false-positive) patients are treated with entrectinib (see Table 2) in this strategy than in the other testing strategies.

Table 5 Budget impact analysis

	Five-year incremental budget impact (int€)			Percentage of incremental budget impact due to test costs			Budget impact as percentage of total health expenditure			Budget impact as percentage of total cancer care expenditures		
	EN	***HU***	***NL***	***EN***	***HU***	***NL***	***EN***	***HU***	***NL***	***EN***	***HU***	***NL***
IHC then RNA-NGS	156,347,606	37,874,049	76,879,546	65.85	52.19	81.39	0.02	0.11	0.03	0.27	1.23	0.29
Stratified	162,707,341	43,612,999	81,027,374	66.50	57.76	81.95	0.02	0.12	0.03	0.28	1.41	0.31
RNA-NGS for all	247,205,447	117,721,977	233,475,628	74.22	81.90	92.65	0.03	0.34	0.08	0.42	3.81	0.88
IHC for all	1,066,761,912	340,863,660	326,279,464	8.80	4.32	15.62	0.11	0.97	0.11	1.82	11.03	1.23

EN = England, HU = Hungary, IHC = immunohistochemistry, int€ = international euros, NL = the Netherlands, RNA-NGS = next-generation sequencing of tumour RNA.

Discussion

While the TRK inhibitors entrectinib and larotrectinib were the first histology-independent therapies on the market, several other pharmaceuticals targeting specific genetic markers are already in use and many more are expected to become available, allowing for more personalised (cancer) care.(37) As we have argued previously, when evaluating the cost-effectiveness of targeted (or "personalised") treatments, it is important to include an accurate representation of the testing pathway that is needed to identify eligible patients.(38) In this study, we incorporated the period in which NTRK testing is performed in a cost-effectiveness analysis of entrectinib treatment and assessed several different testing strategies. Herewith, we provided an example of how the modelling of testing pathways can be approached and how HTA methods can be used to aid decision-making on which testing strategy to implement. We performed our analysis for three countries, to investigate whether there might be differences in the cost-effectiveness of NTRK testing and entrectinib treatment across different settings, and we found differences indeed.

Our results showed that the implementation of entrectinib is unlikely to be cost-effective in Hungary. Even in the subgroup analysis of NTRK+ patients, where the costs and effects of introducing NTRK-testing were excluded, the INMB of entrectinib was far below zero, at int€ –53,088. Indeed, the incremental costs of entrectinib compared to SoC were large in Hungary, at int€ 97,525, compared to int€ 41,439 in England and int€ 50,603 in the Netherlands. Hungary has lower income per capita and lower healthcare expenditure than England and the Netherlands. Our results could imply that expensive targeted treatments like entrectinib may bring limited value in lower-income countries, and more cost-effective healthcare may have to be prioritised first. Nonetheless, (large) discounts on the price of entrectinib in Hungary could improve its cost-effectiveness results.

The results look more promising for England and the Netherlands, where the NTRK+ subgroup analysis showed a positive INMB for entrectinib compared to SoC. This suggests that entrectinib has the potential to be cost-effective. Nonetheless, all the evaluated testing strategies rendered a negative INMB compared to the base case scenario, in which NTRK testing was not performed and entrectinib was not provided. In line with our results, the National Institute for Health and Care Excellence (NICE) has not recommended entrectinib for routine use in the English National Health Service (NHS) because of its unconvincing cost-effectiveness results but, due to its potential to be cost-effective, NICE did recommend entrectinib to be included in the Cancer Drugs Fund. (39) In the Netherlands, too, entrectinib was recommended for temporary conditional reimbursement. In both England and the Netherlands, the national HTA agencies have entered into an agreement with the manufacturer of entrectinib regarding further data collection, both through additional clinical trials (STARTRK-2 and STARTRK-NG) and

through real-world evidence collection among patients receiving entrectinib.(40, 41) After a period of data collection, a final health economic analysis and reimbursement decision will follow. While such conditional reimbursement schemes can ensure timely market access for patients who might benefit from treatment, withdrawing pharmaceuticals from the market once additional evidence negates their cost-effectiveness can be difficult in practice.(42, 43) In an attempt to counter any such issues, the agreement between the Dutch HTA organisation and the manufacturer explicitly states that all parties involved during the conditional reimbursement period will cooperate in case the reimbursement of entrectinib is discontinued.(41) The agreement also includes a communication plan to ensure patients understand the temporary nature of the current reimbursement decision in place.

Note that the incremental costs, effects and ICERs for the NTRK+ subgroup analysis were similar between England and the Netherlands. Yet, because the cost-effectiveness threshold recommended by Dutch HTA guidelines is much higher than the NICE threshold (€80,000 vs. £36,000 in this case), INMB outcomes are different for the countries (int€ 8,405 in England versus int€ 54,372 in the Netherlands), resulting in different conclusions regarding the interventions' cost-effectiveness. These differences could be justifiable if the values of national thresholds were set based on assessments of opportunity cost and/or societal preferences and so reflected real differences between national settings. However, cost-effectiveness thresholds have historically been based on little to no evidence.(44, 45) Our results illustrate that differences in the thresholds are not inconsequential and might cause (poorly justified) differences in reimbursement decisions across countries.

Clear differences can be seen between the cost-effectiveness outcomes for the main analysis, in which the full NTRK test-treatment pathway was assessed, and the outcomes for the subgroup analysis, in which the testing phase was left out of consideration. This illustrates that including required tests in the economic evaluation of a new treatment may alter reimbursement decisions, supporting the argument we have previously made that testing should be incorporated in the economic model if it is part of the decision problem (e.g. if the testing is to be newly introduced in clinical practice, or to be introduced for additional patient groups).(38) Given that national cost-effectiveness analyses are often based on global models from manufacturers that are adapted to the local setting, and given that testing pathways may vary across countries, we encourage modellers working on country adaptations to ensure that national testing strategies are accurately reflected in (country adaptations of) cost-effectiveness models.

Many have argued in favour of the expansion of genetic testing in healthcare.(46) This could mean that, in the future, all cancer patients find out whether they carry NTRK

gene fusions (or other genetic alterations) as part of standard care, therewith changing the cost-effectiveness outcomes for entrectinib. However, more widespread use of genetic testing does not automatically mean NTRK gene fusions are identified. First, note that whole-genome sequencing (WGS) is DNA-based and, in contrast to RNA-NGS, is not able to determine if a detected NTRK fusion is functional and indeed causing overexpression of TRK proteins.(4) That is, if WGS were to be implemented for all cancer patients, an additional RNA-NGS panel would be needed to identify NTRK+ patients. Also, if broad genetic testing to establish a genetic profile or "passport" for all citizens were implemented, as has been suggested by some, separate genetic tests (including RNA-NGS panels) would be necessary to enable targeted cancer care. This is because many cancer drugs target genetic markers in the tumour DNA, which is different from the patients' germline DNA that would be sequenced when creating a genetic profile.

Finally, our finding that the provision of RNA-NGS to all patients is not cost-effective in England and Hungary even at zero cost for the RNA-NGS test, suggests that the expected price drops in NGS technology alone are insufficient to make the implementation of NTRK testing and subsequent treatment cost-effective. Further changes may be needed in the test-treatment pathway, such as a price reduction for entrectinib.

Limitations

For the Netherlands, we previously illustrated how registry data containing genomic and clinical parameters can be used to construct a control arm for single-arm trial data. As explained in the Methods, we used the Dutch registry data to estimate survival and time to treatment discontinuation among NTRK- patients and to estimate a hazard ratio for NTRK+ patients.(9) In this study, where we also assess cost-effectiveness for England and Hungary, we were unable to perform a similar analysis for England and Hungary, due to data and resource limitations. This shows one of the limitations of single-arm trial data; local data and resources may be insufficient to estimate a control arm for the national setting, causing uncertainty about the treatment's effectiveness. Moreover, the publicly available data on the patient populations in clinical trials tends to be limited and insufficient to determine the extent to which patients in the trial population and the control population are comparable, so that issues of confounding and selection bias cannot be sufficiently addressed.(47, 48) Indeed, even after estimating a control arm using a Dutch population of cancer patients, large uncertainty remains about entrectinib's effectiveness in the Netherlands as we cannot establish to what extent the patients and the healthcare practices in the US-based entrectinib trials can be compared to the patients and healthcare practices in the Dutch database. Other approaches for estimating a control arm to single-arm trial data have been suggested, including the use of trial patients who did not respond to the treatment as a proxy for control arm

patients,(49) and basing control arm estimations on trial patients' progression-free survival during the most recent prior therapy.(49, 50) However, these methods require patient-level trial data, which may not always be available to modellers. As argued by many others, randomised controlled trials have many advantages over its alternatives and are generally preferred.(51)

Furthermore, because of the small sample size of the entrectinib trials, no tumour type-specific estimates of treatment effectiveness were available. We therefore had to assume that the effectiveness of entrectinib is homogeneous across tumour types, despite prior research suggesting this may be inaccurate.(52) Although we did estimate tumour type-specific effectiveness for the SoC arm, we deemed it inappropriate to present tumour type-specific ICER and INMB estimates, given the possibility that entrectinib effectiveness is heterogeneous. If indeed there is such heterogeneity, the single ICER values we estimated might be biased, as the proportional distribution of tumour types in the entrectinib trials may differ from the proportional distribution in clinical practice.

Another data limitation in this study was that we assumed the Dutch costs for adverse events (adjusted using PPP conversion factors) and Dutch estimates of the relationship between SF-6D utility and proximity to death to apply to England and Hungary as well. We also used list prices to obtain the costs of entrectinib, while actual costs may be lower due to discounts, potentially improving the cost-effectiveness of the implementation of entrectinib.

Conclusion

Health technology assessments of histology-independent cancer treatments have proven challenging, for various reasons. Our study has helped to provide possible approaches to address some of the challenges. Furthermore, we have provided cost-effectiveness estimates for the implementation of the histology-independent therapy entrectinib in three different countries.

Out of four NTRK-testing strategies assessed, the optimal strategy was the same in England, Hungary and the Netherlands. The strategy starts with an IHC test (which assesses TRK protein expression) for all patients, followed by an RNA-NGS test (which looks at the tumour RNA to identify NTRK gene fusions) for patients who receive a positive result on the IHC test.

Nonetheless, the implementation of NTRK testing followed by treatment with entrectinib is likely not cost-effective in Hungary. In England and the Netherlands, the implementation of entrectinib was also not found to be cost-effective, though the results from a subgroup analysis of NTRK+ patients suggested that entrectinib has the potential to be cost-effective.

References

1. Assessment Report VITRAKVI International Non-Proprietary Name: Larotrectinib, (2019).
2. Rozlytrec (Entrectinib), (2020).
3. Amatu A, Sartore-Bianchi A, Bencardino K, Pizzutilo EG, Tosi F, Siena S. Tropomyosin receptor kinase (TRK) biology and the role of NTRK gene fusions in cancer. Ann Oncol. 2019;30(Suppl_8):viii5-viii15.
4. Hsiao SJ, Zehir A, Sireci AN, Aisner DL. Detection of Tumor NTRK Gene Fusions to Identify Patients Who May Benefit from Tyrosine Kinase (TRK) Inhibitor Therapy. J Mol Diagn. 2019;21(4):553-71.
5. Gatalica Z, Xiu J, Swensen J, Vranic S. Molecular characterization of cancers with NTRK gene fusions. Mod Pathol. 2019;32(1):147-53.
6. Drilon A. TRK inhibitors in TRK fusion-positive cancers. Ann Oncol. 2019;30(Suppl_8):viii23-viii30.
7. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and External Controls in Clinical Trials - A Primer for Researchers. Clin Epidemiol. 2020;12:457-67.
8. Ghadessi M, Tang R, Zhou J, Liu R, Wang C, Toyoizumi K, et al. A roadmap to using historical controls in clinical trials - by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). Orphanet J Rare Dis. 2020;15(1):69.
9. Huygens S, Vellekoop H, Versteegh M, Santi I, Szilberhorn L, Zelei T, et al. Cost-effectiveness analysis of treating NTRK-positive cancer patients with the histology-independent therapy entrectinib. Submitted.
10. Aerts J, Dinjens W, Evers M, Grünberg K, van Herpen C, Kapiteijn H, et al. Consensus diagnose en behandeling van NTRK-genfusie gerelateerde solide tumoren. 2020.
11. Kroneman M, Boerma W, van den Berg M, Groenewegen P, de Jong J, van Ginneken E, et al. Netherlands: health system review. Health Systems in Transition. 2016;18.
12. Gaál P, Szigeti S, Csere M, Gaskins M, Panteli D. Hungary: Health system review. Health Systems in Transition. 2011;13.
13. Cylus J, Richardson E, Findley L, Longley M, O'Neill C, Steel D. United Kingdom: Health system review. Health Systems in Transition. 2015;17.
14. Doebele RC, Drilon A, Paz-Ares L, Siena S, Shaw AT, Farago AF, et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. Lancet Oncol. 2020;21(2):271-82.
15. NICE health technology evaluations: the manual, (2022).
16. Nutrition NIoPa. Gyógyszereink. 2017.
17. Zorginstituut Nederland. Guideline for economic evaluations in healthcare. 2016.
18. Solomon JP, Benayed R, Hechtman JF, Ladanyi M. Identifying patients with NTRK fusion cancer. Ann Oncol. 2019;30(Suppl_8):viii16-viii22.
19. Wong D, Yip S, Sorensen PH. Methods for Identifying Patients with Tropomyosin Receptor Kinase (TRK) Fusion Cancer. Pathol Oncol Res. 2020;26(3):1385-99.
20. Kirchner M, Glade J, Lehmann U, Merkelbach-Bruse S, Hummel M, Lehmann A, et al. NTRK testing: First results of the QuiP-EQA scheme and a comprehensive map of NTRK fusion variants and their diagnostic coverage by targeted RNA-based NGS assays. Genes Chromosomes Cancer. 2020;59(8):445-53.
21. Aerts JG, Dinjens WN, Evers MP, Grünberg K, van Herpen C, Kapiteijn H. Consensus Diagnose en Behandeling van NTRK-Genfusie Gerelateerde Solide Tumoren. 2020.
22. Bins S, Cirkel GA, Gadellaa-Van Hooijdonk CG, Weeber F, Numan IJ, Bruggink AH, et al. Implementation of a Multicenter Biobanking Collaboration for Next-Generation Sequencing-Based Biomarker Discovery Based on Fresh Frozen Pretreatment Tumor Tissue Biopsies. Oncologist. 2017;22(1):33-40.
23. Williams HL, Walsh K, Diamond A, Oniscu A, Deans ZC. Validation of the Oncomine() focus panel for next-generation sequencing of clinical tumour samples. Virchows Arch. 2018;473(4):489-503.
24. Bazhenova L, Lokker A, Snider J, Castellanos E, Fisher V, Fellous M, et al. TRK Fusion Cancer: Patient Characteristics and Survival Analysis in the Real-World Setting. Target Oncol. 2021;16(3):389-99.
25. World Bank. Global Economic Monitor (GEM) [Available from: https://databank.worldbank.org/source/global-economic-monitor-(gem).

26. Organisation for Economic Co-Operation and Development (OECD). OECD.Stat [Available from: https://stats.oecd.org/.
27. World Bank. World Development Indicators 2020 [Available from: https://databank.worldbank.org/source/world-development-indicators.
28. Luta X, Diernberger K, Bowden J, Droney J, Howdon D, Schmidlin K, et al. Healthcare trajectories and costs in the last year of life: a retrospective primary care and hospital analysis. BMJ Support Palliat Care. 2020.
29. van Baal PH, Wong A, Slobbe LC, Polder JJ, Brouwer WB, de Wit GA. Standardizing the inclusion of indirect medical costs in economic evaluations. Pharmacoeconomics. 2011;29(3):175-87.
30. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ. 1995;14(2):171-89.
31. Versteegh M, van der Helm I, Mokri H, Oerlemans S, Blommestein H, van Baal P. Estimating Quality of Life Decrements in Oncology Using Time to Death. Value Health. 2022.
32. Paulden M. Calculating and Interpreting ICERs and Net Benefit. Pharmacoeconomics. 2020;38(8):785-807.
33. Forsythe A, Zhang W, Phillip Strauss U, Fellous M, Korei M, Keating K. A systematic review and meta-analysis of neurotrophic tyrosine receptor kinase gene fusion frequencies in solid tumors. Ther Adv Med Oncol. 2020;12:1758835920975613.
34. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. 2013.
35. World Health Organization (WHO). Global Health Expenditure Database: Current Health Expenditure (CHE) per Capita 2019 [Available from: https://apps.who.int/nha/database/ViewData/Indicators/en.
36. Hofmarcher T, Lindgren P, Wilking N, Jonsson B. The cost of cancer in Europe 2018. Eur J Cancer. 2020;129:41-9.
37. Goetsch CM. Genetic tumor profiling and genetically targeted cancer therapy. Semin Oncol Nurs. 2011;27(1):34-44.
38. Vellekoop H, Huygens S, Versteegh M, Szilberhorn L, Zelei T, Nagy B, et al. Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine. Pharmacoeconomics. 2021;39(7):771-88.
39. (NICE) NIfHaCE. Entrectinib for Treating NTRK Fusion Positive Solid Tumours [ID1512]. 2020.
40. Excellence NIfHaC. Cancer Drugs Fund - Data Collection Arrangement: Entrectinib for treating NTRK fusion-positive solid tumours (ID1512). 2020.
41. Zorginstituut Nederland. Vervolgadvies voorwaardelijke toelating van entrectinib (Rozlytrek®) bij solide tumoren met een NTRK-genfusie (procedure: weesgeneesmiddelen, conditionals en exceptionals). 2021.
42. McCabe C, Chilcott J, Claxton K, Tappenden P, Cooper C, Roberts J, et al. Continuing the multiple sclerosis risk sharing scheme is unjustified. BMJ. 2010;340:c1786.
43. van de Wetering EJ, van Exel J, Brouwer WB. The Challenge of Conditional Reimbursement: Stopping Reimbursement Can Be More Difficult Than Not Starting in the First Place! Value Health. 2017;20(1):118-25.
44. Cameron D, Ubels J, Norstrom F. On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. Glob Health Action. 2018;11(1):1447828.
45. Vallejo-Torres L, Garcia-Lorenzo B, Castilla I, Valcarcel-Nazco C, Garcia-Perez L, Linertova R, et al. On the Estimation of the Cost-Effectiveness Threshold: Why, What, How? Value Health. 2016;19(5):558-66.
46. Pritchard DE, Moeckel F, Villa MS, Housman LT, McCarty CA, McLeod HL. Strategies for integrating personalized medicine into healthcare practice. Per Med. 2017;14(2):141-52.
47. Lu CY. Observational studies: a review of study designs, challenges and strategies to reduce confounding. Int J Clin Pract. 2009;63(5):691-7.
48. McNamee R. Confounding and confounders. Occup Environ Med. 2003;60(3):227-34; quiz 164, 234.
49. Briggs A, Wehler B, Gaultney JG, Upton A, Italiano A, Bokemeyer C, et al. Comparison of Alternative Methods to Assess the Cost-Effectiveness of Tumor-Agnostic Therapies: A Triangulation Approach Using Larotrectinib as a Case Study. Value Health. 2022;25(6):1002-9.
50. Krebs MG, Blay JY, Le Tourneau C, Hong D, Veronese L, Antoniou M, et al. Intrapatient comparisons of efficacy in a single-arm trial of entrectinib in tumour-agnostic indications. ESMO Open. 2021;6(2):100072.

51. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. BJOG. 2018;125(13):1716.
52. Murphy P, Claxton L, Hodgson R, Glynn D, Beresford L, Walton M, et al. Exploring Heterogeneity in Histology-Independent Technologies and the Implications for Cost-Effectiveness. Med Decis Making. 2021;41(2):165-78.
53. de Groot S, Santi I, Bakx P, Wouterse B, van Baal P. Informal care costs according to age and proximity to death to support cost-effectiveness analyses. PharmacoEconomics. 2023;41(9):1137-49.

Supplementary materials

Supplementary materials can be found at https://doi.org/10.2217/pme-2022-0070

Contents

CHAPTER 7

General discussion

In this PhD thesis, a health economics perspective was applied to the field of personalised medicine (PM). The aim of the thesis was threefold: to gain a better understanding of the added value of personalised medicine; to outline methodological challenges in economic evaluations of personalised medicine as well as ways to address them; and to conduct a case study to evaluate the cost-effectiveness of a personalised medicine intervention.

The added value of personalised medicine

To better understand the value of PM, a systematic literature review of economic evaluations of PM was performed. As described in **Chapter 2**, we extracted data on the incremental health benefits and incremental costs for all two-way comparisons between a PM intervention on the one hand and a non-PM intervention on the other. The incremental net monetary benefit (ΔNMB) was subsequently calculated for each PM intervention, using standardised estimates of national cost-effectiveness thresholds.

The median incremental health benefit of PM interventions was found to be 0.03 quality-adjusted life year (QALY) per patient, or roughly 11 quality-adjusted days. The mean incremental benefit was 0.26 QALY, or roughly 3 quality-adjusted months. The added health benefits of PM seem therefore modest on average. They also seem not too dissimilar from the added health benefits of other new healthcare interventions, as a literature review of economic evaluations of all types of interventions found a median QALY increase of 0.06 and a mean QALY increase of 0.31.[1]

The results also showed that the ΔNMB of PM interventions centred around zero, with a median of Int$ 18. This implies that the health benefits of PM tend to be counterbalanced by the associated cost increases. High prices for new pharmaceuticals may be part of the reason for this finding. Although price negotiations can be used to lower prices, additional measures aiming at curbing the hegemony of the pharmaceutical industry may be needed to achieve more modest price-setting.[2,3]

Large heterogeneity among PM interventions

The findings of the systematic review suggest that, so far, there has been a limited actualisation of the high hopes and expectations surrounding PM into (large) added value to the healthcare system. Nonetheless, large heterogeneity in outcomes was also identified, with ΔNMB as high as Int$ 21,615 at the 95th percentile and ΔNMB as low as Int$ –91,832 at the 5th percentile. This implies that not all forms of "personalising" healthcare are worthwhile – from a healthcare system perspective at least. It also seems that the term "personalised medicine" may be too broad to indicate whether individual interventions falling in this category will be valuable or not. The term may therefore not be suitable when deciding upon resource allocation. That is, the sizeable allocation of

public resources toward the field of PM over recent years, based on high expectations for PM at large, may have partially ended up funding PM interventions and innovations with limited value.

We performed a regression analysis to try and identify possible factors explaining why some PM interventions seem so much more valuable than others. Gene therapies were found to be a distinct subcategory of PM, as gene therapies were associated with much larger health benefits than other PM interventions. Indeed, most of the gene therapies included in the review were treatments for early onset conditions with high morbidity and mortality, meaning there is the potential for large QALY gains. However, gene therapies were also associated with higher costs and lower ΔNMB than other PM interventions. This suggests that gene therapies have the potential to bring large health benefits but tend to render a net loss to the healthcare system at current price-setting. Nonetheless, the number of gene therapies included in our analysis was limited and the findings were found to be uncertain.

The regression analysis also suggested that PM interventions aimed at preventing adverse drug reactions (ADR) tend to have higher ΔNMB, while PM interventions focussing on the identification (and treatment) of likely responders to therapy tend to have lower ΔNMB. A possible explanation for this is that many of the interventions in the former category stratify patients to existing (off-patent) treatments (e.g. introducing DPYD-profiling prior to administering fluoropyrimidines), whereas many interventions in the latter category stratify toward newly developed treatments (e.g. the provision of osimertinib to non-small cell lung cancer patients with EGFR-positive tumours). Further research and development to better stratify existing care therefore seems valuable. However, it is currently more profitable for pharmaceutical companies to invest in new drugs for genetic subsets of patients than to invest in research on how to better target existing therapies. The onus is therefore on policymakers to encourage improved stratification in existing care, possibly through rules and regulations as well as the allocation of research funding.

Opportunity costs of investment in PM

Most (60%) of the economic evaluations of PM interventions that were identified in the systematic review were cancer treatments. Indeed, as mentioned in **Chapter 1**, many of the developments in PM so far have been in the field of oncology. This is not necessarily undesirable, as cancer is a disease with high morbidity and mortality. However, a study investigating funding priorities in medical research found that cancers are overfunded relative to both their morbidity and mortality.[4] Disease areas that were found to be underfunded compared to their global disease burden were, among others, maternal sepsis, maternal haemorrhage, several sexually transmitted diseases, all tropical

diseases, psychiatric conditions such as panic disorder and insomnia, rheumatic heart disease, and hypertensive heart disorder. More generally, the field of PM, with its focus on gene profiling, implicitly prioritises disease areas where gene profiling has more potential to be valuable (e.g. cancer and rare genetic disorders) over other disease areas. Therefore, the large amounts of research funding that have been allocated to PM over the past years may have been at the expense of disease areas with limited use for genetic profiling yet with high unmet need. Apart from risk profiling tests, PM has also focussed on curative care more than preventive care. Primary prevention, especially, has received little attention in PM, despite substantial possible benefits from more prevention efforts. Indeed, it is estimated that around 80% of major chronic diseases can be prevented through lifestyle factors.[5,6] In oncology it is estimated that only 5-10% of cancer cases are due to genetic defects, while 90-95% of cases are caused by lifestyle and environmental factors.[7] As mentioned in **Chapter 1**, the term "personalised medicine" has various interpretations. The ambiguity and the appeal of the word "personalised" may have contributed to the popularity of PM. However, with its emphasis on oncology and genetic testing, and its lack of attention for primary prevention, PM is relatively specific in practice. Arguably, the focal areas of PM and the related opportunity costs are to some extent hidden by the broad and multi-interpretable word "personalised". The use of a more specific term, such as "stratified medicine" or " genomic medicine" may therefore be more appropriate.

Methodological guidance for economic evaluations of personalised medicine

In response to concerns that existing health technology assessment (HTA) frameworks might be inadequate for assessing the cost-effectiveness of PM interventions, we investigated the methodological challenges that PM may bring. In **Chapter 3**, an overview of possible challenges was presented, as well as guidance on how to proceed in the face of these challenges. The guidance included 23 recommendations and was based on extensive research. We performed a targeted literature review for an initial overview of possible methodological challenges, held interviews with 18 experts, recorded the methods used in the economic evaluations of PM identified through the systematic literature review mentioned previously, and held a workshop with around 30 experts to discuss and refine our initial set of recommendations. **Chapters 4-6** described a case study, regarding the cost-effectiveness of NTRK-testing and subsequent treatment with entrectinib, where the guidance was applied as much as possible. Below follows a recap of key elements of the guidance, including a description of how its recommendations were incorporated in the case study. Further thoughts and recommendations are also given.

Modelling test-treatment pathways

Most health economic models to date start from the moment that patients begin treatment (whether pharmaceutical or otherwise), meaning that the population eligible for treatment has already been identified. In PM, however, the (genetic) testing of patients and subsequent stratification to different treatments plays a central role and leaving out the "testing"-part from the health economic model may give an incomplete depiction of reality. The guidance therefore advocated for the inclusion of the testing period in economic evaluations of PM and offered eight recommendations for the accurate modelling of test-treatment pathways.

Given that different technologies might be used for genetic profiling and given that tests might be applied at different times and in varying combinations, modellers were urged to not only include one of the available options in the economic model but to identify all relevant test-treatment pathways and to justify why the pathways included in the model were selected (Recommendation [Rec.] 3).

In our study evaluating the cost-effectiveness of NTRK testing and treatment for the Netherlands (**Chapter 5**), the assumed test-treatment pathway was based on a consensus report by Dutch experts outlining the envisioned NTRK-testing protocol in Dutch clinical practice. Given the variety of experts involved and the consensus reached among them, the suggested protocol was deemed the most likely scenario for the clinical implementation of NTRK testing hence was used for the model. In a follow-up study (**Chapter 6**), the cost-effectiveness of three additional NTRK-testing strategies (which were inspired by interviews with clinical geneticists, oncologists, and other experts) was assessed. This was done for the Netherlands and two additional countries: Hungary and England.

The aim of genetic tests is generally to determine whether specific markers are present or not. However, tests tend to not be 100% accurate, leading to false positive (FP) and false negative (FN) results. Our guidance recommended the explicit modelling of outcomes for patients with false positive and false negative test results (Rec. 4). Modellers were also urged to make sure that the data used to estimate the diagnostic accuracy of a testing technology aligns with the patient population in the model (Rec. 5).

Two testing technologies were included in our analysis: next-generation sequencing of RNA (RNA-NGS) and immunohistochemistry (IHC). RNA-NGS is currently the gold standard for detecting NTRK gene fusions (i.e. sensitivity and specificity are assumed to be 100%), while IHC testing is less accurate.

Because of its imperfect accuracy, IHC was assumed to render both FP and FN results. Patients with an FN result, i.e. unidentified NTRK+ patients, were assumed to be treated according to the standard of care (SoC) and have the survival probabilities of NTRK+ patients receiving SoC. Patients with an FP result, i.e. NTRK- patients wrongly seen as NTRK+, were assumed to be treated with entrectinib. For a lack of data on the effect of entrectinib on NTRK- patients, it was assumed that they have the same survival probabilities as NTRK- patients receiving SoC. However, the latter only applied to one of the testing scenarios, as in all other scenarios patients received follow-up RNA-NGS testing after a positive IHC test, meaning that FP results would be filtered out by the RNA-NGS results.

To calculate the probabilities of FP and FN results, a literature search was performed on the sensitivity and specificity of IHC testing for elevated levels of the TRK protein. Tumour type-specific estimates were gathered, as the accuracy of IHC testing may differ across tumour types (despite the tumour-agnostic label of entrectinib). The prevalence of the different tumour types in our patient population was subsequently accounted for.

Further recommendations for ensuring that the testing pathway is accurately modelled concerned the possible consequences of waiting times during the testing phase (Rec. 8) and possible variation in the costs associated with testing (Rec. 9).

Because there is relatively high short-term mortality in our patient population with advanced cancer, patients might die during the waiting period associated with testing. To estimate the impact of the waiting period, the waiting times for various steps of the testing phase were included in the model. We considered the waiting time for a biopsy, the waiting time between the biopsy and test results being available, and the waiting time between test results being available and the start of treatment. Since not all patients go through the exact same process (e.g. some patients still have tissue available from a previous biopsy and do not need a new one), the model allowed for varying waiting times depending on the different steps that patients went through. For each total waiting time, the percentage of patients that would die during the waiting period was calculated. NTRK+ patients who died during the waiting period were assumed to incur the costs associated with testing but receive no entrectinib treatment.

Possible variation in testing costs was addressed in two ways. First, total testing costs were varied according to the steps patients had gone through, ranging from 'only one test being performed' to 'both IHC and RNA-NGS as well as three biopsies being performed'. For the Netherlands (**Chapter 5**), differences in testing costs between hospitals were also considered by obtaining list prices from a geographically spread selection of academic, teaching, and general hospitals and calculating weighted averages.

Incorporating patient and clinician compliance

Test-treatment protocols may be imperfectly adhered to, both by clinicians and patients. This is the case in healthcare more broadly, for example with screening programmes such as for colorectal cancer. Nonetheless, the issue may be more pronounced in PM because of the complicated nature of genetics as well as the presence of many alternative test-treatment pathways. Depending on the aim of a given health economic model, it may be valuable to account for patient and clinician compliance. Our guidance encouraged the inclusion of parameters reflecting patient and clinician compliance in economic evaluations for decision-makers who require cost-effectiveness results under realistic circumstances (Rec. 19). Comparing outcomes under current levels of compliance to a "perfect implementation" scenario may also provide insights in the value of further efforts to increase compliance. Additionally, it was recommended that modellers consider possible variation in compliance across societal groups (Rec. 20).

Patient and clinician compliance were not accounted for in our case study. The reasoning for this was that high compliance was expected because of the severity of disease and the lack of treatment alternatives for our patient population. We were also faced with limited data availability regarding the uptake of genetic testing for targeted cancer treatment. Nonetheless, several studies have shown that non-white patients and patients with a lower socioeconomic status are much less likely to be referred for genetic testing.[8–10] Possible explanations have been suggested, both on the patient side (e.g. mistrust of genetic testing, limited health literacy) and on the provider side (e.g.

implicit bias,[11] poor judgement regarding which patients would benefit most from genetic testing,[12] insufficient ability to identify patients with limited health literacy and clearly communicate with them).[8] Nonetheless, limited research has been done to determine the causes of the identified disparities. Indeed, a systematic review of studies investigating attitudes toward cancer genomic testing found that most research participants were white.[13] The increasing use of genetic testing in clinical practice therefore carries a significant risk of exacerbating existing health inequalities. Further research into the causes of the differing referral rates is strongly encouraged. Subsequent policy changes may be needed to ensure that the benefits of genetic testing reach all.

(Lack of) effectiveness data

The more patients are stratified into specific subgroups, the smaller the sample size of each patient population becomes. The large randomised controlled trials (RCTs) that have been used over the past decades to measure comparative effectiveness may therefore not always be feasible in PM. Although alternative, more flexible versions of RCTs have been developed, pharmaceutical companies have so far mostly resorted to single-arm trials to estimate the treatment effectiveness of interventions aimed at (relatively) small patient groups. Comparative effectiveness cannot be established from single-arm trials, for their lack of a control arm. External data may therefore be used to support the estimation of comparative effectiveness. Our guidance stated that when external data is used to estimate treatment effectiveness for patients with a specific genetic marker, the prognostic value of the genetic marker should be accounted for (Rec. 14).

In **Chapter 4** the prognostic value of NTRK gene fusions was estimated. The clinicogenomic Hartwig Medical Foundation (HMF) database was used to obtain a sample of cancer patients who were treated in Dutch clinical practice. We identified 24 NTRK+ patients and matched them to 96 NTRK- patients in a propensity score matching analysis. A subsequent survival analysis estimated a hazard ratio of 1.44 (95% CI 0.81-2.55) for NTRK+ patients.

The HMF database was also used to estimate tumour type-specific survival curves for NTRK- patients receiving SoC (**Chapter 5**). To calculate survival for NTRK+ patients receiving SoC, the estimated hazard ratio of 1.44 was applied to the survival curves of NTRK- patients.

In our case study, we tried our best to construct an appropriate comparator arm to the single-arm entrectinib trials. Besides accounting for the prognostic value of NTRK gene fusions, we selected a patient population that matched the inclusion criteria of the entrectinib trials (patients with locally advanced or metastatic disease who had received at least one previous line of treatment). We also only included patients in the comparator arm who had one of the tumour types appearing in the trial population.

However, important differences are likely to have remained, for example because trial patients were treated in the US, Korea, Spain, and Italy, while patients in our comparator arm were treated in the Netherlands. With access to patient-level trial data, we might have been able to use a more advanced approach than the naïve comparison we conducted. Nonetheless, the combination of single-arm trial data with external, real-world data for the estimation of comparative treatment effectiveness would have had many remaining limitations. Among these are unobserved confounders[14] as well as the Hawthorne effect (i.e. the setting of clinical trials affects patient and clinician behaviour because participants know they are being observed) possibly skewing the estimated difference between the trial arm and the real-world comparator arm. Beaulieu-Jones et al., in a study on the use of real-world evidence in healthcare regulation, also point out that multiple hypothesis testing is more likely to go unreported or unnoticed when real-world data is used, especially when it pertains large, widely distributed datasets.[15] They write: "[M]ultiple chances and analytical approaches to explore an observational hypothesis may result in different point estimates. Given the financial stakes in regulatory outcomes, there are strong incentives for reporting of positive results".[15]

Given the many drawbacks of using real-world data to estimate treatment effectiveness, forgoing an RCT should be avoided as much as possible. Commonly cited justifications for a lack of RCT evidence are the rarity of the condition in question and the absence of approved comparator treatments.[16] However, an analysis of EMA and FDA approvals found that these arguments are inconsistently applied, with examples of RCT evidence submitted for treatments for highly rare conditions and both controlled and uncontrolled studies submitted under similar availability of comparator treatments.[16] Another study found that, between 2014 and 2019, more than half of the oncology drugs receiving FDA approval based on single-arm trial data had approved comparators available that they should have been compared to.[17] More stringent guidance on when a departure from RCT evidence is justified is therefore encouraged. For small patient groups, the use of more flexible RCT designs may be stimulated.[18,19] Stricter guidelines on the use of observational data are also encouraged. For single-arm trials, for example, preregistration may be required of the methods to be used in constructing a synthetic control arm, possibly with the additional requirement to perform a follow-up study with a *prospective* cohort if the original observational data was retrospective.[15]

To achieve a stronger evidence base, tighter collaboration may be needed between marketing authorisation agencies (e.g. EMA, FDA) and reimbursement authorities. There has been an increasing use of expedited drug approval programmes by both EMA and FDA over recent years.[20,21] The rationale for expedited approval is generally that there is an "unmet clinical need" and that patients should receive access to the pharmaceutical as quickly as possible. It has however been argued that the argument

of "unmet clinical need" was inappropriately used in several expedited approvals.[22] Allowing pharmaceuticals onto the market with limited clinical evidence also has numerous disadvantages. There may be harm to patients, as severe side effects may have gone undetected in the submitted evidence.[22] Expedited approvals may also cause a suboptimal allocation of healthcare resources, as treatment effectiveness may turn out to be worse than the preliminary estimates. Although marketing authorisation agencies and reimbursement agencies have different roles and aims, a better streamlining of their respective rules and requirements may be needed for improved population health.

