Eyes on the prize: early economic evaluation to guide translational research

de Graaf, Gimon

DOI:
10.33612/diss.100467716

Eyes on the prize: early economic evaluation to guide translational research

Examples from the development of biomarkers for type 2 diabetes

Gimon de Graaf

Colofon

Eyes on the prize: early economic evaluation to guide translational research
Examples from the development of biomarkers for type 2 diabetes
ISBN/EAN: 978-94-034-2057-8 (printed version)
ISBN/EAN: 978-94-034-2056-1 (electronic version)

The research presented in this thesis was performed within the framework of the Center for Translational Molecular Medicine (CTMM); project PREDICCt, and supported by the Dutch Heart Foundation, Dutch Diabetes Research Foundation, and Dutch Kidney Foundation.

Cover design by Seroj de Graaf | Seroj.nl
Layout and design by David de Groot | Persoonlijkproefschrift.nl
Printed by Ridderprint BV | Ridderprint.nl

rijksuniversiteit
groningen

Eyes on the prize: early economic evaluation to guide translational research

Examples from the development of biomarkers for type 2 diabetes

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 13 november 2019 om 16.15 uur

door

Gimon de Graaf

geboren op 15 november 1984
te Nijmegen

Promotor

Prof. dr. E. Buskens

Copromotor

Dr. D. Postmus

Beoordelingscommissie

Prof. dr. H.W. Frijlink

Prof. dr. K.G.M. Moons

Prof. dr. J.L. Severens

CONTENTS

CHAPTER 1 **7**
General introduction

CHAPTER 2 **23**
Using multi-criteria decision analysis to support research priority setting in biomedical translational research projects

CHAPTER 3 **47**
The early economic evaluation of novel biomarkers to accelerate their translation into clinical applications

CHAPTER 4 **69**
A method for the early health technology assessment of novel biomarker measurement in primary prevention programs

CHAPTER 5 **97**
Design of stepwise screening for prediabetes and type 2 diabetes based on costs and cases detected

CHAPTER 6 **121**
The impact of short lead time for diabetes on the cost-effectiveness of treating patients Identified in prediabetes screening

CHAPTER 7 **141**
General discussion

CHAPTER 8 **167**
Summary 168
Nederlandse samenvatting 172
Dankwoord 176
Other SHARE dissertations 184
About the author 188

Chapter 1

General introduction

THE CHANGING CONSTRAINTS FOR HEALTHCARE INNOVATION

We may have reached a point in healthcare where we are able to do more than we can or are willing to afford. As a consequence of the ongoing technological development in healthcare, the number of diseases, syndromes, and conditions for which no form of intervention is available, has become very limited. In the not too distant past, the arrival of a new healthcare technology almost always represented new treatment possibilities for patients that could not be treated before or a drastic improvement to what was previously possible. This has created a persistent positive attitude towards healthcare innovations among doctors, patients, and the general public that lasts until the present day.[1] Nowadays, however, new technologies that enter the market often present only a minor benefit over existing ones, if at all. This is especially true for the major disease fields such as cardiovascular disease and cancer, which, due to their potential large target market, receive the most interest from researchers, funding agencies, pharmaceutical companies, and device manufacturers.[2] Independent of the magnitude of their added clinical benefit, new technologies almost always come at a higher cost than available ones. The welcoming attitude towards new technologies is therefore a substantial driver of the increase in healthcare costs that has been observed in most developed countries during the past decades.[3,4]

The rising healthcare costs are increasingly seen as a problem and a threat to the sustainability of healthcare systems. As a response, governments are increasingly initiating cost containment actions. These often take the form of budget cuts, caps, or maximum growth agreements. This means that when new, expensive medical technologies are incorporated in the care practice, spending on other modes of care provision has to be reduced. This is known as displacement.[5,6] The health benefits foregone because of the displacement of existing modes of health care provision are known as the opportunity cost of the new technology.[6,7] When the new intervention that is funded produces less

health than the displaced care, the amount of health in the total population is reduced. Therefore, new technologies such as the expensive cancer medicines that have been introduced on the market over the past years do not only pose a threat to financial sustainability, but in fact also to population health.

In order to prevent a reduction in population health through displacement, only those new technologies that produce more health for a given amount of financial resources should be introduced in the health care system. This requires a thorough assessment of the impact of a new technology on resource use and health effects, which can be obtained using Health Technology Assessment (HTA).

HEALTH TECHNOLOGY ASSESSMENT

HTA is a multidisciplinary method of evidence synthesis that considers evidence on safety, clinical effectiveness, and cost of health technologies.[8] The term technology should be interpreted in the broadest possible way. It refers to all proceedings and means used in healthcare, including pharmaceuticals, diagnostic tests, and medical devices, but also treatment protocols or the choice between immediate action and watchful waiting. In a broader application, HTA can include social, ethical, and legal aspects of the use of health technologies. Which aspects are included in an HTA depends on the purpose of the evaluation, i.e., the decision it aims to inform. In practice, costs and health effects are the dominant aspects in HTAs, as their purpose is most often to inform decisions on the reimbursement and adoption of new medical technologies. The evidence synthesized in an HTA often comes from epidemiological studies or clinical trials (evidence on health effects), and costing studies or other economic evaluations (evidence on costs). An HTA is always an incremental analysis, meaning that it will compare two or more competing alternatives. Most often these are a new intervention and the current way patients are treated (referred to as care as usual or current care). The dominant outcome measure used in HTA is the ratio of additional

cost per unit of health effect gained. The latter can be a disease-specific effect (such as the number of exacerbations in COPD), but it is more often a general effect (life years or quality-adjusted life years). This outcome ratio is referred to as an incremental cost-effectiveness ratio. Using a general effect measure allows comparing interventions for different diseases and is therefore almost always demanded by regulatory authorities for decisions on adoption and reimbursement. A reference cost-effectiveness threshold based on the overall production efficiency of the healthcare system can be used to determine whether the new technology will produce more health than the displaced care modalities. Governments and market regulators increasingly use such insights produced by HTAs in their decision to adopt and reimburse new technologies.

THE RELEVANCE OF HTA TO RESEARCHERS, DEVELOPERS, AND INVESTORS

When the cost-effectiveness of a new intervention is one of the criteria that determine its adoption and reimbursement, it becomes a factor critical to commercial success. Therefore, in order to make sound decisions on whether a new concept is worth developing or investing in, developers and investors must assess the potential of a new technology to be a cost-effective intervention. Likewise, when selecting from multiple targets, prototypes, or development portfolios, an estimate of potential cost-effectiveness of the alternatives is an important decision criterion. HTA performed in this setting – before or during development – is referred to as early HTA.