Reflecting and addressing uncertainty

Large uncertainty is often mentioned as a key challenge in PM.[23,24] Although uncertainty is present in all health economic models, characteristics particular to PM may give rise to increased parameter uncertainty as well as structural uncertainty. Parameter uncertainty (i.e. uncertainty as to how closely the estimated values for input parameters approximate the true values) may for example be increased by the unclear causal status between genetic markers and clinical outcomes, or by unknown genetic data quality due to a lack of standardisation across laboratories and across genomic databases.[25] Parameter uncertainty is affected by sample size as well as by variance in the data. Therefore, opposing factors are at play in the estimation of treatment effectiveness in PM. On the one hand, increased patient stratification decreases heterogeneity within each patient population, thereby decreasing parameter uncertainty. On the other hand, increased patient stratification may also lead to smaller sample sizes, thereby increasing parameter uncertainty.

Structural uncertainty (i.e. uncertainty whether the model assumptions and selected methods are appropriate) may arise in PM because of complex testing procedures with many possible test-treatment pathways. Structural uncertainty also arises in the estimation of comparative effectiveness, as method selection may have a large impact on the estimated outcomes.[26] For example, in a study of six clinical areas, Sacks et al. found that the results of clinical trials were more dependent on the method for selecting control groups than on the therapy under consideration.[27] This issue is exacerbated in PM, as various different methodological choices may have to be made to estimate the outcome of interest (e.g. lifetime incremental QALYs) based on limited available data. In our guidance we therefore encouraged modellers to identify uncertainties in structural assumptions and decisions and investigate their impact on cost-effectiveness results through a sensitivity analysis (Rec. 22).

An elaborate sensitivity analysis was performed in the case study to evaluate the impact of parameter uncertainty on the results. However, apart from a scenario analysis regarding the costs and health consequences of the testing phase, structural uncertainty was not quantified. Resource constraints were the main reason for this. Generally, evaluating the impact of structural assumptions and decision on cost-effectiveness requires extensive additional modelling work. For example, evaluating the impact of our method selection for constructing a synthetic control arm may require constructing one or multiple new control arms from scratch (using different methods). Building several different versions of a single health economic model may in practice not always be feasible. Also, while the inclusion of all elements of uncertainty in sensitivity analyses provides a more complete reflection of reality, cost-effectiveness results surrounded by a lot of uncertainty may offer little help in decision-making. Uncertainty in PM is therefore best addressed at the source. As argued in the previous section, there is still much room for improvement in the generation of clinical evidence. Beyond stricter requirements for the evidence that is submitted to marketing authorisation agencies and reimbursement agencies, policymakers may encourage more widespread collection of clinical and genomic registry data (while ensuring data protection of course) and more standardisation of methods across clinicogenomic databases. Improved clinicogenomic data collection is likely to increase our understanding of the relationships between genetic markers and clinical outcomes. For the insights from improved data collection to benefit all, it is key that included sample populations are representative of the wider population. Given large historic differences in research inclusion based on ethnicity, sex, age, socioeconomic status, and geography, among other things, additional efforts may be needed to ensure inclusivity in clinical data collection.[28–34]

In cases where uncertainty is simply too large for a reimbursement decision to be made, managed entry agreements may be struck between manufacturer and healthcare payer. Various types of managed entry agreement exist (e.g. payment by result, price-volume agreements, etc.), with a range of conditions that can be included. Health economic modelling can be used to optimise the conditions of a managed entry agreement. The guidance stated that if a managed entry agreement is being considered for an intervention, its conditions may be included in the model evaluating the intervention (Rec. 23). Given that clinical guidelines and reimbursement status are not always adjusted after updated estimates of treatment effectiveness become available, the use of MEAs risks wasting resources on insufficiently effective (or harmful) treatments long-term.[22,35] To mitigate this risk, standardised quality protocols for further data collection may need to be developed and standardised re-assessment procedures may be put in place to ensure clinical guidelines and reimbursement status are adjusted when necessary.

Additional value elements

In 2018 a report was released by the ISPOR Value Assessment Framework Special Task Force arguing that the QALY insufficiently reflects the value of healthcare interventions.[36] The report listed several "additional elements of value", including value of knowing, fear of contagion, insurance value, value of hope, real option value, severity of disease, equity, and scientific spillovers.[36] The report received a relatively large amount of attention in the pharmacoeconomic field, with some arguing that several of the suggested value elements may be particularly important in PM.[37] However, in our guidance we identified several flaws to the suggested value elements (**Chapter 3**). A key limitation we pointed out is that many of the elements are conceptually ambiguous. They lack clear definitions, and some appear to partially overlap. Below follows a more detailed discussion of each suggested element of value.

Note first that the value elements focus on slightly different target groups. *Fear of contagion* and *insurance value* describe interventions' effect on the general public, i.e. people who are not currently affected by the condition the intervention is for. *Value of knowing*, *value of hope*, and *real option value* concern patients who are in care for the condition in question. Finally, *severity of disease*, *equity*, and *scientific spillovers* are mostly related to societal goalsetting.

According to the task force, *fear of contagion* relates to the reduced anxiety people may experience once interventions become available that decrease their chances of being infected by a contagious condition (e.g. Ebola vaccine, COVID-19 vaccine). *Insurance value* refers to the comfort people may experience from knowing that if they were to get the condition in question, (improved) treatment would be available. Considering psychological effects among the general public in a quantitative manner in decision making would require a complete overhaul of the system, whereby all public resources are allocated with the aim of maximising the well-being of the overall population.[38,39] Although this might sound theoretically appealing to some, doing so would be dauntingly complex in practice and no pragmatic solution has been proposed to date.

The *value of knowing* reflects the effect of getting (reliable) test results. On the one hand, people might feel comforted by the reduced uncertainty. On the other hand, people might experience increased anxiety, for example because variants related to a heightened cancer risk were identified. Several authors have argued that the value of knowing is an important concept in the context of genetic testing.[40,41] Indeed, in our review of economic evaluations of PM, we found multiple studies that had included the value of knowing (**Chapter 3**). Although there is some debate as to whether the QALY is sufficiently sensitive to measure the psychological effects of test results[24,42,43],

all the identified studies included the value of knowing in the form of a QALY increase or decrease.

Both *value of hope* and *real option value* relate to patients' attitudes toward risk. Patients might experience value of hope prior to starting treatments that have a long right tail in terms of survival, meaning there is a (small) chance that they will live markedly longer than the expected median. Real option value refers to the possibility of benefitting from future advances in medicine when receiving life-extending treatment. Allocating real option value to life-extending treatments implicitly prioritises length of length over quality of life, in return for a chance that relevant advances in medicine happen in the additional life time. Some patients may indeed prefer more uncertain outcomes with a small chance of long-term survival or prefer prioritising length of life over quality of life. As was argued in **Chapter 3**, however, it is debateable whether all elements that bring value to (some) patients should be paid for by national healthcare payers. Additionally, including real option value in price-setting may result in the healthcare system paying for the same innovation twice: first for the possibility of a new treatment and second for the actualised treatment.

Severity of disease, whereby treatments for very severe diseases are valued more highly than treatments for mild diseases, reflects an equity concern. Arguably, severity of disease is therefore already covered by the *equity* element of value and does not need to be listed separately. Equity concerns, including severity of disease, unmet clinical need, and others, are not uncommonly considered in HTAs.[44,45] Nonetheless, HTA agencies tend to focus on specific elements of equity, with a standardised method for the comprehensive (yet pragmatic) quantification of interventions' equity effects yet to be implemented. Non-disease related health disparities (e.g. socioeconomic, ethnic, geographic) have also received limited attention in HTAs.[46,47] More research on the role of marketing authorisation and reimbursement agencies in decreasing inequities may therefore be valuable.

Finally, the element *scientific spillovers* aims to reward innovation, for example by valuing pharmaceuticals with a new mechanism of action more highly. While innovation may be a legitimate policy goal, several studies have brought into question whether allowing higher prices for pharmaceuticals with a new mechanism of action is the desired way to encourage innovation.[48–51] Research has shown that innovation efforts focus on a few areas of expected high return, while there is virtually no innovation in other areas of large unmet need.[49,52–54] It has been argued that drug prices in the former should in fact be lowered to encourage innovation in the latter.[48,54,55] Drastic innovation also appears to be hampered by the incentive for "me-too" drugs, which are drugs similar to existing

drugs (with a high price) but with a mechanism of action that is different enough to warrant a patent protection.[56,57]

Beyond *value of knowing* and *severity of disease/equity* (both of which are already commonly considered), there may be little merit in including the suggested additional value elements in HTAs. Our guidance recommended that modellers only include elements of value recommended by national HTA guidelines in the base case (Rec. 18).

> The HTA guidelines in both the Netherlands and the UK prescribe severity-weighting, where the value of the cost-effectiveness threshold is varied according to the severity of the treated disease. We used country-specific tools to calculate the QALY shortfall of our patient population (i.e. the difference between expected life expectancy with and without advanced cancer, given the current SoC). In the Netherlands, the estimated QALY shortfall fell in the highest category, implying a threshold value of €80,000. In the UK, the estimated QALY shortfall implied a severity weight of 1.2 and a threshold value of £36,000.

A perhaps somewhat cynical interpretation of the amplification that the task force report received is that pharmaceutical companies may be looking to justify high prices for drugs with limited health benefit by arguing that the drugs offer additional value elements. While this is likely not the sole reason for the publicity around additional value elements, policymakers are encouraged to keep their eyes on the prize: what are the main goals for the healthcare sector? And what might be the best tools to achieve those goals? Additional measures may be needed to minimise the effects of lobbying efforts by the pharmaceutical industry on policymaking and research funding.

Cost-effectiveness outcomes for NTRK-testing and treatment

The ICER of implementing NTRK-testing and subsequent entrectinib treatment was estimated to be €169,957/QALY in the Netherlands (**Chapter 5**). This value is much higher than the cost-effectiveness threshold of €80,000/QALY. Indeed, the probabilistic sensitivity analysis (PSA) suggested that the implementation of NTRK-testing and treatment had a 0.2% chance of being cost-effective. In a scenario analysis that focussed only on the (supposedly already identified) population of NTRK+ patients, the ICER dropped steeply to €38,563/QALY. This illustrates that the consequences of neglecting the introduction of further testing as part of the decision problem may be substantial. It also suggests that the cost-effectiveness of TRK inhibitors may improve as genetic testing becomes more widespread in cancer care and the cost per genetic test decreases. Nonetheless, even if all cancer patients were offered genetic testing as part of standard care, additional testing may be needed prior to starting entrectinib treatment hence

test-related costs and health consequences may still be incurred. This is because RNA-based testing (as opposed to DNA-based testing) may be needed to identify functional NTRK gene fusions and/or because the genetic profile of tumours may change over time (e.g. NTRK fusions can emerge in response to targeted treatment for EGFR mutations).[58,59]

In **Chapter 5**, the focus was on the Netherlands, and the modelled testing pathway was based on the testing protocol suggested in a consensus report by Dutch experts. In **Chapter 6**, the scope of the analysis was expanded to also include Hungary and the UK. Furthermore, the cost-effectiveness of four different testing strategies was compared. In all three countries, the optimal testing strategy was found to be *IHC then RNA-NGS*, where all patients receive IHC testing and those with a positive test result receive confirmatory RNA-NGS testing. Although the testing strategy proposed by the Dutch expert group (*Stratified*) rendered slightly better QALY outcomes than *IHC then RNA-NGS*, its ΔNMB was lower. This suggests that health economic modelling may be a valuable addition to expert opinion when deciding on the optimal implementation of (genetic) tests.

All evaluated testing strategies rendered negative ΔNMB in all three countries. When focussing only on the NTRK+ population (i.e. assuming that no additional testing is needed to establish eligibility for entrectinib treatment), ΔNMB turned positive in the Netherlands and in the UK but remained (deeply) negative in Hungary. This suggests that there is some potential for entrectinib to become cost-effective in the Netherlands and the UK, while in Hungary the implementation of entrectinib would most likely cause a net loss to the healthcare system unless the price of entrectinib was significantly reduced. The cost-effectiveness outcomes for entrectinib fit in a broader pattern, where most of the new oncology drugs come at high prices yet offer limited health benefit.[22,60] The high prices for oncology drugs affect access to cancer care in low- and middle-income countries (LMICs), as LMICs account for only 5% of global expenditure on cancer despite incurring an estimated 80% of the global cancer disease burden.[61,62] In high-income countries, too, the price-setting for oncology drugs may prove unsustainable. For example, a study focussing on the German healthcare system estimated that, based on the current price-setting for cancer drugs relative to their health benefit, paying for *curative* cancer drugs would more than triple national healthcare expenditure.[63] Measures may be needed to moderate cancer drug prices, possibly including a more restricted use of expedited approval pathways and more education of health professionals and the public alike to prevent unrealistic expectations about the benefits of new pharmaceuticals.[22,64]

Finally, the “tumour-agnostic” or “histology-independent” label of entrectinib and larotrectinib seems to imply that tumour location and/or histology are irrelevant to the choice for TRK inhibitor treatment. Indeed, for a lack of tumour type specific information on the effectiveness of entrectinib, we estimated a single ICER for the cost-effectiveness

of NTRK-testing and treatment. However, evidence suggests there may be significant heterogeneity in (cost-)effectiveness across tumour types.[65] Allowing single estimates of (cost-)effectiveness for a pool of tumour types is therefore likely inappropriate and marketing authorisation and reimbursement agencies are encouraged to ensure that heterogeneity is appropriately reflected in clinical evidence for histology-independent therapies.

Concluding remarks

As explained in **Chapter 1**, "personalised medicine" is a future-oriented term. De Grandis & Halgunset describe the term as a "cluster of visions for the future" and argue that the term is purposefully rhetorically appealing to rally support and further investment.[66] **Chapter 2** showed that so far there has been a limited actualisation of the high hopes and expectations surrounding PM, with a median QALY gain of 0.03 and median ΔNMB of Int$ 18 across PM interventions. **Chapter 5** and **Chapter 6** showed that the implementation of NTRK-testing and subsequent treatment with the histology-independent therapy entrectinib is likely not cost-effective (at least in Hungary, the Netherlands, and the UK) given the current context and price-setting. However, all hope is not lost. The regression analysis in **Chapter 2** suggested that genetic testing to prevent adverse drug reactions may be particularly valuable, and several PM interventions with high ΔNMB were identified. There appears to be value in some forms of PM but not in others, suggesting that "personalised medicine" may be too broad a term to guide investment and funding decisions. Also, as argued above, the word "personalised" can be interpreted in many ways and to some extent conceals the field's focus on genetic tests, pharmaceuticals, and oncology. Using more modest, specific terms, such as "stratified medicine" and "genomic medicine" may lead to more modest and realistic expectations regarding the value of new interventions.

Several recommendations were made in this chapter (**Chapter 7**) for a (more) successful implementation of PM. Stricter requirements ought to be set for clinical evidence submitted to marketing authorisation agencies and reimbursement agencies. There also need to be stricter requirements regarding follow-up studies to be conducted in cases where the original evidence is limited. Standardised procedures should be developed to ensure that clinical guidelines and reimbursement status are updated if follow-up studies show insufficient treatment effectiveness or severe side effects. More widespread collection of clinical and genomic registry data may be needed. National HTA guidelines may have to be updated with additional guidance on the modelling of testing pathways, to ensure testing pathways are consistently included and modelled across reimbursement dossiers. More research funding and policy measures are needed to ensure that the implementation of PM does not exacerbate existing health inequalities.

The issue of unequal referral rates to genetic testing needs to be addressed, and patient populations both in clinical trials and in registry data need to be more representative of real-life patient populations. Measures are needed to ensure that areas of technological innovation better align with areas of high unmet clinical need. Measures are needed to temper prices for (oncology) drugs with limited health benefit. And finally, policymakers, research funders, and researchers alike are encouraged to exercise caution around the topic of PM.[67] Not all PM interventions offer added value to the healthcare system and society at large. So far, the field of PM has also had little attention for primary prevention, despite a huge potential for health gains through improved (targeting and tailoring of) primary prevention. Let us keep our eyes on the prize and only allocate resources to interventions and areas of innovation and implementation that are likely to offer (sizeable) added value.

References

1. Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY Gains in Applied Studies: A Review of Cost-Utility Analyses Published in 2010. ***PharmacoEconomics***. 2014;32(4):367-375. doi:10.1007/s40273-014-0136-z
2. Kapczynski A. The Political Economy of Market Power in Pharmaceuticals. ***J Health Polit Policy Law***. 2023;48(2):215-239. doi:10.1215/03616878-10234184
3. Busfield J. Pills, Power, People: Sociological Understandings of the Pharmaceutical Industry. ***Sociology***. 2006;40(2):297-314. doi:10.1177/0038038506062034
4. Vanderelst D, Speybroeck N. Scientometrics reveals funding priorities in medical research policy. ***J Informetr***. 2013;7(1):240-247. doi:10.1016/j.joi.2012.10.004
5. Ford ES, Bergmann MM, Kröger J, Schienkiewitz A, Weikert C, Boeing H. Healthy living is the best revenge: findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam study. ***Arch Intern Med***. 2009;169(15):1355-1362. doi:10.1001/archinternmed.2009.237
6. World Health Organization, Public Health Agency of Canada, eds. ***Preventing Chronic Diseases: A Vital Investment***. World Health Organization ; Public Health Agency of Canada; 2005.
7. Anand P, Kunnumakara AB, Sundaram C, et al. Cancer is a Preventable Disease that Requires Major Lifestyle Changes. ***Pharm Res***. 2008;25(9):2097-2116. doi:10.1007/s11095-008-9661-9
8. van der Giessen JAM, van Riel E, Velthuizen ME, van Dulmen AM, Ausems MGEM. Referral to cancer genetic counseling: do migrant status and patients' educational background matter? ***J Community Genet***. 2017;8(4):303-310. doi:10.1007/s12687-017-0326-4
9. Muller C, Lee SM, Barge W, et al. Low Referral Rate for Genetic Testing in Racially and Ethnically Diverse Patients Despite Universal Colorectal Cancer Screening. ***Clin Gastroenterol Hepatol***. 2018;16(12):1911-1918.e2. doi:10.1016/j.cgh.2018.08.038
10. Manrriquez E, Chapman JS, Mak J, Blanco AM, Chen L may. Disparities in genetics assessment for women with ovarian cancer: Can we do better? ***Gynecol Oncol***. 2018;149(1):84-88. doi:10.1016/j.ygyno.2017.10.034
11. Dharwadkar P, Greenan G, Stoffel EM, et al. Racial and Ethnic Disparities in Germline Genetic Testing of Patients With Young-Onset Colorectal Cancer. ***Clin Gastroenterol Hepatol***. 2022;20(2):353-361.e3. doi:10.1016/j.cgh.2020.12.025
12. Baldwin LM, Trivers KF, Andrilla CHA, et al. Accuracy of Ovarian and Colon Cancer Risk Assessments by U.S. Physicians. ***J Gen Intern Med***. 2014;29(5):741-749. doi:10.1007/s11606-014-2768-2
13. Smith-Uffen M, Bartley N, Davies G, Best M. Motivations and barriers to pursue cancer genomic testing: A systematic review. ***Patient Educ Couns***. 2021;104(6):1325-1334. doi:10.1016/j.pec.2020.12.024
14. Pearl J, Glymour M, Jewell NP. ***Causal Inference in Statistics: A Primer***. John Wiley & Sons; 2016.
15. Beaulieu-Jones BK, Finlayson SG, Yuan W, et al. Examining the Use of Real-World Evidence in the Regulatory Process. ***Clin Pharmacol Ther***. 2020;107(4):843-852. doi:10.1002/cpt.1658
16. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. ***BMJ Open***. 2016;6(6):e011666. doi:10.1136/bmjopen-2016-011666
17. Haslam A, Gill J, Prasad V. The response rate of alternative treatments for drugs approved on the basis of response rate. ***Int J Cancer***. 2021;148(3):713-722. doi:10.1002/ijc.33231
18. Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. ***BMC Med***. 2018;16(1):29. doi:10.1186/s12916-018-1017-7
19. Thall PF, Wathen JK. Practical Bayesian adaptive randomisation in clinical trials. ***Eur J Cancer***. 2007;43(5):859-866. doi:10.1016/j.ejca.2007.01.006
20. Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz HM, Ross JS. Assessment of Clinical Trials Supporting US Food and Drug Administration Approval of Novel Therapeutic Agents, 1995-2017. ***JAMA Netw Open***. 2020;3(4):e203284. doi:10.1001/jamanetworkopen.2020.3284
21. Hwang TJ, Ross JS, Vokinger KN, Kesselheim AS. Association between FDA and EMA expedited approval programs and therapeutic value of new medicines: retrospective cohort study. ***BMJ***. 2020;371:m3434. doi:10.1136/bmj.m3434

22. Schnog JJB, Samson MJ, Gans ROB, Duits AJ. An urgent call to raise the bar in oncology. ***Br J Cancer***. 2021;125(11):1477-1485. doi:10.1038/s41416-021-01495-7
23. Love-Koh J, Peel A, Rejon-Parrilla JC, et al. The Future of Precision Medicine: Potential Impacts for Health Technology Assessment. ***PharmacoEconomics***. 2018;36(12):1439-1451. doi:10.1007/s40273-018-0686-6
24. Annemans L, Redekop K, Payne K. Current Methodological Issues in the Economic Assessment of Personalized Medicine. ***Value Health***. 2013;16(6, Supplement):S20-S26. doi:10.1016/j.jval.2013.06.008
25. Lohse S. Mapping uncertainty in precision medicine: A systematic scoping review. ***J Eval Clin Pract***. 2023;29(3):554-564. doi:10.1111/jep.13789
26. 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(5):739-749. doi:10.1111/j.1524-4733.2008.00502.x
27. Sacks Henry, Chalmers TC., Smith Harry. Randomized versus historical controls for clinical trials. ***Am J Med***. 1982;72(2):233-240. doi:10.1016/0002-9343(82)90815-4
28. Guerrero S, López-Cortés A, Indacochea A, et al. Analysis of Racial/Ethnic Representation in Select Basic and Applied Cancer Research Studies. ***Sci Rep***. 2018;8(1):13978. doi:10.1038/s41598-018-32264-x
29. Sharrocks K, Spicer J, Camidge DR, Papa S. The impact of socioeconomic status on access to cancer clinical trials. ***Br J Cancer***. 2014;111(9):1684-1687. doi:10.1038/bjc.2014.108
30. Knepper TC, McLeod HL. When will clinical trials finally reflect diversity? ***Nature***. 2018;557(7704):157-159. doi:10.1038/d41586-018-05049-5
31. Alpert AB, Scout N f. n., Schabath MB, Adams S, Obedin-Maliver J, Safer JD. Gender- and Sexual Orientation– Based Inequities: Promoting Inclusion, Visibility, and Data Accuracy in Oncology. ***Am Soc Clin Oncol Educ Book***. 2022;(42):542-558. doi:10.1200/EDBK_350175
32. Eskander MF, Gil L, Beal EW, et al. Access Denied: Inequities in Clinical Trial Enrollment for Pancreatic Cancer. ***Ann Surg Oncol***. 2022;29(2):1271-1277. doi:10.1245/s10434-021-10868-4
33. Seidler EM, Keshaviah A, Brown C, Wood E, Granick L, Kimball AB. Geographic distribution of clinical trials may lead to inequities in access. ***Clin Investig***. 2014;4(4):373-380. doi:10.4155/cli.14.21
34. Pittell H, Calip GS, Pierre A, et al. Racial and Ethnic Inequities in US Oncology Clinical Trial Participation From 2017 to 2022. ***JAMA Netw Open***. 2023;6(7):e2322515. doi:10.1001/jamanetworkopen.2023.22515
35. 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(1):118-125. doi:10.1016/j.jval.2016.09.001
36. Lakdawalla DN, Doshi JA, Garrison LP, Phelps CE, Basu A, Danzon PM. Defining Elements of Value in Health Care—A Health Economics Approach: An ISPOR Special Task Force Report [3]. ***Value Health***. 2018;21(2):131-139. doi:10.1016/j.jval.2017.12.007
37. Garrison LP, Towse A. A Strategy to Support Efficient Development and Use of Innovations in Personalized Medicine and Precision Medicine. ***J Manag Care Spec Pharm***. 2019;25(10):1082-1087. doi:10.18553/jmcp.2019.25.10.1082
38. Brouwer WBF, Culyer AJ, van Exel NJA, Rutten FFH. Welfarism vs. extra-welfarism. ***J Health Econ***. 2008;27(2):325-338. doi:10.1016/j.jhealeco.2007.07.003
39. Culyer AJ. ***The Dictionary of Health Economics, Third Edition***. Edward Elgar Publishing; 2014.
40. Grosse SD, Wordsworth S, Payne K. Economic methods for valuing the outcomes of genetic testing: beyond cost-effectiveness analysis. ***Genet Med***. 2008;10(9):648-654. doi:10.1097/GIM.0b013e3181837217
41. Kohler JN, Turbitt E, Biesecker BB. Personal utility in genomic testing: a systematic literature review. ***Eur J Hum Genet***. 2017;25(6):662-668. doi:10.1038/ejhg.2017.10
42. Bouttell J, Heggie R, Oien K, et al. Economic evaluation of genomic/genetic tests: a review and future directions. ***Int J Technol Assess Health Care***. 2022;38(1):e67. doi:10.1017/S0266462322000484
43. Payne K, McAllister M, Davies LM. Valuing the Economic Benefits of Complex Interventions: When Maximising Health Is Not Sufficient. ***Health Econ***. 2013;22(3):258-271. doi:10.1002/hec.2795
44. Guindo LA, Wagner M, Baltussen R, et al. From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking. ***Cost Eff Resour Alloc***. 2012;10(1):9. doi:10.1186/1478-7547-10-9

45. Skedgel C, Henderson N, Towse A, Mott D, Green C. Considering Severity in Health Technology Assessment: Can We Do Better? ***Value Health***. 2022;25(8):1399-1403. doi:10.1016/j.jval.2022.02.004
46. Johri M, Norheim OF. CAN COST-EFFECTIVENESS ANALYSIS INTEGRATE CONCERNS FOR EQUITY? SYSTEMATIC REVIEW. ***Int J Technol Assess Health Care***. 2012;28(2):125-132. doi:10.1017/S0266462312000050
47. Lane H, Sarkies M, Martin J, Haines T. Equity in healthcare resource allocation decision making: A systematic review. ***Soc Sci Med***. 2017;175:11-27. doi:10.1016/j.socscimed.2016.12.012
48. Canoy M, Tichem J. Lower drug prices can improve innovation. ***Eur Compet J***. 2018;14(2-3):278-304. doi:10.1080/17441056.2018.1512231
49. Angelis A, Polyakov R, Wouters OJ, Torreele E, McKee M. High drug prices are not justified by industry's spending on research and development. ***BMJ***. 2023;380:e071710. doi:10.1136/bmj-2022-071710
50. ***Drug Industry: Profits, Research and Development Spending, and Merger and Acquisition Deals***. Accessed August 27, 2023. https://apps.dtic.mil/sti/citations/AD1174880
51. Light DW, Warburton R. Demythologizing the high costs of pharmaceutical research. ***BioSocieties***. 2011;6(1):34-50. doi:10.1057/biosoc.2010.40
52. Barrenho E, Miraldo M, Smith PC. Does global drug innovation correspond to burden of disease? The neglected diseases in developed and developing countries. ***Health Econ***. 2019;28(1):123-143. doi:10.1002/hec.3833
53. Barrenho E, Halmai R, Miraldo M, et al. Inequities in cancer drug development in terms of unmet medical need. ***Soc Sci Med***. 2022;302:114953. doi:10.1016/j.socscimed.2022.114953
54. Panteli D, Edwards S. ***Ensuring Access to Medicines: How to Stimulate Innovation to Meet Patients' Needs?*** (Richardson E, Palm W, Mossialos E, eds.). European Observatory on Health Systems and Policies; 2018. Accessed September 2, 2023. http://www.ncbi.nlm.nih.gov/books/NBK526400/
55. Herper M. Could High Drug Prices Be Bad For Innovation? Forbes. Accessed September 2, 2023. https://www.forbes.com/sites/matthewherper/2014/10/23/could-high-drug-prices-be-bad-for-innovation/
56. Vincent Rajkumar S. The high cost of prescription drugs: causes and solutions. ***Blood Cancer J***. 2020;10(6):71. doi:10.1038/s41408-020-0338-x
57. Leoni P, Sandroni A. Can patent duration hinder medical innovation. ***Int J Health Econ Manag***. 2016;16(4):397-406. doi:10.1007/s10754-016-9198-0
58. Zhu L, Jiang M, Wang H, et al. A narrative review of tumor heterogeneity and challenges to tumor drug therapy. ***Ann Transl Med***. 2021;9(16):1351. doi:10.21037/atm-21-1948
59. Liu F, Wei Y, Zhang H, Jiang J, Zhang P, Chu Q. NTRK Fusion in Non-Small Cell Lung Cancer: Diagnosis, Therapy, and TRK Inhibitor Resistance. ***Front Oncol***. 2022;12. Accessed August 27, 2023. https://www.frontiersin.org/articles/10.3389/fonc.2022.864666
60. Vokinger KN, Hwang TJ, Daniore P, et al. Analysis of Launch and Postapproval Cancer Drug Pricing, Clinical Benefit, and Policy Implications in the US and Europe. ***JAMA Oncol***. 2021;7(9):e212026. doi:10.1001/jamaoncol.2021.2026
61. Ocran Mattila P, Ahmad R, Hasan SS, Babar ZUD. Availability, Affordability, Access, and Pricing of Anti-cancer Medicines in Low- and Middle-Income Countries: A Systematic Review of Literature. ***Front Public Health***. 2021;9. Accessed August 27, 2023. https://www.frontiersin.org/articles/10.3389/fpubh.2021.628744
62. World Health Organization. ***Technical Report: Pricing of Cancer Medicines and Its Impacts: A Comprehensive Technical Report for the World Health Assembly Resolution 70.12: Operative Paragraph 2.9 on Pricing Approaches and Their Impacts on Availability and Affordability of Medicines for the Prevention and Treatment of Cancer***. World Health Organization; 2018. Accessed August 27, 2023. https://apps.who.int/iris/handle/10665/277190
63. Gandjour A. The price of curing cancer. ***BMC Health Serv Res***. 2021;21:1328. doi: 10.1186/s12913-021-07327-x
64. Prasad V, Mailankody S. Research and Development Spending to Bring a Single Cancer Drug to Market and Revenues After Approval. ***JAMA Intern Med***. 2017;177(11):1569-1575. doi:10.1001/jamainternmed.2017.3601

65. Murphy P, Claxton L, Hodgson R, et al. Exploring Heterogeneity in Histology-Independent Technologies and the Implications for Cost-Effectiveness. ***Med Decis Making***. 2021;41(2):165-178. doi:10.1177/0272989X20980327
66. De Grandis G, Halgunset V. Conceptual and terminological confusion around personalised medicine: a coping strategy. ***BMC Med Ethics***. 2016;17(1):43. doi:10.1186/s12910-016-0122-4
67. Joyner MJ, Paneth N. Promises, promises, and precision medicine. ***J Clin Invest***. 2019;129(3):946-948. doi:10.1172/JCI126119

CHAPTER 8

Summary, Samenvatting,
Acknowledgements and About the author

Summary

In this thesis a health economics perspective is applied to the phenomenon of personalised medicine.

Chapter 1: General introduction

Chapter 1 contains a detailed look at the origin and definition of the concept "personalised medicine". The term first appeared in a 1999 Wall Street Journal article about a research project that was aiming to develop targeted pharmaceuticals "for each unique genetic profile". Over time, the initial focus of personalised medicine – on pharmaceuticals and on genetic profiles – expanded. The term came to be associated with the broader idea that healthcare should be (more) tailored to our individual characteristics. Yet debate emerged about the exact meaning of "personalised medicine". Philosophers De Grandis & Halgunset have argued that the term describes a cluster of visions for the future of healthcare more than an existing reality. They suggest that the meaning of personalised medicine is therefore fluid and may shift over time, depending on which technologies actualise and political developments such as "who gets involved", "who pays, and for what". Based on a systematic literature review on the various uses of the term "personalised medicine" by Schleidgen et al., and to best capture the available technologies so far, the following definition of personalised medicine is adopted for this thesis:

> *a medical model that bases therapeutic choice on the result of gene profiling or aims to correct pathogenic gene mutations*

Chapter 1 also outlines key concepts and methods in the field of health economics or, more specifically, its subfield of health technology assessment (HTA). HTA is used to assess health technologies, also called "interventions". The properties and effects of different healthcare interventions are evaluated, using the same methodological approach and the same criteria each time. This way, HTA provides insights into the value for money that different healthcare interventions provide *relative to each other* and hence can support decision making on resource allocation.

The main objectives of the thesis are as follows: to provide insights into the added value of personalised medicine and to investigate methodological challenges in evaluating personalised medicine. A cost-effectiveness analysis of a personalised medicine intervention is performed as a case study.

Chapter 2: The net benefit of personalised medicine

Personalised medicine (PM) has been surrounded by excitement and anticipation. Indeed, a better tailoring of healthcare may have many positive consequences – such as improved treatment effectiveness, less adverse drug reactions, more diseases prevented – that in turn could lead to increased quality and length of life and/or a reduction in wasteful healthcare spending. At the same time, though, there are concerns about the potentially high costs of (implementing) PM. Some of the PM interventions that have come onto the market in recent years have come with large price tags. And more widespread implementation of gene testing in clinical practice may be rather expensive.

Chapter 2 presents an analysis aiming to get a better understanding of the added value, or the net benefit, of PM. A systematic literature review of economic evaluations of PM interventions was performed. For all pair-wise comparisons between PM interventions and non-PM interventions that were identified, the estimated incremental costs (Δcosts) and health effects (ΔQALY) were captured and used to calculate incremental net monetary benefit (ΔNMB). To investigate whether certain types of PM might offer more benefit than others, regression analyses were performed with Δcosts, ΔQALY and ΔNMB as dependent variables and various intervention characteristics as independent variables. Random intercepts were used to cluster studies according to country.

Out of 4,774 studies reviewed, 128 met the inclusion criteria, providing cost-effectiveness outcomes for 279 PM interventions. Most studies were set in the United States (48%) and the United Kingdom (16%). No economic evaluations of PM interventions were found for low- or lower-middle income countries. Cancer treatments (60%) and pharmaceutical interventions (72%) occurred frequently. Approximately 32% of studies were found to have some form of sponsorship from the pharmaceutical industry.

The median incremental health benefit of PM interventions was found to be 0.03 quality-adjusted life year (QALY) per patient, or roughly 11 quality-adjusted days. The added health benefits of PM seem therefore modest on average. The results also show that the ΔNMB of PM interventions centred around zero, with a median of 18 international dollars (Int$). This implies that the health benefits of PM tend to be counterbalanced by the associated cost increases.

Nonetheless, there is a variety in outcomes, with ΔNMB as high as Int$ 21,615 at the 95th percentile and ΔNMB as low as Int$ –91,832 at the 5th percentile. The regression analysis showed gene therapies to be a distinct subcategory of PM. Gene therapies are associated with (much) larger health benefits, higher costs and lower ΔNMB than other PM interventions. This suggests that gene therapies have the potential to bring large health benefits but tend to render a net loss to the healthcare system at current price-

setting. However, the number of gene therapies included in our analysis was limited and the findings were found to be uncertain.

The regression analysis also suggests that PM interventions aimed at preventing adverse drug reactions (ADR) tend to have higher ΔNMB, while PM interventions focussing on the identification (and treatment) of patients likely to respond to treatment tend to have lower ΔNMB. A possible explanation for this is that many of the interventions in the former category stratify patients to existing (off-patent) treatments, whereas many interventions in the latter category stratify toward newly developed treatments. While much of the focus in PM has been on developing new treatments for new subsets of patients, further research and development to better stratify *existing* care may be valuable.

Also, despite the general tendency of PM interventions toward a ΔNMB of zero, we identified various interventions with (very) negative ΔNMB. National and international pricing policies may be needed to reduce the costs of these interventions, to prevent healthcare systems facing negative net benefit when implementing the interventions.

Chapter 3: Guidance for the harmonisation and improvement of economic evaluations of personalised medicine

Personalised medicine, with its heavy focus on genetic markers and innovative approaches, differs from much of the healthcare that came before. In response to concerns that existing HTA methods might be inadequate for assessing the cost-effectiveness of PM interventions, we investigated the methodological challenges that PM may bring. An overview of the possible challenges as well as guidance on how to proceed in the face of these challenges are presented in Chapter 3.

We performed a targeted literature review and held interviews with 18 experts for an initial overview of possible methodological challenges and ways to address them. The systematic literature review of economic evaluations of PM (Chapter 2) was also used to investigate the methodological approaches that have been used so far. Once an initial set of recommendations was drafted, they were discussed in a workshop with around 30 stakeholders and subsequently adjusted. The final guidance contained 23 recommendations.

Although we found some studies arguing that adjustments to standard HTA approaches may be needed for PM interventions, many of the interviewed experts stated that PM should be subject to the same methodological framework as other interventions to ensure consistency and comparability. The first two recommendations of the guidance reflect this sentiment. The use of the standard perspective (Recommendation [Rec.] 1) and standard discount rates (Rec. 2) as indicated by national HTA guidelines is encouraged, for PM and non-PM interventions alike.

The (genetic) testing of patients and subsequent stratification to different treatments plays a central role in PM. Yet most health economic models to date start from the moment that patients begin treatment (i.e. after any testing has concluded). To ensure a more complete reflection of the reality, the guidance advocates for the inclusion of the testing period in economic evaluations of PM. Eight recommendations are given for the accurate modelling of test-treatment pathways.