Public investors in research have an obligation to maximize the societal benefit of their investments. For them, an assessment of potential cost-effectiveness is critical to fulfilling that obligation. Public or public-private funders of translational research such as the Center for Translational Molecular Medicine (CTMM) or the European Commission (Horizon 2020) allocate large sums to address an abstract societal goal (such as the reduction

of burden from diabetes). In practice, there are often many ways in which such an abstract goal could be reached, not all of which have the same expected impact or likelihood to succeed. Their responsibility towards society obliges public investors to select those research proposals that have the highest expected societal benefit. Early HTA can be used to make an early assessment of the potential impact of translational research projects on quality of life and healthcare costs.

The difference between early HTA and mainstream HTA

Early and mainstream HTA differ on two main aspects. First, the aim of the analysis and research questions are different.[9] Mainstream HTA is most often used to support adoption or reimbursement decisions. Early HTA, on the other hand, is used to inform decisions on investment, portfolio management, and price setting, among other strategic business decisions. Second, the available evidence at the time of analysis is different. For mainstream HTA, the intervention is clearly defined, and there is almost always trial or other experimental data on the impact of the intervention on costs and effects. In early HTA, the intervention is not well defined. Rather, the research question of an early HTA could be to identify the most promising form of the intervention. Also, data on the impact of an intervention is seldom available. This, however, does not mean that no useful analysis can be performed. Valuable insights can be obtained by collecting information on the current care setting of the intended target population, such as epidemiological data and the costs and health effects of the current intervention. Synthesis of such evidence in a model enables the testing of the central premise of the mechanism by which a novel intervention might improve health and cost outcomes. This compels the formulation of a clear definition of a set of key characteristics of the new intervention, such as a precise definition of the patients who should receive the intervention and how and by whom the intervention should be provided, a process that is informative and thus valuable in itself.

Because early and mainstream HTA have different objectives, they have different outputs. The central outcome of a mainstream HTA is most often the aforementioned incremental cost-effectiveness ratio. Due to the large uncertainty in the input data in an early HTA, this outcome is not the most informative. Instead, indicating boundaries or tipping points of key parameters are more informative as they can be used as input during research and development processes.

As a scientific sub-field, early HTA is still very young, with most papers being published during the past ten years.[9,10] Many of the methods for early HTA are still in concept or pilot phase.[11] Their application by investors and developers for investment decisions, portfolio management, and R&D decisions is still very limited. A strong catalyst for the development of early HTA methods is the demand for the incorporation of early HTA in research projects by several large public-private partnerships and international funding agencies.

THE CTMM PREDICCT PROJECT

The Center for Translational Molecular Medicine (CTMM) was a large Dutch public-private partnership, consisting of several partners from academia (25% of funding), industry (25% of funding), and government (50% of funding). The rationale was that translational research could be done more effectively if experts from these partners cooperated in all phases of development. Historically, translational research is meant to bridge the so-called bench-bedside gap.[12] This gap is perceived to exist between the vast amount of knowledge on the biomedical processes underlying disease produced by fundamental research on one hand, and the slow progress in clinical care which is supposed to benefit from this knowledge on the other. Many different definitions of and approaches to translational research exist.[12] Within CTMM, the goal was to develop novel techniques based on insights from molecular medicine to improve diagnostic and treatment capabilities in the most prominent disease areas in western society, i.e., cardiovascular disease,

oncology, degenerative disease, and auto-immune disease. These improved capabilities were expected to improve the health outcomes for patients as well as the sustainability of healthcare systems. Due to the translational nature of the CTMM projects, early HTA was considered an important tool to inform strategic decisions and provide early estimations of the potential impact on the set objectives. As a result, an HTA work package was part of every CTMM project. This approach gave a substantial impulse to the development and application of methods for early HTA in translational research.

One of the CTMM research consortia was the PREdiction and early diagnosis of DIabetes and diabetes-related Cardiovascular Complications (PREDICCt) project. This project was initiated with the aim to develop innovative biomarker-based technologies to allow identification of individuals at increased risk of type-2 diabetes mellitus (DM2) and related complications.[13] The research presented in this thesis was conducted as part of the CTMM PREDICCt project.

Type-2 diabetes

Diabetes Mellitus is a group of metabolic disorders in which the regulation of blood glucose levels is disrupted. This leads to high blood sugar levels over prolonged time periods. In DM2 this is caused by insulin resistance, whereby cells in the body are less responsive to insulin. Lack of physical exercise and obesity are important factors contributing to the development of DM2. As obesity rates rise around the world, so does the incidence of DM2. The worldwide prevalence is estimated to rise to 642 million people by 2040.[14] The burden of DM2, both for patients as well as society, is for the largest part caused by its complications. Complications are usually categorized into microvascular (damage to small blood vessels) and macrovascular (damage to large blood vessels). The most common microvascular complications are damage to the eyes, kidneys, and nerves (called retinopathy, nephropathy, and neuropathy, respectively). This can lead to blindness, kidney failure, skin

damage, and amputation of extremities. Macrovascular complications include coronary artery disease, stroke, and peripheral vascular disease. Diabetes patients have a 2 to 4 fold increased risk for coronary heart disease.[15]

Because of their contribution to the burden of disease, diagnostic and treatment protocols for DM2 are to a large extent focused on the prevention of complications (tertiary prevention). Treatment of DM2 patients is aimed at regulating glucose levels in order to minimize vascular damage. In addition, complication risk is reduced by treating hypertension and dyslipidemia. Also, DM2 patients are regularly screened for the occurrence of complications such as retinopathy.