Given that different technologies might be used for genetic profiling and given that tests might be applied at different times and in varying combinations, modellers are urged to not consider just one of the available test options. Instead, all relevant test-treatment pathways should be identified, and a justification should be given as to why the pathways included in the model were selected (Recommendation [Rec.] 3). We also stress that including the costs and health effects of testing only for the patients who end up receiving a specific treatment, as is sometimes done, gives an incomplete depiction of the consequences of introducing a test-treatment combination into clinical practice. Rather, the cost and health effects for *all* patients who receive testing should be considered in the model, including for patients who get false-positive or false-negative test results (Rec. 4). Furthermore, when introducing a test for hereditary conditions, a downstream consequence of introducing the test may be that relatives of the initial patients with a positive result (the "index patients") also become eligible for testing. Modellers are encouraged to include the cost and health outcomes of testing relatives in the economic evaluation of the index patients (Rec. 10). Other topics that modellers are encouraged to consider are: relevance of the available data on diagnostic accuracy to the patient population in the model (Rec. 5); the effect of different possible cut-off values for the test results on cost-effectiveness outcomes (Rec. 6); possible interdependence between test results when multiple tests are used (Rec. 7); morbidity/mortality resulting from waiting periods related to the testing procedure (Rec. 8); and possible variation in testing costs across laboratories (Rec. 9).

The more patients are stratified into specific subgroups, the smaller the sample size of each patient population becomes. The large randomised controlled trials (RCTs) that have been used over the past decades to measure comparative effectiveness may therefore not always be feasible in PM. Indeed, submissions for marketing authorisation and reimbursement without RCT evidence seem increasingly common. However, doubts have been raised whether foregoing an RCT was justified in all of the non-RCT dossiers that have so far been submitted to regulators such as the European Medicines Agency and the US Food and Drug Administration. Also, there are concerns about the consequences of allowing interventions onto the market based on limited clinical evidence, including possible harm to patients and a suboptimal allocation of healthcare resources. We therefore encourage the use of effectiveness data from trials with two (or

more) alternative treatment strategies (Rec. 11). Nonetheless, given that those tasked with the evaluation of an intervention's cost-effectiveness may not have control over the available effectiveness data, the guidance offers several recommendations aimed at upholding quality standards as much as possible. These include recommendations urging modellers to be transparent about data sources and assumptions, when using surrogate outcomes to estimate final outcomes (Rec. 12) and when estimating the association between a genetic marker and clinical outcomes (Rec. 15). When external data is used to estimate comparator effectiveness, modellers should account for possible changes in treatment effectiveness over time (Rec. 13) as well as the prevalence of relevant genetic markers in the intervention vs. comparator populations (Rec. 14). Additional recommendations are made to improve the extrapolation of short-term data from clinical studies to long-term health outcomes. Modellers are encouraged to use expert opinion alongside statistical fit to choose the survival model (Rec. 16) and to account for any excess mortality (compared to the general population) for long-term survivors of a disease (Rec. 17).

In 2018 a report was released by the ISPOR Value Assessment Framework Special Task Force arguing that the quality-adjusted life year (QALY) insufficiently reflects the value of healthcare interventions. The report listed several "additional elements of value", including value of knowing, fear of contagion, insurance value, value of hope, real option value, severity of disease, equity, and scientific spillovers. Some have argued the additional value elements may be particularly important in PM. However, there are several concerns with the suggested list of value elements. These include conceptual ambiguity of the different value elements and possible overlap between them, a one-sided focus on positive elements of value, and a blurring of individual preferences versus societal priority-setting. The guidance therefore discourages the inclusion of value elements beyond the those recommended by national HTA guidelines and states that if additional value elements are included in a sensitivity analysis, possible elements of negative value should be equally considered (Rec. 18).

Patient and clinician compliance in clinical practice may be far removed from the optimised conditions in clinical trials, especially when genetic tests are involved. For example, there may be patient mistrust toward genetic testing, or the complicated nature of genetic testing may hamper its adoption by clinicians. We therefore recommend that, if an insight into the intervention's cost-effectiveness *under realistic circumstances* is required, parameters reflecting patient and clinician compliance should be included in the economic evaluation (Rec. 19). Previous studies have shown that non-white patients and patients with a lower socioeconomic status are much less likely to be referred for genetic testing. We therefore also recommended that modellers consider possible variation in compliance across societal groups (Rec. 20).

Large uncertainty is often mentioned as a key challenge in PM. Both parameter uncertainty (i.e. uncertainty as to how closely the estimated values for input parameters approximate the true values) and structural uncertainty (i.e. uncertainty whether the model assumptions and selected methods are appropriate) are likely to arise. Parameter uncertainty may for example be exacerbated by the unclear causal status between genetic markers and clinical outcomes. Structural uncertainty may arise because of complex testing procedures with many possible test-treatment pathways, or because various methodological choices have to be made to estimate the outcome of interest based on limited available data. The guidance states that when expert judgement is used to estimate values for the input parameters in the model, the elicited values should be synthesised into a probability distribution so that the uncertainty in expert judgement can be reflected (Rec. 21). Modellers are also encouraged to identify the uncertainties in structural assumptions and decisions and to investigate the impact on the cost-effectiveness results through a sensitivity analysis, ideally by parameterising the structural aspects (Rec. 22).

In cases where uncertainty is too large for a reimbursement decision to be made, managed entry agreements may be struck between manufacturer and healthcare payer. Various types of managed entry agreement exist (e.g. payment by result, price-volume agreements, etc.), with a range of conditions that can be included. Health economic modelling can be used to optimise the conditions of a managed entry agreement. The guidance states that if a managed entry agreement is being considered for an intervention, its conditions may be included in the model evaluating the intervention (Rec. 23).

Chapter 4: Prognostic value of the NTRK fusion biomarker in the Netherlands

The methodological guidance of Chapter 3 is applied to a case study of a PM intervention: testing for oncogenic neurotrophic tyrosine receptor kinase (NTRK) gene fusions among cancer patients and subsequently providing the histology-independent treatment entrectinib to those with a positive test result.

NTRK fusions result from chromosomal rearrangements that cause the 3′ region of the NTRK gene to join the 5′ region of a fusion partner gene. Such fusions may result in TRK fusion proteins with constitutively active tyrosine kinases, which can lead to tumour growth. Entrectinib is an inhibitor for the three most common types of TRK protein: TRKA, TRKB and TRKC (encoded by the NTRK1, NTRK2 and NTRK3 gene, respectively). A pooled analysis for three phase I/II trials for entrectinib appeared promising. However, the trials were single-arm trials. To estimate entrectinib's comparative effectiveness, we had to rely on historical data. A key issue is that the entrectinib trial population consists of patients

with tumours with NTRK gene fusions ("NTRK+ patients"), while the available comparator data contains mostly patients without such tumours ("NTRK- patients"). To be able to account for this difference, we estimated the prognostic value of the NTRK gene fusion.

The clinicogenomic Hartwig Medical Foundation (HMF) database was used to obtain a sample of cancer patients who were treated in Dutch clinical practice. In a sub-cohort of the HMF database, the Center for Personalised Cancer Treatment (CPCT-02) study, whole-genome sequencing of tumour DNA was performed for thousands of patients from 44 hospitals in the Netherlands. For all patients in the CPCT-02 study, we extracted data on various genetic markers of the tumour DNA as well as several clinical variables.

Patients were then classified into two cohorts: NTRK+ patients and NTRK- patients. Patients who had NTRK fusions in the tumour DNA that were deemed "likely functional" (i.e. leading to the expression of TRK fusion proteins) were included in the NTRK+ cohort. Patients without NTRK fusions in the tumour DNA, or with NTRK fusions that did not meet the conditions for likely functionality, were included in the NTRK- cohort. We performed a propensity score matching analysis to identify a subgroup of NTRK- patients similar to the group of NTRK+ patients. NTRK+ patients were matched to NTRK- patients using a 1:4 matching ratio. Within each tumour type, patients were matched on the variables age, sex, year of biopsy, and number of previous lines of therapy. We subsequently analysed the survival of NTRK+ and NTRK- patients using the Kaplan-Meier method.

Among 3,556 patients from the CPCT-02 study with known tumour location, 24 NTRK+ patients were identified. NTRK+ patients were distributed across nine different tumour types: bone/soft tissue, breast, colorectal, head and neck, lung, pancreas, prostate, skin and urinary tract. NTRK fusions involving the NTRK3 gene (46%) and NTRK1 gene (33%) were most common.

The survival analysis rendered a hazard ratio (HR) of 1.44 (95% CI 0.81–2.55) for NTRK+ patients. HR values smaller than 1 are included in the 95% CI, meaning the estimated point value is statistically non-significant. Nonetheless, three prior studies (in slightly different settings) produced similar estimates, making it plausible that NTRK+ patients indeed have a worse prognosis than NTRK− patients. In a sensitivity analysis where NTRK+ patients with concurrent oncogenic biomarkers were excluded, we found a lower HR than in the main analysis (1.20, 95% CI 0.61–2.36). This suggests that the main analysis may have overestimated the prognostic value of the NTRK gene fusion *per se*.

An important limitation of the study is that we estimated a single HR value for all NTRK+ patients, despite prior studies suggesting that there may be heterogeneity in the prognostic value of NTRK gene fusions across different tumour types. The small sample

of 24 NTRK+ patients left us with little option, yet caution should be exercised when using the results of this study (and those of the three prior studies).

Chapter 5: Cost-effectiveness analysis of treating patients with NTRK-positive cancer with the histology-independent therapy entrectinib

Entrectinib and fellow TRK inhibitor larotrectinib came to the market as the first two molecularly targeted therapies with a histology-independent (also called tissue- or tumour-agnostic) label. Histology-independent therapies can be prescribed regardless of the tissue of origin of the tumour, reflecting an innovative approach to cancer care. Chapter 5 presents a cost-effectiveness analysis of treating NTRK+ patients with entrectinib. The intervention is defined as: NTRK gene fusion testing for all patients with locally advanced or metastatic solid tumours, followed by treatment with entrectinib in NTRK+ patients and standard of care (SoC) in NTRK– patients. In the comparator arm, all patients receive SoC and no NTRK testing is performed. Larotrectinib ideally would have been included as another comparator. However, single-arm trials were performed for both entrectinib and larotrectinib. Because of key differences in the respective trial populations, larotrectinib could not be included in the economic evaluation of entrectinib.

The introduction of entrectinib into clinical practice requires the introduction (/ expansion) of NTRK testing to identify eligible patients. We therefore included the testing phase in the model, reflected by a decision tree. Various chance nodes were included in the decision tree, including the probabilities of (re-)biopsies being needed and the probability of patients dying by the end of the testing period. Patients who survive the testing period subsequently enter a microsimulation model. The individual-level state transition model includes the health states “alive and on treatment”, “alive and off treatment”, and “dead”. Patients receive entrectinib if they tested NTRK-positive in the decision tree and are treated with SoC if they tested NTRK-negative. In the comparator group, patients do not receive testing (i.e. they skip the decision tree and go straight into the microsimulation model) and are all given SoC.

Input data on the effectiveness of entrectinib for NTRK+ patients were obtained from the entrectinib basket trials, in the form of overall survival (OS) and time-to-treatment discontinuation (TTD). Because of the small sample size of the entrectinib trials and the resulting lack of reliable tumour-specific estimates, we had to assume that the effectiveness of entrectinib was constant across tumour types. Nonetheless, for the comparator arm (SoC) we estimated tumour-specific effectiveness. The HMF database was used to estimate OS and TTD for NTRK- patients receiving SoC. To calculate outcomes for NTRK+ patients receiving SoC, the estimated HR of 1.44 (Chapter 4) was applied to the OS and TTD estimates for NTRK- patients. We also obtained tumour-

specific cost estimates for both testing and SoC treatments. Finally, based on previous research findings that the quality of life decreases as patients approach death, we incorporated (age- and sex-specific) patient utility in our model as a function of time to death.

The results show that the testing for NTRK fusions among patients with advanced cancer and subsequent treatment with entrectinib in NTRK+ patients and SoC in NTRK– patients is associated with a QALY gain of 0.0043 at an increased cost of €732 per patient as compared with no testing and SoC for all patients. The incremental cost-effectiveness ratio (ICER) is €169,957/QALY, far above the Dutch (severity-based) threshold of €80,000/QALY. Probabilistic sensitivity analysis suggests that the implementation of NTRK-testing and treatment has a 0.2% chance of being cost-effective. In a scenario analysis that focuses only on the population of NTRK+ patients and excludes the costs and health effects associated with NTRK testing, the ICER drops steeply to €38,563/QALY. This illustrates that the consequences of neglecting the introduction of further testing as part of the decision problem may be substantial. It also suggests that the cost-effectiveness of TRK inhibitors may improve if genetic testing becomes more widespread in cancer care and the cost per genetic test decreases.

We present a single ICER in this study, implying an all-or-nothing decision regarding the reimbursement of entrectinib for NTRK+ patients. However, it might be more appropriate to differentiate between tumour types, for example because of heterogeneity in the (relative) treatment effect and incremental costs. That is, even when histology-independent therapies receive marketing authorisation for all histologies, reimbursement might be warranted only for a subset of histologies.

Chapter 6: Cost-effectiveness of alternative NTRK-testing strategies in three countries

In Chapter 5, the focus was on the Netherlands, and the modelled testing pathway was based on the testing protocol suggested in a consensus report by Dutch experts. In Chapter 6, the scope of the analysis is expanded to also include Hungary and the UK. Furthermore, the cost-effectiveness of four different testing strategies is compared. A “no testing” strategy is also considered.

All testing strategies include next generation sequencing panels used to screen tumour ribonucleic acid (RNA-NGS) for NTRK fusions and/or immunohistochemistry (IHC) tests to assess TRK protein expression. RNA-NGS tests can detect oncogenic NTRK fusions with high sensitivity and specificity. However, RNA-NGS tends to be expensive. IHC tests are generally more affordable than RNA-NGS tests, but also less accurate. The following testing strategies were included: (1) IHC test for all tumour types, (2) RNA-NGS test for

all tumour types, (3) IHC test followed by RNA-NGS in patients with a positive IHC test result for all tumour types, and (4) Stratified test strategies. Strategy 4 uses RNA-NGS tests (without preceding IHC tests) for patients with tumour types with high NTRK fusion prevalence (>90%) or wild-type TRK protein expression. All other patients in strategy 4 are subject to the sequential testing protocol used in strategy 3.

We applied the perspective, discount rate(s), and threshold value recommended by the respective HTA guidelines of England, Hungary, and the Netherlands. Country-specific values for the input parameters were used as much as possible. These included country-specific waiting times for the various stages of the testing phase and country-specific costs for biopsies, tests, entrectinib, and SoC pharmaceuticals.

The results show the same ranking of strategies for all countries, both in terms of ICERs and ΔNMB. The *No testing* strategy renders the lowest costs and lowest amount of QALYs, while *RNA-NGS for all* (Strat. 2) has the highest costs and QALYs. For all countries, the *IHC then RNA-NGS* (Strat. 3) option has the highest ΔNMB. The strategy results in less QALYs than *Stratified* (Strat. 4) and *RNA-NGS for all*, caused by a higher number of unidentified NTRK+ patients (i.e. those with a false-negative IHC test result) not receiving entrectinib treatment. Yet the issue of false negatives appears to be offset by the cost-savings from performing RNA-NGS for a smaller group of patients (i.e. only those who receive a positive IHC result first).

However, all estimated ICERs (int€ 89,196 for England, int€ 138,135 for Hungary, and int€ 142,663 for the Netherlands) are above national cost-effectiveness thresholds and all estimated ΔNMB values are negative. When focussing only on the NTRK+ population (i.e. assuming that no additional testing is needed to establish eligibility for entrectinib treatment), ΔNMB turns positive in the Netherlands and in the UK. This suggests that there is some potential for entrectinib to become cost-effective in the Netherlands and the UK. Nonetheless, even in a future with more widespread genetic testing, the assumption that no additional testing is needed to establish entrectinib eligibility may not hold. This is because RNA-based testing (as opposed to DNA-based testing) may be needed to identify functional NTRK gene fusions and/or because the genetic profile of tumours may change over time, meaning multiple tests are needed over time. In Hungary the ΔNMB of entrectinib remains (deeply) negative in the subgroup analysis of NTRK+ patients. This may mean that expensive targeted treatments like entrectinib bring limited value in lower-income countries, and more cost-effective healthcare may have to be prioritised first.

Chapter 7: General discussion

Chapter 7 provides a summary of the main findings. Based on Chapters 1 and 2, it is suggested that "personalised medicine" may be too broad a term and that using more modest, specific terms, such as "stratified medicine" and "genomic medicine" may lead to more realistic expectations regarding the value of innovative treatments. It is also explained how the different methodological recommendations of Chapter 3 are addressed in the economic model for NTRK testing and treatment that is described in Chapters 4-6.

Various recommendations are given for a (more) successful implementation of PM, as outlined below:

Stricter requirements ought to be set for clinical evidence submitted to marketing authorisation agencies and reimbursement agencies. There also need to be stricter requirements regarding follow-up studies to be conducted in cases where the original evidence is limited. Standardised procedures should be developed to ensure that clinical guidelines and reimbursement status are updated if follow-up studies show insufficient treatment effectiveness or severe side effects. More widespread collection of clinical and genomic registry data may be needed. National HTA guidelines may have to be updated with additional guidance on the modelling of testing pathways, to ensure testing pathways are consistently included and modelled across reimbursement dossiers. More research funding and policy measures are needed to ensure that the implementation of PM does not exacerbate existing health inequalities. The issue of unequal referral rates to genetic testing needs to be addressed, and patient populations both in clinical trials and in registry data need to be more representative of real-life patient populations. Measures are needed to ensure that areas of technological innovation better align with areas of high unmet clinical need. Measures are also needed to temper prices for (oncology) drugs with limited health benefit. And finally, policymakers, research funders, and researchers alike are encouraged to exercise caution around the topic of personalised medicine. Not all PM interventions offer added value to the healthcare system and society at large. So far, the field of PM has also had little attention for primary prevention, despite a huge potential for health gains through improved (targeting and tailoring of) primary prevention. Let us keep our eyes on the prize and only allocate resources to interventions and areas of innovation and implementation that are likely to offer (sizeable) added value.

Samenvatting

In dit proefschrift wordt een gezondheidseconomisch perspectief toegepast op het fenomeen *personalised medicine*, ofwel gepersonaliseerde geneeskunde.

Hoofdstuk 1: Algemene introductie

In hoofdstuk 1 werpen we een blik op de oorsprong en definitie van het concept 'gepersonaliseerde geneeskunde'. De term verscheen voor het eerst in 1999. De Wall Street Journal schreef toen over een onderzoeksproject dat tot doel had om gerichte geneesmiddelen te ontwikkelen "voor elk uniek genetisch profiel". In de loop van de tijd breidde de oorspronkelijke focus van gepersonaliseerde geneeskunde op geneesmiddelen en op genetische profielen zich uit. De term raakte verbonden met het meer algemene idee dat de gezondheidszorg (meer) moet worden toegespitst op onze individuele kenmerken. Tegelijkertijd ontstond er discussie over de exacte betekenis van 'gepersonaliseerde geneeskunde'. Filosofen De Grandis en Halgunset betoogden dat de term meer een verzameling van visies op de toekomst van de gezondheidszorg beschrijft dan een bestaande realiteit. Ze suggereren dat de betekenis van gepersonaliseerde geneeskunde daarom veranderlijk is en in de loop van de tijd aangepast kan worden, afhankelijk van welke technologieën worden gerealiseerd en politieke ontwikkelingen zoals 'wie erbij betrokken raakt', 'wie betaalt en waarvoor'. Op basis van een systematisch literatuuronderzoek naar de verschillende toepassingen van de term 'gepersonaliseerde geneeskunde' door Schleidgen et al., en om de tot nu toe beschikbare technologieën zo goed mogelijk weer te geven, nemen we voor dit proefschrift de volgende definitie van gepersonaliseerde geneeskunde aan:

> *een medisch model dat de therapeutische keuze baseert op het resultaat van gen-profilering of tot doel heeft pathogene gen-mutaties te corrigeren*

Verder worden in hoofdstuk 1 belangrijke concepten en methoden binnen de gezondheidseconomie uitgelegd. Gezondheidseconomie, en in het bijzonder het deelgebied van *health technology assessment* (HTA), wordt gebruikt om gezondheidstechnologieën te beoordelen. De eigenschappen en effecten van verschillende technologieën, ook wel 'interventies' genoemd, worden geëvalueerd, waarbij steeds dezelfde methodologische aanpak en dezelfde criteria worden gebruikt. Op deze manier wordt inzicht verkregen in de prijs-kwaliteitverhouding die verschillende zorginterventies ten opzichte van elkaar bieden. Hiermee kan besluitvorming over de verdeling van middelen ondersteund worden.

De belangrijkste doelstellingen van dit proefschrift zijn als volgt: het verschaffen van inzicht in de toegevoegde waarde van gepersonaliseerde geneeskunde, en het

onderzoeken van methodologische uitdagingen bij het evalueren van gepersonaliseerde geneeskunde. Als casestudie voeren we een kosteneffectiviteitsanalyse uit van een interventie die gerekend wordt tot de gepersonaliseerde geneeskunde.

Hoofdstuk 2: De netto baten van gepersonaliseerde geneeskunde

Het onderwerp 'gepersonaliseerde geneeskunde' (GG) is omgeven door hoopvolle verwachting. Een betere afstemming van de gezondheidszorg op ieder individu kan dan ook veel positieve gevolgen hebben – zoals een hogere effectiviteit van behandelingen, minder bijwerkingen van geneesmiddelen en verbeterde ziektepreventie –, wat vervolgens kan leiden tot een langere levensduur, hogere kwaliteit van leven, en/of een vermindering van onnodige uitgaven in de gezondheidszorg. Er zijn echter ook zorgen over de potentieel hoge kosten van (de implementatie van) GG. Aan sommige GG-interventies die de afgelopen jaren op de markt zijn gekomen, zijn hoge prijskaartjes verbonden. Ook een bredere implementatie van gen-testen in de klinische praktijk is mogelijk behoorlijk duur.

In hoofdstuk 2 wordt een onderzoek gepresenteerd naar de toegevoegde waarde, of de netto baten, van GG. We voerden een systematisch literatuuronderzoek uit naar economische evaluaties van GG-interventies. Voor alle paarsgewijze vergelijkingen tussen GG-interventies en niet-GG-interventies die werden geïdentificeerd, werden de incrementele kosten (Δkosten) en gezondheidseffecten (ΔQALY) vastgelegd en gebruikt om de incrementele netto monetaire baten (ΔNMB) te berekenen. Om te onderzoeken of bepaalde typen GG mogelijk meer voordeel bieden dan andere, voerden we regressieanalyses uit met Δkosten, ΔQALY en ΔNMB als afhankelijke variabelen en verschillende interventiekenmerken als onafhankelijke variabelen. *Random intercepts* werden gebruikt om de verschillende studies te clusteren op basis van land.

Van de 4.774 beoordeelde studies voldeden er 128 aan de inclusiecriteria, wat kosteneffectiviteitsresultaten opleverde voor 279 GG-interventies. De meeste studies betroffen de Verenigde Staten (48%) en het Verenigd Koninkrijk (16%). We vonden geen economische evaluaties van GG-interventies voor lage- of lagere-middeninkomenslanden. Kankerbehandelingen (60%) en farmaceutische interventies (72%) kwamen frequent voor. Ongeveer 32% van de studies bleek enige vorm van sponsoring van de farmaceutische industrie te hebben.

De mediane gezondheidswinst van GG-interventies bleek 0,03 *quality-adjusted life year* (QALY) per patiënt te bedragen, of ongeveer 11 (voor-kwaliteit-van-leven-gecorrigeerde) dagen. Gemiddeld genomen lijken de gezondheidswinsten van GG daarmee bescheiden. Ook tonen de resultaten aan dat de ΔNMB van GG-interventies rond nul ligt, met een

mediaan van 18 internationale dollar (Int$). Dit impliceert dat de gezondheidswinsten van GG-interventies doorgaans worden vergezeld door even grote kostenstijgingen.

Tegelijkertijd zijn er grote verschillen in de uitkomsten, met een ΔNMB zo hoog als Int$ 21.615 op het 95e percentiel en ΔNMB zo laag als Int$ –91.832 op het 5e percentiel. Uit de regressieanalyse komt naar voren dat gentherapieën een aparte subcategorie van GG vormen. Gentherapieën zijn geassocieerd met (veel) grotere gezondheidsvoordelen, hogere kosten en een lagere ΔNMB dan andere GG-interventies. Dit suggereert dat gentherapieën het potentieel hebben om grote gezondheidswinsten te bieden, maar bij de huidige prijsstelling doorgaans een nettoverlies voor het zorgsysteem opleveren. Het aantal gentherapieën in onze analyse was echter beperkt en de bevindingen zijn onzeker.

De regressieanalyse suggereert ook dat GG-interventies gericht op het *voorkomen van bijwerkingen* gemiddeld een hogere ΔNMB hebben. Anderzijds lijken GG-interventies die zich richten op de *identificatie (en behandeling) van patiënten die waarschijnlijk op de behandeling zullen reageren* doorgaans een lagere ΔNMB te hebben. Een mogelijke verklaring hiervoor is dat bij veel van de interventies in de eerste categorie sprake is van bestaande behandelingen (zonder actief patent), terwijl veel interventies in de laatste categorie patiënten stratificeren naar nieuw ontwikkelde behandelingen. Tot nu toe ligt de focus van GG op het ontwikkelen van nieuwe behandelingen voor nieuwe subgroepen van patiënten, maar verder onderzoek (en de implementatie ervan) om de *bestaande* zorg beter te stratificeren is mogelijk waardevol.

Tot slot, ondanks de algemene tendens van GG-interventies naar een ΔNMB van nul, hebben we ook verschillende interventies met (zeer) negatieve ΔNMB geïdentificeerd. Er kan nationaal en internationaal prijsbeleid nodig zijn om de kosten van deze interventies terug te dringen, om te voorkomen dat zorgsystemen een nettoverlies ondervinden bij de implementatie van de interventies.

Hoofdstuk 3: Leidraad voor de harmonisatie en verbetering van economische evaluaties van gepersonaliseerde geneeskunde

Sommigen menen dat gepersonaliseerde geneeskunde zodanig anders is dan andere vormen van geneeskunde, dat de bestaande HTA-methoden ontoereikend zijn voor het beoordelen van de kosteneffectiviteit van GG-interventies. Om hier meer inzicht in te krijgen, hebben we de methodologische uitdagingen die GG met zich mee kan brengen onderzocht. Hoofdstuk 3 biedt een overzicht van de mogelijke uitdagingen en aanbevelingen over hoe om te gaan met deze uitdagingen.

We voerden een pragmatisch literatuuronderzoek uit en hielden interviews met 18 experts voor een eerste overzicht van mogelijke methodologische uitdagingen en manieren om deze aan te pakken. Het systematische literatuuronderzoek naar

economische evaluaties van GG (hoofdstuk 2) werd ook gebruikt om te bekijken welke methodologische benaderingen tot nu toe zijn gebruikt. Nadat een eerste set aanbevelingen was opgesteld, zijn deze in een workshop met zo'n dertig stakeholders besproken en vervolgens aangepast. De uiteindelijke leidraad bevatte 23 aanbevelingen.

Enkele van de studies die we vonden stellen dat aanpassingen aan de standaard HTA-methoden nodig kunnen zijn voor GG-interventies. Veel van de deskundigen die we interviewden gaven echter aan dat GG onderworpen zou moeten zijn aan hetzelfde methodologische raamwerk als andere interventies, om consistentie en vergelijkbaarheid te garanderen. De eerste twee aanbevelingen van de leidraad weerspiegelen dit sentiment. Het gebruik van het standaard perspectief (Aanbeveling 1) en de standaard disconteringsvoet(en) (Aanbeveling 2) zoals aangegeven in de nationale HTA-richtlijnen wordt aangemoedigd, zowel voor GG- als voor niet-GG-interventies.

Het (genetisch) testen van patiënten en de daaropvolgende stratificatie naar verschillende behandelingen speelt een centrale rol bij GG. Toch gaan tot nu toe veel gezondheidseconomische modellen uit van het moment dat patiënten met de behandeling beginnen (d.w.z. nadat eventuele tests zijn afgerond). Om de (klinische) realiteit beter weer te geven, pleit de leidraad voor het opnemen van de testperiode in economische evaluaties van GG. Er worden acht aanbevelingen gedaan voor het accuraat modelleren van test-behandelingstrajecten.

Verschillende technologieën kunnen gebruikt worden voor genetische profilering en tests kunnen worden toegepast op verschillende momenten en in wisselende combinaties. Daarom wordt modelleurs aangeraden om niet slechts één van de beschikbare testopties overwegen. In plaats daarvan dienen alle relevante test-behandelingstrajecten te worden geïdentificeerd en moet worden onderbouwd waarom de in het model opgenomen opties zijn geselecteerd (Aanbeveling 3). We benadrukken verder dat er, om een zo goed mogelijk beeld te krijgen van de klinische realiteit, bij het bepalen van de kosten en gezondheidseffecten van testen rekening moet worden gehouden met *alle* patiënten die de tests ondergaan, inclusief patiënten die fout-positieve of (fout-) negatieve testresultaten krijgen (Aanbeveling 4). De implementatie van een test op erfelijke aandoeningen kan betekenen dat familieleden van de initiële patiënten met een positieve uitslag (de "indexpatiënten") ook in aanmerking komen voor de test. Bij een economische evaluatie voor indexpatiënten wordt daarom aangeraden om de kosten en gezondheidseffecten van het testen van eventuele familieleden ook mee te nemen (Aanbeveling 10). Andere onderwerpen die overwogen dienen te worden zijn: de relevantie van de beschikbare data over diagnostische nauwkeurigheid voor de betreffende patiëntenpopulatie (Aanbeveling 5); het effect van verschillende mogelijke drempelwaarden voor de testresultaten op de kosteneffectiviteitsuitkomsten

(Aanbeveling 6); mogelijke onderlinge samenhang tussen testresultaten wanneer meerdere tests worden gebruikt (Aanbeveling 7); morbiditeit/sterfte als gevolg van wachttijden in verband met de testprocedure (Aanbeveling 8); en mogelijke variatie in testkosten tussen laboratoria (Aanbeveling 9).

Hoe meer patiënten in specifieke subgroepen worden ingedeeld, hoe kleiner de steekproefomvang van elke patiëntenpopulatie wordt. De grote *randomised controlled trials* (RCT's) die de afgelopen decennia zijn gebruikt om de relatieve effectiviteit van interventies te meten, zijn daarom mogelijk niet altijd haalbaar bij GG. Steeds regelmatiger lijken aanvragen voor markttoetreding en vergoeding van geneesmiddelen te worden ingediend zonder RCT-bewijs. Echter, bij een deel van de niet-RCT-dossiers die tot nu toe zijn ingediend bij toezichthouders zoals het European Medicines Agency (EMA) en de Amerikaanse Food and Drug Administration (FDA) zijn er twijfels of het ontbreken van RCT-bewijs inderdaad gerechtvaardigd was. Ook zijn er zorgen over de gevolgen van het op de markt brengen van interventies op basis van beperkt klinisch bewijs, waaronder mogelijke gezondheidsschade voor patiënten en een suboptimale verdeling van beschikbare middelen. Wij moedigen daarom aan dat modelleurs effectiviteitsdata gebruiken uit studies met twee (of meer) alternatieve behandelstrategieën (Aanbeveling 11). Tegelijkertijd hebben degenen die de kosteneffectiviteit van interventies evalueren vaak geen controle over de beschikbare effectiviteitsdata. Daarom biedt de leidraad verschillende aanbevelingen om de kwaliteit van effectiviteitsschattingen zoveel mogelijk te waarborgen. Zo zijn er aanbevelingen die aandringen op transparantie over aannames en databronnen, zowel bij het gebruik van surrogaatuitkomsten om de uiteindelijke uitkomsten te schatten (Aanbeveling 12) als bij het schatten van het verband tussen genetische markers en klinische uitkomsten (Aanbeveling 15). Als een externe databron wordt gebruikt om de effectiviteit van andere mogelijke behandelingen (*comparators*) te schatten, moeten modelleurs rekening houden met mogelijke veranderingen in de effectiviteit van de comparatorbehandelingen sinds de studies ernaar plaatsvonden (Aanbeveling 13) en met de prevalentie van relevante genetische markers in de interventie- versus comparatorpopulaties (Aanbeveling 14). Verder biedt de leidraad aanbevelingen ter verbetering van de extrapolatie van de korte-termijngegevens uit klinische onderzoeken naar gezondheidsuitkomsten op de lange termijn. Modelleurs worden aangemoedigd om naast de statistische fit ook de mening van deskundigen te gebruiken bij het kiezen van een overlevingsmodel (Aanbeveling 16) en om rekening te houden met eventuele oversterfte (vergeleken met de algemene bevolking) onder langdurige overlevenden van een ziekte (Aanbeveling 17).

In 2018 werd een rapport uitgebracht door de *ISPOR Value Assessment Framework Special Task Force*, waarin werd betoogd dat de QALY onvoldoende de waarde van gezondheidsinterventies weerspiegelt. Het rapport noemde verschillende 'aanvullende

waarde-elementen', waaronder de waarde van weten (*value of knowing*), angst voor besmetting (*fear of contagion*), verzekeringswaarde (*insurance value*), waarde van hoop (*value of hope*), reële optiewaarde (*real option value*), ernst van de ziekte (*disease severity*), rechtvaardigheid (*equity*) en wetenschappelijke neveneffecten (*scientific spillovers*). Er is door sommigen betoogd dat zulke aanvullende waarde-elementen in het bijzonder relevant zijn voor GG. Er zijn echter verschillende kanttekeningen te plaatsen bij de voorgestelde lijst met waarde-elementen. Zo zijn veel van de waarde-elementen conceptueel onduidelijk en is er mogelijk overlap tussen de verschillende waarde-elementen. Ook is er een eenzijdige focus op positieve elementen van waarde, en wordt er voorbijgegaan aan het verschil tussen individuele voorkeuren enerzijds en het bepalen van maatschappelijke prioriteiten anderzijds. De leidraad ontmoedigt daarom het opnemen van waarde-elementen anders dan de elementen die worden aanbevolen in nationale HTA-richtlijnen. Mochten aanvullende waarde-elementen worden opgenomen in een gevoeligheidsanalyse, dan dienen mogelijke elementen met een negatieve waarde ook te worden overwogen (Aanbeveling 18).

De therapietrouw van patiënten en artsen in de klinische praktijk kan behoorlijk verschillen van de geoptimaliseerde omstandigheden in klinische trials. Dit geldt in het bijzonder als er ook genetische tests nodig zijn. Er kan bijvoorbeeld sprake zijn van wantrouwen bij patiënten tegenover genetische tests. Of de acceptatie onder artsen wordt belemmerd door de complexiteit van het interpreteren van testresultaten. De leidraad stelt daarom dat, als er inzicht nodig is in de kosteneffectiviteit van een interventie *onder realistische omstandigheden*, therapietrouw-parameters moeten worden opgenomen in de economische evaluatie (Aanbeveling 19). Uit eerder onderzoek is gebleken dat niet-witte patiënten en patiënten met een lagere sociaaleconomische status minder vaak worden doorverwezen voor genetisch onderzoek. We raden modelleurs daarom ook aan rekening te houden met mogelijke variatie in naleving afhankelijk van de maatschappelijke groep (Aanbeveling 20).

Grote onzekerheid wordt vaak genoemd als een belangrijke uitdaging bij GG. Zowel parameteronzekerheid (d.w.z. onzekerheid over hoe dicht de geschatte waarden voor modelparameters de daadwerkelijke waarden benaderen) als structurele onzekerheid (d.w.z. onzekerheid of de modelaannames en geselecteerde methoden geschikt zijn) doen zich voor. Parameteronzekerheid kan bij GG bijvoorbeeld worden veroorzaakt door de onduidelijke causale status tussen genetische markers en klinische uitkomsten. Structurele onzekerheid kan ontstaan vanwege complexe testprocedures met veel mogelijke test-behandelingsroutes, of omdat er tussen verschillende methodologische opties gekozen moet worden bij het schatten van uitkomsten op basis van beperkte data. De leidraad stelt dat wanneer de input van deskundigen wordt gebruikt om parameterwaarden te schatten, de onzekerheid in het deskundigenoordeel moet

worden weergegeven (Aanbeveling 21). Dit kan door de input van de deskundigen te synthetiseren in een waarschijnlijkheidsverdeling. Modelleurs worden ook aangemoedigd om de onzekerheden in structurele aannames en beslissingen te identificeren en de impact op de kosteneffectiviteitsresultaten te onderzoeken door middel van een gevoeligheidsanalyse, idealiter door de structurele aspecten te parametriseren (Aanbeveling 22).

Soms is de onzekerheid zodanig groot dat er geen besluit kan worden genomen over de vergoedingsstatus van een interventie (bijvoorbeeld wel of geen opname in het basispakket). In zulke gevallen kan er een overeenkomst voor voorwaardelijke toelating worden afgesloten tussen de leverancier en de zorgbetaler. Er bestaan verschillende soorten toelatingsovereenkomsten (bijvoorbeeld betaling op basis van resultaat, prijs-volumeafspraken etc.), waarin een breed scala aan voorwaarden kan worden opgenomen. Gezondheidseconomische modellen kunnen worden gebruikt om de voorwaarden van een toelatingsovereenkomst te optimaliseren. In de leidraad wordt daarom gesteld dat als voorwaardelijke toelating wordt overwogen voor een interventie, de toelatingsvoorwaarden kunnen worden opgenomen in het model waarmee de interventie wordt geëvalueerd (Aanbeveling 23).

Hoofdstuk 4: Prognostische waarde van de NTRK-fusie biomarker in Nederland

In hoofdstukken 4-6 voeren we een casestudie uit, waarbij we de methodologische leidraad uit hoofdstuk 3 toepassen op een voorbeeld van een GG-interventie: het testen op oncogene neurotrofe tyrosinereceptorkinase (NTRK) genfusies bij kankerpatiënten en vervolgens het geven van de tumor-agnostische behandeling entrectinib aan mensen met een positief testresultaat.