Strategies to reduce the burden of disease from DM2

The rise in prevalence of obesity and DM2 calls for improved strategies to prevent DM2 and its complications in order to avoid a large societal burden. Several strategies are possible, ranging from primary prevention (aiming to reduce the incidence of DM2) to better disease management and early detection of (people at risk for) complications (tertiary prevention). The target population for primary prevention is the general population. Therefore, strategies in primary prevention are generalized to a broad audience (e.g., lifestyle advice). Most often it is proposed to target a subgroup of patients who are at increased risk to develop diabetes for such interventions. A well-established high-risk group are patients with impaired glucose regulation, also known as prediabetes. In this condition, glucose regulation is abnormal, but not yet so severe that it can be classified as diabetes. On the other hand, tertiary prevention has to be more specific to individual patient characteristics, in order to take into account specific disease risk, risk factors, and comorbidities. A challenge in that area is to obtain a detailed profile of individual risk factors in order to provide an effective intervention for that individual. Historically, characteristics such as age, anthropometric measurements (e.g., height, weight, waist circumference), and lifestyle (e.g.,

smoking, diet) have been used to determine a personal risk profile. More recently, advances in molecular diagnostics have engendered enthusiasm and high expectations on the possibilities for personalized medicine.

PERSONALIZED MEDICINE AND BIOMARKERS

The mapping of the human genome (genomics), the increased insight in the regulation of the transcription of the genome (transcriptomics), and expanding knowledge on the function of proteins in the body (proteomics) have repeatedly challenged conventional definitions of diseases. Increasingly, different pathological mechanisms are identified within what was previously seen as one disease. These differences in pathological mechanisms at a molecular level are hypothesized to be driving differences in disease progression and response to treatment that are observed in patient populations with seemingly the same disease. As such, these discoveries have led to a new paradigm in medical science that foresees improved treatments and outcomes by means of grouping patients based on their risk for disease or response to a therapy. Personalized medicine, precision medicine, and stratified medicine are all labels for this paradigm. Within the paradigm of personalized medicine, many research efforts are aimed at identifying novel biomarkers. A biomarker is a substance, structure, or process that can be measured in or on a person or specimen, which can provide information on the incidence or outcome of a disease.[16] From a clinical perspective, biomarkers can be considered diagnostic tests: they are used to obtain information on the risk or stage of disease or treatment response, in order to optimize the care for a patient. Besides the role of a diagnostic test, biomarkers have many different applications in the disease-therapy continuum (Figure 1).

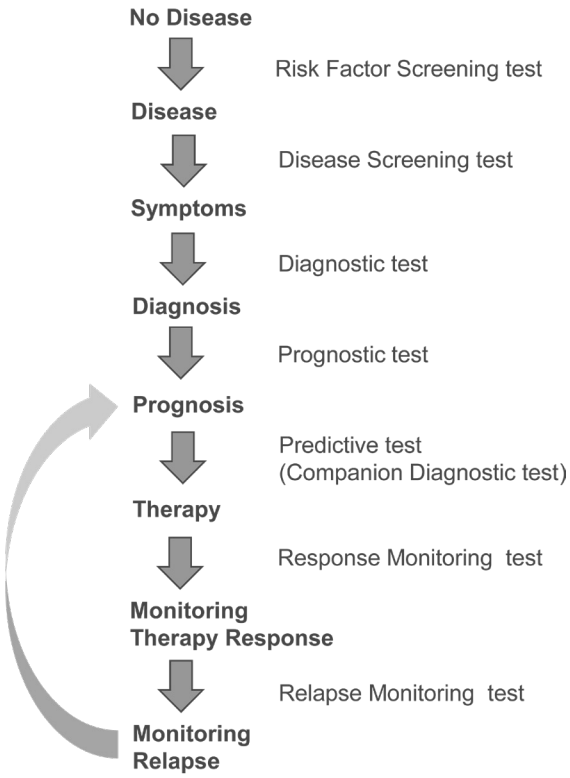


Figure 1: Possible applications of biomarker-based tests in the disease-treatment continuum. [31]

The unfulfilled potential of biomarkers

The hopes that newly discovered biomarkers enable personalized medicine strategies and therefore improved clinical outcomes, fewer side effects, and more cost-effective treatments have spawned a massive effort to identify new biomarkers for a wide variety of diseases.[17–20] Unfortunately, the vast amount of biomarker research fails to live up to the expectations.[20–27] This can be explained in part by the fact that much less effort has been put in translating newly discovered biomarkers into clinical applications than in discovering new biomarker candidates. The translational process from newly discovered biomarker to a diagnostic or prognostic test used in the clinic is a long and complex process requiring substantial financial investments. It

requires several different types of studies generating evidence on diagnostic accuracy, clinical effectiveness, and finally cost-effectiveness. Much like in the sequence of clinical trials used to determine the safety and effectiveness of novel pharmaceuticals, each step presents a hurdle that some candidates will fail to pass.[21,28,29] Only very few discovered biomarker candidates make it to the clinic (Figure 2).[30] Therefore, in order to support strategic decision making, each step requires an (updated) assessment to determine which candidates have enough potential to justify the required investments, and to determine their most promising clinical application. Thus far, well described and proven methods to generate evidence inform these decisions are lacking, leading to poor research and investment decisions and a stagnation of biomarkers in the translational process. In the end, this entails both a loss in health potential for patients and society, as well as wasted resources for public and private investors in research. Novel early HTA methods are therefore urgently needed.

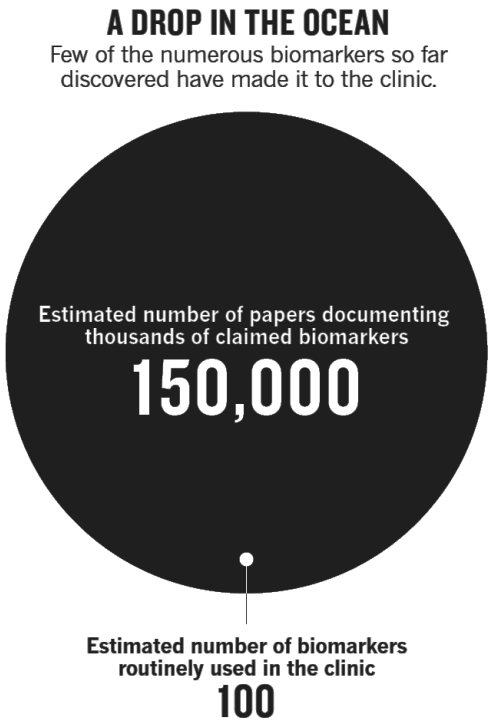


Figure 2: The personalized medicine paradigm has resulted in countless biomarker publications, but so far has made little impact in the clinic. [30]

AIM OF THIS THESIS

Our primary objective was to assess the clinical and economic value of the biomarkers and biomarker-based technologies that were developed within the CTMM PREDICCt project. As we set out to do this, we found that methods to perform such analyses were lacking. As a result, our second objective was to further the methodology for the early economic evaluation of biomarkers so that future R&D and investment decisions can be better informed.

OVERVIEW OF THIS THESIS

The two aims are entwined throughout this thesis, as methods were developed to address specific research needs for the PREDICCt project. Most chapters present a novel method for the early economic evaluation within translational research projects and demonstrate this method by applying it to the PREDICCt project. Chapter 6 is an exemption in that it focuses on a key issue of DM2 screening using established methodology. The chapters in this thesis are ordered in chronological order from the perspective of a translational research project, starting with an abstract societal objective and working towards specific biomarker-based technologies. When a project is selected for funding or when a project commences, a translation of the abstract research objectives into concrete research activities has to be made in the form of priority setting. In **chapter 2** we demonstrate how research priority setting can be done using multi-criteria decision analysis. When a specific research target is chosen, biomarkers are identified through their association with the relevant clinical endpoint. **Chapter 3** demonstrates how the clinical application of a biomarker candidate can be defined and how the data from an association study can be used to make an early estimate of the clinical and economic impact of a biomarker candidate. Similarly, **chapter 4** demonstrates an early estimate of the cost-effectiveness specifically for biomarkers that are to be applied in the context of primary prevention. Continuing further towards the application of a new biomarker-based

technology in primary prevention, **chapter 5** demonstrates a method for the optimization of a 2-step screening program on costs and number of cases detected. The case study presented in this chapter estimates the efficiency of currently available screening techniques and thereby provides a benchmark for potential new biomarkers in this field. Finally, **chapter 6** assesses the effects of different lengths of lead-time of DM2 on the cost-effectiveness of a screening program for patients with impaired glucose regulation.

REFERENCES

1. Deyo, RA & Patrick, DL. *Hope or hype: the obsession with medical advances and the high cost of false promises*. (AMACOM, American Management Association, 2005).
2. Salas-Vega, S, Iliopoulos, O & Mossialos, E. Assessment of Overall Survival, Quality of Life, and Safety Benefits Associated With New Cancer Medicines. *JAMA Oncology* **3**, 382 (2017).
3. Baker, L, Birnbaum, H, Geppert, J, Mishol, D & Moyneur, E. The relationship between technology availability and health care spending. *Health Affairs* **Suppl Web**, W3-537–551 (2003).
4. Sorenson, C, Drummond, M & Khan, BB. Medical technology as a key driver of rising health expenditure: disentangling the relationship. *ClinicoEconomics and Outcomes Research* **5**, 223–234 (2013).
5. Claxton, K, Sculpher, M, Palmer, S & Culyer, AJ. Comment: Causes for concern: Is nice failing to uphold its responsibilities to all NHS patients? *Health Economics* **24**, 1–7 (2015).
6. van Baal, P, Perry-Duxbury, M, Bakx, P, *et al*. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Economics* **28**, 87–100 (2019).
7. Palmer, S & Raftery, J. Economic Notes: opportunity cost. *BMJ* (*Clinical research ed.*) **318**, 1551–2 (1999).
8. Luce, BR, Drummond, M, Jönsson, B, *et al*. EBM, HTA, and CER: Clearing the Confusion. *Milbank Quarterly* **88**, 256–276 (2010).
9. IJzerman, M & Steuten, LM. Early Assessment of Medical Technologies to Inform Product Development and Market Access: A Review of Methods and Applications. *Applied Health Economics and Health Policy* **9**, 331–347 (2011).
10. IJzerman, MJ, Koffijberg, H, Fenwick, E & Krahn, M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *PharmacoEconomics* **35**, 727–740 (2017).
11. Pham, B, Tu, HAT, Han, D, *et al*. Early economic evaluation of emerging health technologies: Protocol of a systematic review. *Systematic Reviews* **3**, 1–7 (2014).
12. van der Laan, AL & Boenink, M. Beyond Bench and Bedside: Disentangling the Concept of Translational Research. *Health Care Analysis* **23**, 32–49 (2015).
13. CTMM Predicct. *Biomarkers for the Prediction and Early Diagnosis of Diabetes and Diabetes-related Cardiovascular Complications: Output Report*. (2015).
14. International Diabetes Federation. *IDF Diabetes Atlas, 7th edition*. (2015).

15. Huxley, R, Barzi, F & Woodward, M. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. *BMJ (Clinical research ed.)* **332**, 73–78 (2006).

16. International Epidemiological Association. *A Dictionary of Epidemiology*. (2008).

17. Ludwig, J & Weinstein, J. Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer* **5**, 845–856 (2005).

18. Nordström, A & Lewensohn, R. Metabolomics: Moving to the clinic. *Journal of Neuroimmune Pharmacology* **5**, 4–17 (2010).

19. Moons, KGM. Criteria for scientific evaluation of novel markers: A perspective. *Clinical Chemistry* **56**, 537–541 (2010).

20. Anderson, NL, Ptolemy, AS & Rifai, N. The riddle of protein diagnostics: Future bleak or bright? *Clinical Chemistry* **59**, 194–197 (2013).

21. Ioannidis, JPA. Is molecular profiling ready for use in clinical decision making? *The Oncologist* **12**, 301–311 (2007).

22. Ioannidis, JPA & Panagiotou, OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA: The Journal of the American Medical Association* **305**, 2200–2210 (2011).

23. Lumbreras, B, Parker, LA, Porta, M, *et al*. Overinterpretation of clinical applicability in molecular diagnostic research. *Clinical Chemistry* **55**, 786–794 (2009).

24. Vickers, AJ, Jang, K, Sargent, D, Lilja, H & Kattan, MW. Systenatic Review of Statistical Methods Used in Molecular Marker Studies in Cancer. *Cancer* **112**, 1862–1868 (2008).

25. Williams, PM, Lively, TG, Jessup, JM & Conley, BA. Bridging the Gap: Moving Predictive and Prognostic Assays from Research to Clinical Use. *Clinical Cancer Research* **18**, 1531–1539 (2012).

26. Zolg, W. The proteomic search for diagnostic biomarkers: lost in translation? *Molecular & Cellular Proteomics* **5**, 1720–1726 (2006).