NTRK-fusies zijn het resultaat van chromosomale herschikkingen die een combinatie opleveren van het 3'-gebied van een NTRK-gen en het 5'-gebied van een fusiepartnergen. Dergelijke fusies kunnen leiden tot TRK-fusie-eiwitten met constitutieve kinase-activiteit, wat kan resulteren in tumorgroei. Entrectinib is een remmer van de drie meest voorkomende typen TRK-eiwit: TRKA, TRKB en TRKC (respectievelijk gecodeerd door het NTRK1-, NTRK2- en NTRK3-gen). Een gepoolde analyse van drie fase I/II-trials voor entrectinib liet positieve resultaten zien. Echter, de studies waren enkelarmig, waardoor de relatieve effectiviteit van entrectinib ten opzichte van andere behandelopties onduidelijk is. Om toch een schatting te kunnen maken, maken we in onze casestudie gebruik van effectiviteitsdata uit een externe databron. Een uitdaging hierbij is dat de onderzoekspopulatie in de entrectinib studies bestond uit patiënten met tumoren met NTRK-genfusies ("NTRK-positief"), terwijl de beschikbare externe data voornamelijk

patiënten zonder dergelijke tumoren ("NTRK-negatief") bevatten. Om voor dit verschil te kunnen corrigeren, hebben we de prognostische waarde van de NTRK-genfusie geschat.

We gebruikten de Hartwig Medical Foundation (HMF) database voor een steekproef van kankerpatiënten uit de Nederlandse klinische praktijk. De HMF-database bevat zowel genetische als klinische data. Een subset van de database, die werd verzameld via de Center for Personalised Cancer Treatment (CPCT-02) studie, bevat *whole-genome sequencing* (WGS)-data over het tumor-DNA van duizenden patiënten verspreid over 44 ziekenhuizen in Nederland. Voor onze analyse bekeken we bij de patiënten uit het CPCT-02-cohort de genetische markers in het tumor-DNA en een aantal klinische variabelen.

De patiënten werden ingedeeld in twee cohorten: NTRK-positief en NTRK-negatief. Patiënten met NTRK-fusies in het tumor-DNA die "waarschijnlijk functioneel" werden geacht (d.w.z. leidend tot de expressie van TRK-fusie-eiwitten) werden opgenomen in het NTRK-positieve cohort. Patiënten zonder NTRK-fusies in het tumor-DNA, of met NTRK-fusies die niet voldeden aan de voorwaarden voor waarschijnlijke functionaliteit, kwamen terecht in het NTRK-negatieve cohort. We voerden een *propensity score matching*-analyse uit om de groep van NTRK-positieve patiënten te koppelen een vergelijkbare subgroep van NTRK-negatieve patiënten. Hierbij werd een 1:4 *matching ratio* aangehouden. Binnen elk tumortype werden patiënten gematcht op de variabelen 'leeftijd', 'geslacht', 'jaar van biopsie' en 'aantal eerdere behandellijnen'. Vervolgens vergeleken we de overleving van NTRK-positieve en NTRK-negatieve patiënten met behulp van de Kaplan-Meier-methode.

Onder de 3.556 patiënten uit de CPCT-02-studie met bekende tumorlocatie waren 24 NTRK-positieve patiënten. Bij de NTRK-positieve patiënten komen negen verschillende tumortypen voor: bot/zacht weefsel, borst, colorectaal, hoofd en nek, long, pancreas, prostaat, huid en urinewegen. NTRK-fusies waarbij het NTRK3-gen (46%) en het NTRK1-gen (33%) betrokken waren, komen het meest voor.

In de overlevingsanalyse (*survival analysis*) vonden we een hazardratio (HR) van 1,44 (95% BI 0,81–2,55) voor NTRK-positieve patiënten. Het 95%-betrouwbaarheidsinterval bevat ook waarden kleiner dan 1, wat betekent dat de geschatte puntwaarde statistisch niet-significant is. Niettemin leverden drie eerdere onderzoeken (in net andere settings) vergelijkbare schattingen op, wat het aannemelijk maakt dat NTRK-positieve patiënten inderdaad een slechtere prognose hebben dan NTRK-negatieve patiënten. We voerden ook een gevoeligheidsanalyse uit waarbij NTRK-positieve patiënten werden uitgesloten als er naast een NTRK-genfusie nog andere oncogene genetische markers waren gevonden in het tumor-DNA. Deze analyse laat een lagere HR zien dan de hoofdanalyse (1,20; 95% BI 0,61–2,36). Dit suggereert dat de hoofdanalyse mogelijk de prognostische waarde van de NTRK-genfusie *an sich* overschat.

Een belangrijke beperking van het onderzoek is dat we één enkele HR-waarde hebben geschat voor alle NTRK-positieve patiënten, terwijl eerdere onderzoeken suggereren dat er mogelijk heterogeniteit bestaat tussen verschillende tumortypen wat betreft de prognostische waarde van NTRK-genfusies. We hadden weinig andere keus omdat er slechts 24 NTRK-positieve patiënten in de database bleken te zitten, maar toch is voorzichtigheid geboden bij het gebruik van de resultaten van dit onderzoek (en die van de drie eerdere onderzoeken).

Hoofdstuk 5: Kosteneffectiviteitsanalyse van de behandeling van patiënten met NTRK-positieve kanker met de tumor-agnostische therapie entrectinib

Entrectinib en collega-TRK-remmer larotrectinib kwamen op de markt als de eerste twee moleculair gerichte therapieën met een tumor-agnostisch label (ook wel “pan-tumor indicatie” genoemd). Tumor-agnostische therapieën kunnen worden voorgeschreven ongeacht de histologie van de tumor, een benadering die niet eerder werd gebruikt bij de kankergeneesmiddelen. In hoofdstuk 5 presenteren we een kosteneffectiviteitsanalyse van de behandeling van NTRK-positieve patiënten met entrectinib. De interventie is als volgt gedefinieerd: NTRK-genfusietesten voor alle patiënten met lokaal gevorderde of gemetastaseerde solide tumoren, gevolgd door behandeling met entrectinib bij NTRK-positieve patiënten en standaardbehandeling (SB) bij NTRK-negatieve patiënten. In de vergelijkingsarm krijgen alle patiënten SB en wordt er geen NTRK-test uitgevoerd. Larotrectinib zou idealiter als een comparator zijn opgenomen. Er zijn echter voor zowel entrectinib als larotrectinib alleen enkelarmige onderzoeken uitgevoerd, wat het combineren van databronnen bemoeilijkt. De verschillen tussen de onderzoekspopulaties van de larotrectinib studies en de entrectinib studies zijn zodanig groot dat larotrectinib niet kon worden opgenomen in de economische evaluatie van entrectinib.

De introductie van entrectinib in de klinische praktijk vereist de introductie (/uitbreiding) van NTRK-testen om geschikte patiënten te identificeren. Daarom hebben we de testfase in het model opgenomen, weergegeven met behulp van een beslisboom. In de beslisboom zitten verschillende kansknooppunten, waaronder de waarschijnlijkheid dat (her)biopten nodig zijn en de waarschijnlijkheid dat patiënten overlijden tijdens de testperiode. Patiënten die de testperiode overleefden, komen vervolgens in een patiëntsimulatiemodel terecht. Het simulatiemodel omvat de gezondheidstoestanden ‘levend en onder behandeling, ‘levend en niet onder behandeling’ en ‘dood’. Patiënten krijgen entrectinib als ze NTRK-positief zijn getest in de beslisboom en worden behandeld met SB als ze NTRK-negatief zijn getest. In de vergelijkingsarm worden de patiënten

niet getest (d.w.z. ze slaan de beslissingsboom over en gaan rechtstreeks naar het patiëntsimulatiemodel) en krijgen allemaal SB.

Uit de entrectinib-trials werden data verkregen over de effectiviteit van entrectinib bij NTRK-positieve patiënten. Het betreft data over de totale overleving, ofwel *overall survival* (OS), en de tijd tot stopzetting van de behandeling, ofwel *time to treatment discontinution* (TTD). De steekproefomvang in de entrectinib-trials was klein. Door het daaruit voortvloeiende gebrek aan betrouwbare *tumor-specifieke* schattingen, hebben we moeten aannemen dat de effectiviteit van entrectinib constant is voor alle tumortypen. Voor de vergelijkingsarm, SB, konden we wel de tumor-specifieke effectiviteit schatten. We gebruikten de HMF-database om OS en TTD te schatten voor NTRK-negatieve patiënten die SB ontvangen. Om vervolgens de uitkomsten te berekenen voor *NTRK-positieve* patiënten die SB krijgen, werd de geschatte HR van 1,44 (hoofdstuk 4) toegepast op de OS- en TTD-schattingen voor NTRK-negatieve patiënten. Verder verzamelden we voor zowel de relevante testen als voor de SB-behandelingen tumor-specifieke kostenschattingen. Ten slotte hebben we, naar aanleiding van eerder onderzoek dat toonde dat de kwaliteit van leven afneemt naarmate patiënten de dood naderen, patiëntutiliteit (*patient utility*) in ons model opgenomen als een functie van de tijd tot aan overlijden. Hierbij hebben we gebruik gemaakt van geslacht- en leeftijd-specifieke schattingen van patiëntutiliteit.

De resultaten laten zien dat het testen op NTRK-fusies bij patiënten met gevorderde kanker en vervolgens behandelen met entrectinib bij NTRK-positieve patiënten en SB bij NTRK-negatieve patiënten geassocieerd is met een gezondheidswinst van 0,0043 QALY en een kostenstijging van €732 per patiënt ten opzichte van de vergelijkingsarm (geen testen en SB voor alle patiënten). De incrementele kosteneffectiviteitsratio (ICER) bedraagt €169.957/QALY, ver boven de Nederlandse (op ziektelast gebaseerde) drempelwaarde van €80.000/QALY. Probabilistische gevoeligheidsanalyse liet zien dat de implementatie van NTRK-testen en -behandeling een kans van 0,2% heeft om kosteneffectief te zijn. In een scenarioanalyse die zich uitsluitend richtte op NTRK-positieve patiënten en waarin de kosten en gezondheidseffecten geassocieerd met NTRK-testen niet werden meegenomen, daalt de ICER naar € 38.563/QALY. Dit laat zien dat het weglaten van de testfase uit het economisch model, waarmee een onvolledig beeld wordt geschetst van het beslisprobleem, grote gevolgen kan hebben voor de geschatte kosteneffectiviteit. Het suggereert ook dat de kosteneffectiviteit van TRK-remmers kan verbeteren als genetische tests meer wijdverspreid worden in de kankerzorg en de kosten per genetische test afnemen.

We presenteren in deze studie één enkele ICER, wat een alles-of-niets-beslissing impliceert wat betreft de vergoeding van entrectinib voor NTRK-positieve patiënten.

Mogelijk is het echter passender om onderscheid te maken tussen tumortypen, bijvoorbeeld vanwege heterogeniteit in het (relatieve) behandeleffect en/of de incrementele kosten. Dit kan betekenen dat hoewel markttoetreding is toegestaan voor alle tumortypen, vergoeding door de zorgverzekeraar/zorgbetaler niet voor alle tumortypen gerechtvaardigd kan worden.

Hoofdstuk 6: De kosteneffectiviteit van verschillende NTRK-teststrategieën in drie landen

In hoofdstuk 5 ligt de focus op Nederland en is het gemodelleerde testtraject gebaseerd op het testprotocol dat werd voorgesteld in een consensusrapport van Nederlandse experts. In hoofdstuk 6 wordt de reikwijdte van de analyse uitgebreid naar Hongarije en Groot-Brittannië. Bovendien wordt de kosteneffectiviteit van vier verschillende teststrategieën vergeleken. Ook wordt er een 'geen testen'-strategie overwogen.

Alle teststrategieën bevatten RNA-gebaseerde *next-generation sequencing* paneltesten om te screenen op NTRK-fusies (RNA-NGS) en/of immunohistochemische (IHC)-tests om de expressie van TRK-eiwitten te beoordelen. RNA-NGS-testen hebben een hoge sensitiviteit en specificiteit. RNA-NGS is echter vaak duur. IHC-tests zijn over het algemeen goedkoper dan RNA-NGS-tests, maar ook minder nauwkeurig. De volgende teststrategieën zijn opgenomen: (1) IHC-test voor alle tumortypen, (2) RNA-NGS-test voor alle tumortypen, (3) IHC-test gevolgd door RNA-NGS bij patiënten met een positief IHC-testresultaat, voor alle tumortypen, en (4) gestratificeerde teststrategie. In strategie 4 ontvangen patiënten met tumortypen met een hoge NTRK-fusieprevalentie (>90%) of wildtype TRK-eiwitexpressie rechtstreeks een RNA-NGS-test (zonder voorafgaande IHC-test). Alle andere patiënten vallen onder het testprotocol dat ook in strategie 3 wordt gebruikt.

We hebben het perspectief, de disconteringsvoet(en) en de drempelwaarde toegepast die worden aanbevolen door de respectieve HTA-richtlijnen van Engeland, Hongarije en Nederland. Voor de modelparameters zijn zoveel mogelijk land-specifieke invoerwaarden gebruikt. Zo gebruikten we land-specifieke wachttijden voor de verschillende fasen van de testperiode en land-specifieke kosten voor biopten, tests, entrectinib en SB-geneesmiddelen.

De resultaten laten voor alle landen dezelfde ranking van strategieën zien, zowel wat betreft ICER's als ΔNMB. De *geen test*-strategie levert de laagste kosten en het laagste aantal QALY's per patiënt op, terwijl *RNA-NGS voor iedereen* (strat. 2) voor de hoogste kosten en grootste QALY-winst zorgt. In alle drie de landen heeft de *IHC en vervolgens RNA-NGS*-optie (strat. 3) de hoogste ΔNMB. Deze strategie resulteert desondanks in minder QALY's dan strategieën *gestratificeerd* (strat. 4) en *RNA-NGS voor iedereen*.

Dit wordt veroorzaakt door een groter aantal niet-geïdentificeerde NTRK-positieve patiënten (d.w.z. degenen met een fout-negatief IHC-testresultaat) die daardoor geen behandeling met entrectinib krijgen. Het hogere aantal fout-negatieven lijkt te worden gecompenseerd door kostenbesparingen doordat de dure RNA-NGS-testen voor een kleinere groep patiënten worden toegepast (d.w.z. alleen degenen die eerst een positief IHC-resultaat ontvangen).

Echter, alle geschatte ICER's (int€ 89.196 voor Engeland, int€ 138.135 voor Hongarije en int€ 142.663 voor Nederland) liggen boven de nationale drempelwaardes voor kosteneffectiviteit en alle geschatte ΔNMB-waarden zijn negatief. In een scenarioanalyse waarbij we ons uitsluitend richtten op de NTRK-positieve populatie (d.w.z. dat we aannamen dat er geen aanvullende tests nodig zijn om vast te stellen of men in aanmerking komt voor een behandeling met entrectinib), vinden we positieve ΔNMB-waarden in Nederland en Engeland. Dit suggereert dat entrectinib potentie heeft om kosteneffectief te worden in Nederland en Engeland. De aanname dat er geen aanvullende testen nodig zijn om in aanmerking te kunnen komen voor entrectinib zal in de praktijk echter niet vaak kloppen, zelfs in een toekomst waarin genetische testen meer gemeengoed zijn. Dit komt o.a. doordat voor het identificeren van functionele NTRK-genfusies doorgaans RNA-gebaseerde testen nodig zijn, terwijl voor andere genetische markers DNA-gebaseerde testen geschikter zijn. Ook kan het genetische profiel van tumoren in de loop van de tijd veranderen, waardoor er meer dan eens getest moet worden. In Hongarije blijft de ΔNMB van entrectinib (extreem) negatief in de scenarioanalyse met enkel NTRK-positieve patiënten. Dit betekent mogelijk dat dure precisiemiddelen zoals entrectinib beperkte waarde hebben in lagere-inkomenslanden en dat de implementatie van meer kosteneffectieve gezondheidszorg eerst prioriteit moet krijgen.

Hoofdstuk 7: Algemene discussie

Hoofdstuk 7 bevat een samenvatting van de belangrijkste bevindingen. Op basis van de hoofdstukken 1 en 2 wordt benoemd dat 'gepersonaliseerde geneeskunde' mogelijk een te brede term is en dat het gebruik van meer bescheiden, specifieke termen zoals 'gestratificeerde geneeskunde' kan bijdragen aan realistischere verwachtingen over de waarde van innovatieve behandelingen. Verder wordt in hoofdstuk 7 toegelicht hoe de verschillende aanbevelingen uit hoofdstuk 3 zijn toegepast in het economische model voor NTRK-testen en -behandeling dat in hoofdstukken 4-6 wordt beschreven.

Er worden in hoofdstuk 7 ook verschillende aanbevelingen gedaan voor een succesvolle(re) implementatie van GG, die luiden als volgt:

Er zijn strengere eisen nodig voor het klinisch bewijsmateriaal dat wordt ingediend bij de instanties die gaan over markttoetreding en vergoedingsstatus. Er moeten ook strengere eisen komen rondom vervolgstudies voor behandelingen waarbij het oorspronkelijke bewijsmateriaal beperkt is. Er moeten gestandaardiseerde procedures worden ontwikkeld om ervoor te zorgen dat klinische richtlijnen en vergoedingsstatus worden herzien als vervolgstudies aantonen dat de behandeling onvoldoende effectief is of ernstige bijwerkingen veroorzaakt. Mogelijk is het waardevol om op grotere schaal klinische en genetische data te verzamelen. Nationale HTA-richtlijnen kunnen worden bijgewerkt om meer richting te geven rondom het modelleren van testtrajecten, zodat testtrajecten consistent worden meegenomen en gemodelleerd in vergoedingsdossiers. Meer onderzoeksfinanciering en extra beleidsmaatregelen zijn nodig om te voorkomen dat de implementatie van GG bestaande gezondheidsverschillen vergroot. Het probleem van ongelijke doorverwijzing voor genetische tests moet worden aangepakt, en patiëntenpopulaties in zowel klinische trials als in registers moeten representatiever worden. Er zijn maatregelen nodig om ervoor te zorgen dat de ziektegebieden waarin technologische innovatie plaatsvindt beter aansluiten bij de gebieden met een grote onvervulde klinische behoefte (*unmet clinical need*). Er zijn ook maatregelen nodig om de prijzen van (oncologische) geneesmiddelen die tot beperkte gezondheidswinst leiden te temperen. En ten slotte worden beleidsmakers, onderzoeksfinanciers en onderzoekers aangemoedigd om voorzichtig te zijn rond het onderwerp ‘gepersonaliseerde geneeskunde’. Niet alle GG-interventies bieden toegevoegde waarde voor het zorgsysteem en/of de samenleving als geheel. Ook heeft het GG-veld tot nu toe weinig aandacht gehad voor primaire preventie, terwijl mogelijk grote gezondheidswinsten behaald kunnen worden door het beter afstemmen van preventiemaatregelen op het individu. Laten we het doel niet uit het oog verliezen en alleen middelen toebedelen aan de interventies en de innovatie- en implementatieactiviteiten die (grote) toegevoegde waarde kunnen bieden.