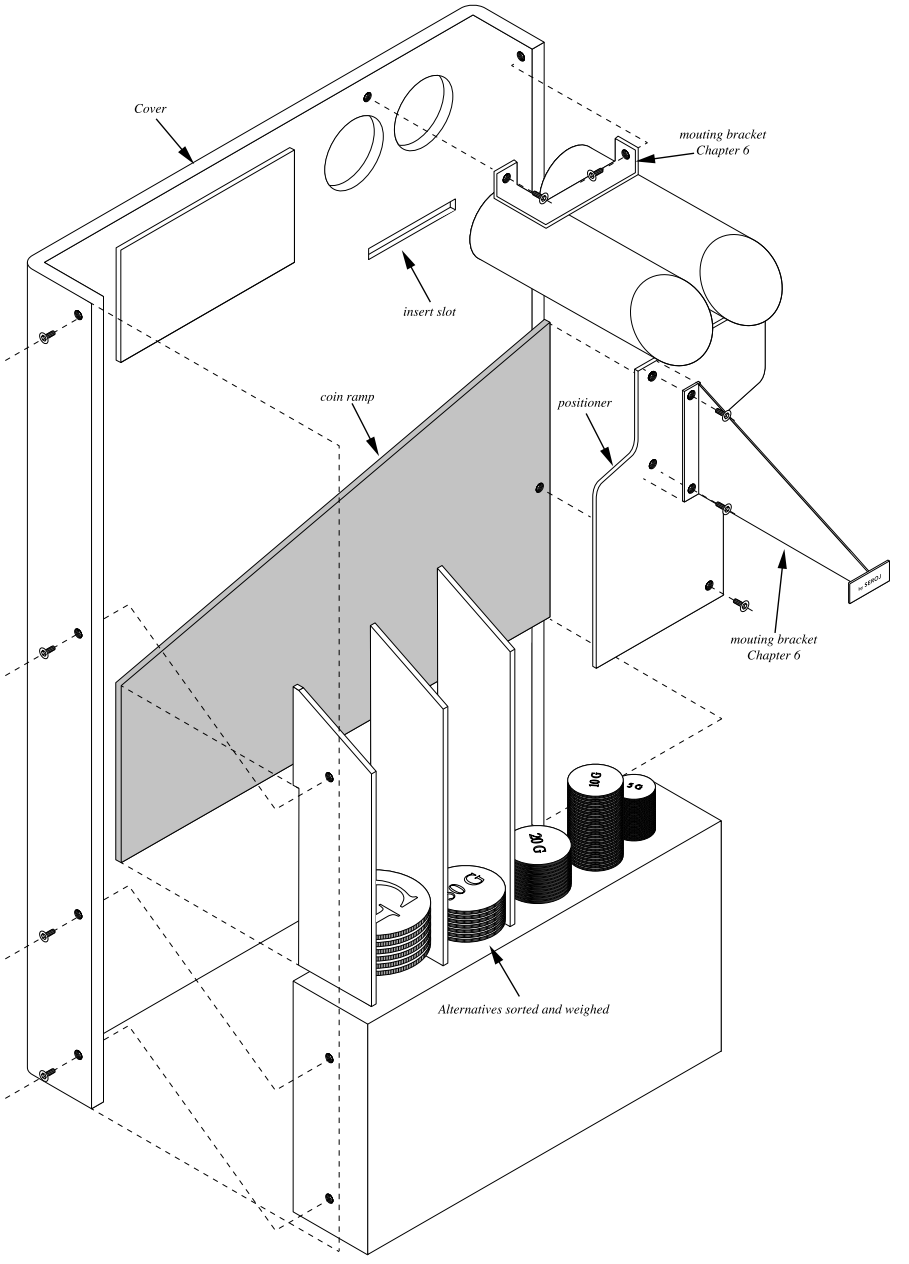
27. Gyawali, B. Point: The imprecise pursuit of precision medicine: Are biomarkers to blame? *Journal of the National Comprehensive Cancer Network* **15**, 859–862 (2017).

28. Van den Bruel, A, Cleemput, I, Aertgeerts, B, Ramaekers, D & Buntinx, F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *Journal of Clinical Epidemiology* **60**, 1116–1122 (2007).

29. Sackett, D & Haynes, R. The architecture of diagnostic research. *BMJ* **324**, 539–541 (2002).

30. Poste, G. Bring on the biomarkers. *Nature* **469**, 156–157 (2011).

31. CTMM. *Medical Technology Assessment for Innovative Molecular Medicine*. (2015).


Cover
mouting bracket
Chapter 6
insert slot
coin ramp
positioner
mouting bracket
Chapter 6
Alternatives sorted and weighed

Chapter 2

Using multi-criteria decision analysis to support research priority setting in biomedical translational research projects

Published as:

De Graaf, G., Postmus, D., & Buskens, E.

Using Multicriteria Decision Analysis to Support Research Priority Setting in Biomedical Translational Research Projects. *Biomed research international* (2015)

ABSTRACT

Translational research is conducted to achieve a predefined set of economic or societal goals. As a result, investment decisions on where available resources have the highest potential in achieving these goals have to be made. In this paper, we first describe how multicriteria decision analysis can assist in defining the decision context and in ensuring that all relevant aspects of the decision problem are incorporated in the decision-making process. We then present the results of a case study to support priority setting in a translational research consortium aimed at reducing the burden of disease of type 2 diabetes. During problem structuring, we identified four research alternatives (primary, secondary, tertiary microvascular, and tertiary macrovascular prevention) and a set of six decision criteria. Scoring of these alternatives against the criteria was done using a combination of expert judgement and previously published data. Lastly, decision analysis was performed using stochastic multicriteria acceptability analysis, which allows for the combined use of numerical and ordinal data. We found that the development of novel techniques applied in secondary prevention would be a poor investment of research funds. The ranking of the remaining alternatives was however strongly dependent on the decision maker's preferences for certain criteria.