Acknowledgements

Bringing about this PhD thesis was a challenge and I am very proud of myself for seeing it through. But it was not a solo effort, and I am incredibly grateful to all the people who provided help, support, or just light-hearted distractions along the way. I feel blessed.

Allereerst hulde aan de mensen met wie ik het HEcoPerMed team vormde en die later, toen ik op de valreep besloot er een PhD van te maken, mijn promotor en copromotoren werden. Maureen, Matthijs en Simone. Een betrokken team dat elkaar goed aanvulde. Met zijn vieren vormden we een mooie regenboog op Jungs kleurencirkel. Maureen, dank voor je betrokkenheid. Als ik iets naar jou opstuurde, wist ik dat je ieder woord met aandacht zou lezen. Ik heb respect voor jouw werklust en je gretigheid om te blijven leren. Geen hoogleraar in een ivoren toren, maar (waar nodig) met de voeten in de klei der gezondheidseconomische evaluaties. Matthijs, tijdens overleggen altijd vol met creatieve ingevingen en slimme ideetjes. Je hebt meerdere gemiste carrières, als stand-up comedian, muzikant, klusser, automonteur, psycholoog en coach; ik bewonder je veelzijdigheid. Dank ook dat je me in bescherming nam. Jij kon vaak beter dan ik inschatten wat realistisch was en floot me terug als ik dreigde te veel op me te nemen. Simone, ik heb je als een waanzinnig fijne collega en begeleider ervaren. *10/10 would recommend*. Jouw rust en overzicht en ijzersterke planningsvaardigheden vormden een goede tegenhanger voor mijn soms wat chaotische (en altijd te tijd-optimistische) brein. Het was fijn om een maatje te hebben om na te praten over consortium-overleggen, expertinterviews, en allerhande frappante voorvallen. Tijdens de coronaperiode zette je deze gewoonte voort door mij na afloop van overleggen te bellen via Teams; tekenend voor jouw attentheid en zorgzaamheid. Many thanks also to Isaac, who helped with the regression analysis for the study on the net benefit of personalised medicine, and to Irene, without whom this PhD thesis might have never existed. When I had finished four papers and was ready to move on to something else altogether, Irene offered to work on a fifth paper together so that I would have enough papers for a PhD thesis. Thank you, Irene, I am grateful.

Another thank you goes out to the wider HEcoPerMed consortium. I was usually the youngest in the room but felt respected and listened to. I very clearly remember our kick-off meeting, in a grey, winterish Brussels. Just weeks prior, I had arrived back in the Netherlands after spending over two years in Ghana (as you could tell by the *kente* jacket I rocked up in). I had gotten used to the animated, expressive style of communicating at Ghana's Ministry of Health and experienced somewhat of a culture shock at the mellow, almost reserved atmosphere at our first meeting. But over the next years our bonds flourished. Partly aided by the in-person meet-ups in Oxford (where the health economists perplexed everyone else with their technical discussions) and

Vienna (where the political scientists bewildered the health economists with their vastly different research methods). Then, 3.5 years after the start of the project, everything came full circle and we returned to Brussels for the final conference, which I can report was lively and colourful. Thank you for the collaboration, Balázs, László, and Tamás (Syreon Research Institute), Apostolos, Rositsa, and Sarah (University of Oxford), Maren and Wolfgang (DLR), and Doris, Manuela, and Susanne (Austrian Institute of Technology).

Beste leden van de beoordelingscommissie, prof.dr. Patrick Bossuyt, prof.dr. Valesca Retel, en prof.dr. Ron van Schaik, heel veel dank dat jullie het nog relatief onopgemaakte PhD manuscript hebben willen doorakkeren. Ik waardeer het zeer. Beste leden van de promotiecommissie, prof.dr. Pieter van Baal, prof.dr. Marcel Dijkgraaf, prof.dr. Maarten IJzerman, prof.dr. Valesca Retel, dr. Lonneke Timmers, en dr. Hans Westgeest, ook aan jullie veel dank. Ik voel me vereerd dat jullie tijd hebben vrijgemaakt in jullie ongetwijfeld drukke schema's om met mij in gesprek te gaan over mijn proefschrift.

Hoewel de coronaperiode al na slechts een jaar in dienst losbarstte, heb ik gelukkig toch ook af en toe in levenden lijve kunnen genieten van mijn lieve en behulpzame collega's bij het institute for Medical Technology Assessment. Dank aan mijn gezellige kamergenootjes, Martine en (nadat de muur met de aangrenzende kamer werd doorgebroken) Philip en Kathi. Speciale dank aan Gimon, de projectleider van het eerste project dat ik draaide. Voor een onderzoek naar de implementatie van voorspellende testen in de Nederlandse zorg reisden we af naar verschillende steden om klinisch genetici, oncologen en andere experts te interviewen. We bleken het, met een gedeelde liefde voor muziek en een gedeelde aversie tegen de vroege ochtend, al snel goed te kunnen vinden. Door jouw kritische vragen kreeg ik vaak mijn gedachten beter op een rijtje. Ook kon ik altijd bij jou terecht voor een flinke portie scepsis over mooie praatjes door partijen met financiële belangen, wat soms erg troostrijk was. Bij je eigen promoveren vond ik je opvallend ontspannen, dus ik heb goede hoop dat het met zo'n koele kikker aan mijn zijde ook bij mij wel goedkomt.

Natuurlijk ook dank aan de collega's van de Erasmus School of Health Policy & Management. Extra dank voor Pieter Bakx, die ik in 2016 al leerde kennen toen ik op bezoek was vanuit Engeland en die mij met zijn betrokkenheid en gedetailleerde, waardevolle feedback op mijn masterscriptie wist te enthousiasmeren voor de Rotterdamse tak van de gezondheidseconomie. Ook bijzondere dank voor Pieter van Baal, die mij vaak wist te vermaken met wederom een uitgesproken mening en met wie ik, zeker richting het einde van de werkdag, soms veel te lang kon kletsen over muziek. En dan natuurlijk nog de lieve Sanna, die altijd mijn Daily Paper drip wist te waarderen. Ik waardeer je openheid, je eigengereidheid, en natuurlijk je kennis over

een rijke verzameling aan random onderwerpen. En ik hoop over een tijdje het hardst te kunnen klappen na jouw eigen “hora est”.

De Introductie en Discussie van dit proefschrift schreef ik toen ik inmiddels op de personeelslijst stond bij Pharos. Een organisatie met een groot sociaal hart en zoveel gezellige en warme collega’s dat het er teveel zijn om op te noemen. Ik moest even mijn draai vinden, maar ben dankbaar voor het vertrouwen en de vrijheid die ik krijg en ben trots dat ik mijn eigen kleine steentje kan bijdragen aan het verkleinen van gezondheidsverschillen.

Enige tijd geleden kwam ik een tweet tegen van scheikundige Dr. Zoë Ayres die zei:

> *The “gifted and talented” to “burnout, anxious and depressed” to “adult ADHD diagnosis” is real huh?*

Voor mij een herkenbare reis, die ik aflegde tijdens het tot stand komen van dit proefschrift. Onderweg kwam ik verschillende zorgverleners tegen. Ik heb me door allen heel erg gezien en geholpen gevoeld en ben eindeloos dankbaar. Mijn dank gaat in het bijzonder uit naar Hayat, een warme persoonlijkheid die mij hielp om van een donkere plek te helen naar een (nog) sprankelendere versie van mezelf. Ook speciale dank voor Ilse, die de tekenen van ADHD wist te herkennen (een diagnose die bij meisjes en vrouwen vaak wordt gemist).

Mijn lieve mam en pap. Allebei enigszins eigenaardige figuren met weinig egard voor “hoe het hoort”. Allebei goeierds, die het beste voor hebben met anderen. Soms voel ik me richting jullie zoals het wellicht ook voelt om volwassen kinderen te hebben; op een afstandje vergenoegd toekijkend hoe jullie levens zich ontvouwen. Mam, net klaar met de opleiding Social Work en met 60 jaar oud aan het begin van een splinternieuwe loopbaan als maatschappelijk werker. Wat een prestatie, ik ben onwijs trots op je. Pap, net met vervroegd (semi-)pensioen en eindelijk tijd voor struintochten door de vele interessante onderwerpen die deze wereld te bieden heeft. Na het in kaart brengen van de wetenschappelijke kennis op het gebied van zwaartekracht ben je via zelfstudie naar de eigenschappen van geluid nu aanbeland in de wondere wereld van de audiofielen. Ik ben benieuwd naar je volgende passieproject. Mam, ik ben dankbaar dat je mij hebt geleerd open en begripvol te zijn. Het heeft me vele mooie interacties en connecties gebracht, misschien legde het zelfs de basis voor de intercontinentale liefde die ik vond. Pap, mijn favoriete hulplijn bij de vraagstukken des volwassen levens. Klushulp, financieel advies, feedback op stukken die ik schrijf – je bent van alle markten thuis. Oneindig veel dank aan jullie allebei voor alles dat jullie voor ons gedaan (en gelaten) hebben. Uiteraard ook eervolle vermeldingen voor Bart, een geduldige man die houdt van grapjes,

grollen en verrassingen, en Marlies, een gezelschapsdier met een uitstekend ontwikkeld vermogen tot moederlijke “maar hoe gaat het nou echt met je?”-ondervragingen. Jullie verschenen op het toneel toen ik reeds een ongeleid projectiel van een jongvolwassene was, maar ik waardeer jullie en alles wat jullie toevoegen aan onze blended family. Ook noem ik graag nog even mijn lieve en veerkrachtige opa Niek en oma Rini.

Mijn allerliefste siblings. Allereerst de zorgzame Diane, met een zwak voor de mens die net even anders is. Goed met woorden, maar neemt het liefst de tijd om er even over na te denken. Open-minded als geen ander. Enigszins ongrijpbare zwerfkat-redder en tegelijkertijd ook een chique corporate dame. Dan Joris, een grote lieverd die slecht kan liegen en goed anderen kan aanvoelen. Een rustige vent met een speciaal talent voor een goed getimed grapje. Beheerder van de door familie en vrienden alom geprezen Pittig Prima-playlist, voor uw dagelijkse dosis groovy hitjes. En Nanette, een verbinder met een ontspannen aanpak van het leven. Een fashionista die creatieve verjaardagskaarten tekent, handycam compilaties maakt van alle vakanties, en die heeft afgedwongen dat we met Sinterklaas nog steeds surprises knutselen (alleen in schrikkeljaren dan, dat bleek na uitgebreid polderen het maximaal haalbare). Ik ben gek op jullie.

As those who know my private persona are aware, I am spectacularly bad at replying to text messages. I seem to have ended up with a bunch of friends who are either very patient or also not great texters. Hence many of you I don’t speak to all that frequently. But it doesn’t matter really. I am grateful to have such a great, eclectic circle of people to hang out with. Here are your shoutouts, roughly in chronological order of appearance in my life.

Kiki, van schoolpleintaferelen naar puberperikelen, naar twijfels over studiekeuze en baankeuze, naar allebei een eerste hypotheek op onze naam. Twee chronisch-te-laat-komers met een voorliefde voor interessante weetjes, het was altijd al een puike match. Ik hou van jouw hart-op-de-tong eerlijkheid en hoop dat ik nog vele jaren van je creatieve gedachtekronkels kan genieten.

Anne, een levensgenieter met een nuchtere inborst en een immer opgeruimd (en stijlvol!) huis. Ook jou ken ik al sinds de tijd dat vele van onze eerste keren nog moesten komen. Met je speciale talent voor luisteren en doorvragen heb je mij al meer dan eens zien huilen, maar ik hoop toch vooral nog heel lang met je te blijven lachen.

Annemiek, attent, gul en loyaal, jij koestert je vrienden. Stoer ben je ook; in de mannenwereld van de muziek schroom jij niet te uit te spreken als er iets niet in de haak is. Al sinds onze tijd bij het NHJO kan ik over jouw schouder meekijken hoe jij je muziekcarrière aanpakt: heel gestructureerd, met veel discipline maar ook ruimte voor rust en een rotsvast geloof in jezelf. Ik hoop het een en ander te gaan kopiëren!

Eleanor, the energy of a happy butterfly with the voice of a fancy British lady (though more recently with a touch of Australian). Always sharp with the social and political analysis and quite the DIY creative. Unfortunately you appear to have strong preference for places that are a full day of travelling away from England (/the Netherlands) but I appreciate you girlie.

Katie, my partner in crime during the undergrad days. I love your dry – bordering on rude – sense of humour. I love your creative thought processes. I love your knack for planning fun activities. I even love your dedication to unhuman levels of physical exercise and shall forever happily rollerblade along as you go running.

Chiara, the gentle voice of reason whenever I say something flippant. I admire your approach to life: full of love and understanding and always ready to have some fun. With you I can talk for hours on end. I am so grateful that years after our joint initial steps in the world of health economics, both in York and Aberdeen, you will be right by my side as I become Dr Vellekoop.

Evelien, iemand die met een zekere lichtheid door het leven beweegt. Een beetje lief zijn voor elkaar, beetje goed voor jezelf zorgen, beetje genieten; zo ingewikkeld is het niet. Een scherpe geest en een warme persoonlijkheid, altijd in voor een grapje en met hart voor de goede zaak. De wereld kan nog wel een paar Eveliens gebruiken.

Lauren, we met in one of the washrooms at Ghana's Ministry of Health, on a day the toilets weren't flushing. In hindsight I think this is kind of fitting. While the average *obroni* lives in a bubble of luxury and comfort with fellow internationals (speaking of "poor integration" huh, but that's a conversation for another time), you have been living like a Ghanaian in so many ways. Although I sometimes wish you would work a little less hard, I know you love what you do, and I am so impressed with your courage and perseverance.

Sekai, a sparkly soul with a colourful wardrobe and the best collections of Instagram Stories. Rarely do I meet people as dedicated to truth-telling, calling out what needs to be called out. I know it can be lonely at times, but you are so loved, in this realm and beyond.

Smood, Sally, Remas and Tahreer, family of fairies. Nadat ik jullie drie jaar geleden als taal- en huiswerkmaatje leerde kennen werd ik al snel gepromoveerd tot "eerstgeboren dochter". Wat een topvrouwen zijn jullie. Slim en grappig, eigenwijs en stoer. Smood, zo liefdevol en tegelijkertijd voor niemand bang. Als ik moeder mag worden, hoop ik een beetje zoals jij te zijn. Sally, een social butterfly met een grote liefde voor hiphop. Remas, een gamer girlie met een fluwelen stem. En Tahreer, een creatieveling (met verbazingwekkend Rotterdamse tongval!) die overal wel een spelletje of een kunstwerk van weet te maken. Ik heb jullie in mijn hart gesloten.

Indra, je inspireert me meid. Onverschrokken spring je keer op keer in het diepe. Van verandermanagement naar evenementenorganisatie, van Connectiediners naar harptherapeut. Vol lef en doortastendheid stap jij op je doelen af. Tegelijkertijd heb je ook veel rust, sta je in contact met je omgeving. Perfect voor een muziektherapeut en een mooi voorbeeld voor Lilou.

Mark, wie had gedacht dat ik bij de dansles pardoes een gezondheidseconoom in het wild zou tegenkomen, en nog een gezellige ook! Ik bewonder hoe jij met je vrienden en familie omgaat. Je brengt mensen bij elkaar, creëert gemeenschap, en hebt altijd een smeuïge anekdote om te delen. Ik waardeer je herhaaldelijke *reality checks* over mijn te harde werken ook zeer en ik haal jouw "wat denk je nou zelf dat je aan het doen bent"-gezicht tegenwoordig weleens voor ogen als ik dreig weer in overdrive te gaan.

Magomed, maatschappelijk betrokken en een open boek. Enkele jaren geleden stuurde ik jou nog een lijstje met tips voorafgaand aan je eerste reis naar Ghana. Inmiddels loop je regelmatig door Accra's straten en wel met een lieve Ghanese vriendin aan je zijde. Ik kan me weinig betere plekken voorstellen voor, met afstand, de grootste *afrobeat*- en *amapiano*-fanaat die ik ken.

Lyasmin, een levendige vrouw met een hart van goud. We begonnen als collega's-op-afstand, maar groeiden met onze gedeelde activistische inborst al snel naar elkaar toe. Ik geniet van onze gesprekken en ik bewonder hoe jij je eigen koers vaart. Ik zou het voor mezelf jammer vinden als je Rotterdam weer verlaat, maar ik wens vooral dat jouw dromen waarheid worden. Je verdient het.

A peculiar pattern that I first noticed when I joined a symphony orchestra as a teenager is that the groups I tend to feel most comfortable in are groups of musicians. My dear Jazz & Pop classmates at the Utrecht Conservatory; despite being one of the older students, I feel at home with you. I am proud of us and excited for all of our journeys still ahead. Ook liefde voor mijn maten van rosetta.beats, en in het bijzonder de Uutje crew. Wat een verzameling lieve en indrukwekkende vrouwen.

Of course there are more family members and friends whom I haven't mentioned by name but whom I hold dear. I appreciate you!

Then finally, the coda of this love song:

Peter, me dɔ. Your presence alone calms my nervous system all the way down. You inspire me to believe in myself and to trust my intuition. I am in awe of you after all these years still. I admire your perceptiveness and your ability to find fun and lightness in any a situation. Thank you for letting me bask in your love, you mean the world to me.

About the author

Heleen Vellekoop holds a BA in Philosophy, Politics & Economics and an MSc in Health Economics, both from the University of York. In between her bachelor's and master's degree she worked as a research intern at the Health Economics Research Unit of the University of Aberdeen. She was involved in a project measuring patient preferences so as to better tailor the support patients receive for the self-management of chronic conditions. Heleen wrote her master's dissertation – about the equity effects of different financing policies for long-term elderly care – during a 3-month placement at the Erasmus School of Health Policy & Management. From 2016 until the end of 2018 Heleen worked at Ghana's Ministry of Health, through the ODI Fellowship Scheme. She was involved in policy-related research, as well as monitoring and evaluation activities. A key achievement was the delivery of an economic evaluation of the benefits package of Ghana's National Health Insurance Scheme, which outlined possible ways to improve the benefits package design. At the start of 2019, Heleen signed on with the institute for Medical Technology Assessment. A large share of her time was spent on the EU-funded project *Health Economics for Personalised Medicine (HEcoPerMed)*, resulting in this PhD thesis. Toward the end of 2022 Heleen joined Pharos, the Dutch Centre of Expertise on Health Disparities. Here she has joined forces with colleagues from a wide variety of backgrounds, aiming to use her skills as a health economist to help reduce health inequity in the Netherlands. Recently, Heleen finished a research project commissioned by the Dutch Ministry of Health that investigated access to long-term care and care use among elderly people with a first- or second-generation migration background. Heleen is also a flutist and is currently studying towards a Bachelor of Music at the Jazz & Pop department of the Utrecht Conservatory.