INTRODUCTION

The difficulty of developing biomedical discoveries into new medical technologies or therapies has been widely recognized, and is often referred to as the 'bench-bed gap' or the 'valley of death'.[1,2] Translational research aims to bridge this gap by integrating the societal needs identified at the bedside with the research done at the bench. It encompasses the entire value chain from basic biomedical research, through epidemiology, clinical testing, product development, policy and regulatory compliance, and marketing. As a result, the overall success of a translational research project is determined by a multitude of technological, clinical, economic, and regulatory factors. All these factors need to be considered when evaluating which of the available research strategies are most likely to yield innovations that will eventually gain widespread adoption in daily clinical practice. This makes priority setting for translational research a complex problem that requires decision makers to gather and synthesize expertise from different fields. Without the use of a formal decision support method, it is generally impossible to simultaneously consider all aspects of such a decision problem, making it likely that too much emphasis is put on a single outcome of the translational research process. In such a setting, the use of multi-criteria decision analysis (MCDA) can assist in structuring the problem and in making the decisions justifiable and replicable, thereby increasing accountability for public resources spend.[3]

In the context of government-sponsored technology development programs, MCDA has previously been applied to support the selection of research and development projects across different industries and focus areas.[4,5] However, these applications are not directly portable to research priority setting in biomedical translational research projects as the healthcare industry has specific properties that were not addressed in these studies. In particular, healthcare markets are heavily regulated and public provision of goods and services plays an important role in these markets. These characteristics impose rather strict constraints with respect to market penetration and

price setting that already need to be considered early during the translational research process. In this paper, we demonstrate how these aspects can be incorporated in a formal way by using MCDA for priority setting at the start of a translational research project. We illustrate this by means of a case study conducted within the context of a translational research project aimed at the prevention of type 2 Diabetes Mellitus (DM2) and its related complications.

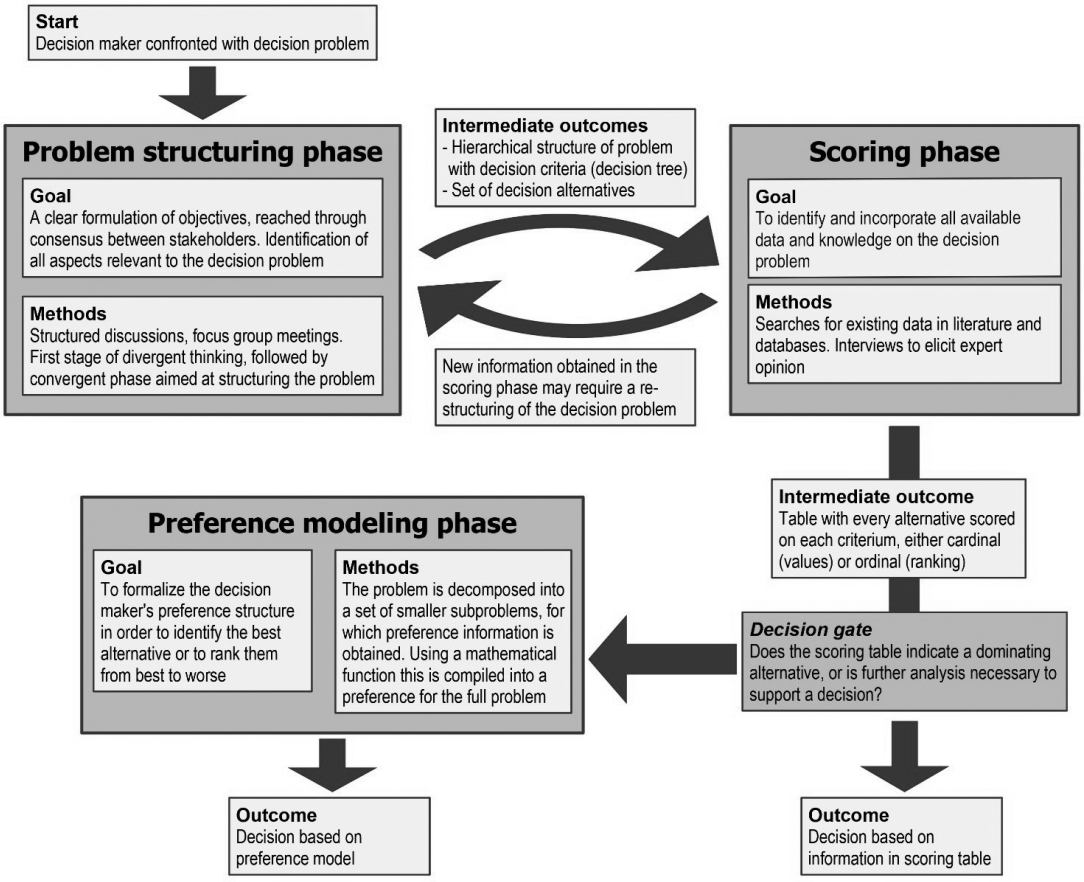


Figure 1: Schematic overview of the application of multi-criteria decision analysis for priority setting.

APPLICATION OF MCDA TO RESEARCH PRIORITY SETTING IN BIOMEDICAL TRANSLATIONAL RESEARCH PROJECTS

Research priority setting for biomedical translational research is a complex problem that requires decision makers to consider a multitude of technological, clinical, economic, and regulatory factors. In such situations, the use of a formal decision support method encourages the incorporation of views and knowledge from experts in different parts of the value chain of biomedical research, thereby reducing the possibility that at later stages in the product development process problems are encountered that in hindsight could already have been foreseen at the start of the project. It can also ensure that all available information related to the decision problem is incorporated into the decision-making process, thereby reducing the chance that the decision focuses too much on a single or narrow set of aspects of the problem. Within the framework of MCDA, this is achieved by sequentially going through the following three phases: problem structuring, scoring of the alternatives against the criteria, and preference modeling (Figure 1). Each of these phases is briefly described in the subsections below.

Problem structuring

During problem structuring, the different stakeholders involved in the decision-making process express their knowledge and views on the context of the decision problem as well as their objectives regarding the decision. Several formats and tools have been proposed to support this idea generation process, including "Post-It" sessions and various checklists and other aids to thinking such as adopting different perspectives and identifying barriers and constraints.[3] This divergent mode of thinking is followed by a convergent phase of idea structuring, in which ideas are clustered and aggregated to arrive at a set of decision alternatives (if not yet clearly defined at the start of the process) and a set of criteria against which these alternatives are to

be evaluated. Depending on the decision context, the definition of these criteria can to an extend be informed by objective knowledge of relevant cause-and-effect mechanisms from scientific literature or other sources. However, the criteria should reflect the objectives of the relevant decision makers and therefore should be derived discussions with the decision makers. Knowledge from outside the decision maker group can be incorporated into these discussions, but should never dictate criteria by itself. The output of the problem structuring phase is often a value tree. This is a graphical representation of the hierarchical ordering of the criteria.

Scoring of the alternatives against the criteria

The next step is to score the alternatives against these criteria, which is done at the lowest level of the value tree. For some criteria (e.g., cost), it may be possible to assess the performance of the alternatives numerically, whereas for others (e.g., quality), it may only be feasible to obtain an ordinal ranking of the alternatives or to allocate them to verbally defined levels of performance (e.g., poor, reasonable, excellent). How the alternatives are scored against the criteria differs from decision context to decision context and depends, amongst others, on the amount of data (e.g., results from observational and/or experimental studies, output from mathematical models, or expert opinion) that is available at the start of the decision-making process and on how many resources one is willing to invest in the collection of more precise measurements. As the information obtained in the scoring phase can change the perspective on the decision problem, it might be necessary to revert to the problem structuring phase in order to incorporate these new insights in the decision context. If this is not the case, the end of the scoring phase concludes the formal specification of the decision problem.

Based on the information in the scoring table, it is sometimes possible to identify one or more alternatives for which there is at least one other alternative that performs better on all of the criteria included in the decision

problem. As it is never optimal to select one of these dominated strategies, they can safely be eliminated from the set of decision alternatives. If there is sufficient budget to fund all the remaining strategies, the decision problem is solved, meaning that the multi-criteria decision-making process can be ended after the scoring phase. If not, the set of decision alternatives needs to be further reduced by making value trade-offs among the performance levels on the different criteria. In such situations, the use of preference modeling can assist in formalizing the decision makers' preference structures, thereby reducing the chance that the decision focuses too much on a single aspect of the decision problem.

Preference modeling

At the research priority setting stage of a translational research project, the amount of developmental uncertainty surrounding the conceived product concepts is usually still enormous. As a result, a full quantitative assessment of the expected clinical and economic benefits from each of the identified decision alternatives is generally not yet possible. It is therefore likely that for some of the criteria the data in the scoring table are solely based on expert opinion. As experts are often more comfortable with producing rankings (e.g., the number of competitor products is larger for alternative A than for alternative B) than with providing exact numerical estimates (e.g., there are 10 competitor products for alternative A and 6 for alternative B), it is important that such ordinal data can be accommodated in the preference modeling phase. For this reason, we will focus in this section on describing SMAA-O[6], a variant of the stochastic multi-criteria acceptability analysis (SMAA) method[7,8] that has been developed for decision problems where the data for some or all criteria is ordinal.

In SMAA-O, it is assumed that the decision maker's preference structure can be represented by means of a mathematical function $v(x)$ that is constructed in such a way that alternative i is preferred over alternative j if and only if

$v(x^i) > v(x^j)$ where x^i denotes the column of the scoring table associated with alternative i. To simplify the construction of $v(x)$, it is generally assumed that the criteria satisfy the independence conditions for applying the additive value function $v(x, w) = w_1 v_1(x_1) + \cdots + w_n v_n(x_n)$, where n is the number of criteria and w_k the weight attached to criterion k. The partial value functions $v_k(x_k)$, normalized so that the worst possible score on each criterion is assigned a value of 0 and the best possible score is assigned a value of 1, reflect the relative desirability of the different levels of achievement on the individual criteria. For numerical criteria, it is usually assumed that equal size ranges on the measurement scales represent the same amount of value to the decision maker, resulting in partial value functions that are linear. For ordinal criteria, the use of such a linear mapping between scale values and partial values is however not directly suitable as the distance between ranks on an ordinal scale is not known. In SMAA-O, this problem is dealt with by randomly assigning the scale values on the ordinal scale to partial values between 0 and 1, in such a way that the rank order between the scale values is maintained. Different ordinal to partial value mappings may translate into a different ranking of the decision alternatives as the overall value associated with each of these alternatives may change. This uncertainty is captured by the *rank acceptability indices* b_i^r, which describe the fraction of Monte Carlo iterations for which alternative i is ranked at place r. The pairwise winning indices c_{ij} describe the fraction of Monte Carlo iterations for which alternative i is ranked at a higher place than alternative j. Missing or imprecise information with respect to the values of the weights can be handled in a similar way by sampling the weight vector from a uniform distribution in the feasible weight space induced by the available preference information.

CASE STUDY

Decision problem

The PREdiction and early diagnosis of DIabetes and diabetes-related Cardiovascular Complications (PREDICCt) project of the Center for

Translational Molecular Medicine (CTMM) was initiated to enhance the possibilities for prevention of DM2 and associated complications through the development of methodologies for molecular diagnostics and molecular imaging of novel biomarkers associated with the development of DM2 and its related complications. DM2 is a complex disease with many genetic, environmental, and behavioral determinants as well as biological pathways involved. Additionally, it is a chronic disease that takes a long time to develop. As a result, there are many different possible target applications for novel diagnostic and imaging techniques. Not all target applications are however equally likely to achieve the objectives of the project to the same extent. As a result, a decision had to be made on the priority setting for the investment of available resources.

Problem structuring

Methods

Several discussion sessions were held with various researchers from the PREDICCt project. During these discussions multiple perspectives on the decision problem were suggested by participants and discussed in the group. Based on these discussions, a set of alternatives was defined. The business plan of CTMM, in which the stakeholders in the project expressed their views and interests, served as the starting point to define a set of criteria. All statements concerning objectives were isolated from the business plan and subsequently ordered and grouped.

Results

As the main aim of the PREDICCt project was the prevention of DM2 and associated complications, the decision alternatives were defined in the scope of the preventive medicine framework. Preventive medicine is often classified in three different levels. Primary prevention targets those in whom the disease is not yet present, with the aim to provide interventions to prevent the disease from manifesting. Secondary prevention targets those who have the disease, but are not yet symptomatic, aiming to reduce the morbidity through early

treatment. Tertiary prevention is aimed at those who are diagnosed with the disease, and enable the provision of interventions limiting further morbidity caused by complications. Complications of DM2 are an important aspect in this case, as most of the burden of the disease is caused by these complications.[9] There are two distinct categories of complications: microvascular (diabetic nephropathy, neuropathy, and retinopathy) and macrovascular (coronary artery disease, peripheral arterial disease, and stroke).[10] These two categories of complications have distinct approaches to prevention, diagnosis, and care. Therefore, it was considered important to make a distinction between tertiary prevention aimed at microvascular complications and tertiary prevention aimed at macrovascular complications. The 4 alternative research approaches identified for the development of a novel biomarker technology in DM2 were thus as follows:

A biomarker technology applied in the general population to

1. select individuals eligible for interventions aimed at preventing or delaying the onset of DM2 (primary prevention)
2. identify those with undiagnosed diabetes in order to initiate treatment earlier (secondary prevention)

A biomarker technology applied in the population of diagnosed DM2 patients to

3. select those that would benefit from interventions aimed at preventing or delaying *microvascular* complications (tertiary prevention)
4. select those that would benefit from interventions aimed at preventing or delaying *macrovascular* complications (tertiary prevention)

The structuring of objectives from the business plan resulted in the identification of four main objectives: reduce the burden of disease, reduce

healthcare costs, increase economic activity, and obtain a high academic profile.

The profile of academic output is to a large extend determined by the novelty and quality of scientific work presented. This is not directly related to the decision alternatives at hand, meaning that a high academic profile could be obtained no matter what alternative is chosen. This objective was therefore not considered relevant for the purpose of the present analysis. For the other three objectives, we conducted a literature review and a brainstorming session to identify a set of factors that are important determinants of these objectives and to identify potential barriers and constraints that hinder their achievement. This resulted in the value tree depicted in Figure 2.

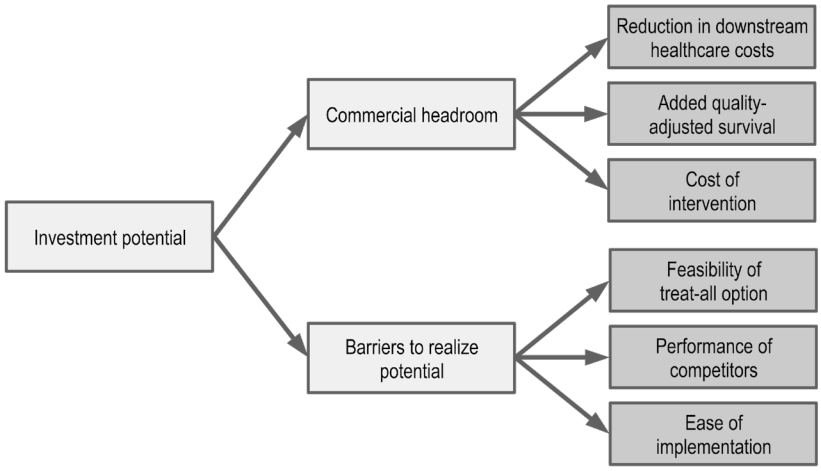


Figure 2: Value tree of overall and lower-level objectives of the public-private partnership.

In the healthcare technology market, the commercial potential of a product is dependent on its clinical value and its impact on the downstream healthcare consumption. The extend of this relation is determined by the level of regulation, which differs between jurisdictions as well as between different

parts of the healthcare system. For highly regulated parts of the healthcare system, the impact of these factors on a technology's commercial potential can be assessed quantitatively by conducting a headroom analysis.[11] The rationale behind this approach is that the estimated change in health effects and healthcare costs, both direct and indirect, resulting from the implementation of a new technology determine the value of the technology for society, and thereby the maximum device-related cost at with the use of this new product will still be reimbursed. As this cost provides an upper-bound for the price that the producer can charge for its product (the principle of value-based pricing), the amount of headroom available is a suitable proxy for the commercial potential of a new medical technology. The upper arm of the value tree therefore consisted of the following 3 determinants of the commercial headroom available: the decrease in downstream healthcare cost, the increase in quality-adjusted survival, and the cost of the intervention associated with the diagnostic or prognostic test. The 3 criteria forming the lower arm of the value tree captured the likelihood that the availability of a more accurate diagnostic or prognostic test will trigger changes in how the healthcare system currently operates. The feasibility of a treat-all option indicated the added value of the ability to treat specific patients as opposed to treating all patients. This provided an indication of the value stemming from better discrimination or prediction. Furthermore, the existence of high-quality competitor technologies, or lack thereof, was considered a major driver for the success of a novel technology to gain market share. Lastly, not all decision alternatives were considered equal in terms of the accessibility of the market and the ease of implementation in the clinical protocol. Technologies that readily fit within the practice as outlined by current guidelines can be implemented with relative ease. Contrarily, those that require a major change in clinical or public health protocols, for example the initiation of a universal screening program, cannot fulfill their potential until such changes are established.

Scoring of the alternatives against the criteria

Methods

For each of the decision alternatives, quantitative estimates of the decrease in downstream healthcare costs, the increase in quality-adjusted survival, and the intervention costs were available in the literature. The performance of the decision alternatives on these criteria was therefore expressed numerically. The performance on the other 3 criteria are strongly dependent on the type of technology developed and can therefore not be quantified at this stage. We therefore used expert opinion to formulate an ordinal ranking of the decision alternatives with respect to these criteria.

Results

The complete scoring matrix is shown in Table 1. Estimates of the effects of primary prevention of diabetes and tertiary prevention of macrovascular complications on the reduction of downstream healthcare costs, gain of quality-adjusted survival, and the costs of interventions were based on a modeling study.[12] For the primary prevention scenario, a lifestyle intervention program in obese individuals was modeled, and for the tertiary prevention of macrovascular complications, a multi-factorial treatment scenario combining intensive glycemic control, cholesterol-lowering treatment, and antihypertensive treatment was modeled. Estimates of the reduction in downstream healthcare cost, gain of quality-adjusted survival, and the costs of interventions for tertiary prevention of microvascular complications was based on a study that modeled the results of intensive blood glucose control and use of ACE-inhibitors on nephropatic complications.[13] As studies have found that secondary prevention of DM2 has little to no effect on downstream healthcare costs and quality-adjusted survival, the performance of this alternative on these two criteria was set equal to 0.[14] However, in case screening is performed and patients are discovered, they will be treated. Therefore, the treatment costs of diabetes patients without complications were included.[15]

Table 1: Scoring of the decision alternatives against the evaluation criteria

	Preference direction	Primary prevention	Secondary prevention	Tertiary prevention microvascular	Tertiary prevention macrovascular
Reduction in downstream healthcare costs	Increasing	€ 658M	€ 0	€ 73M	€ 312M
Added quality-adjusted survival	Increasing	€ 280K	€ 0	€ 1K	€ 80K
Cost of related intervention	Decreasing	€ 792	€ 663	€ 155	€ 561
Feasibility of Treat-All option		2	1	4	3
Performance of existing tests		3	4	1	2
Ease of implementation		2	2	1	1

Two main aspects contributed to the ranking of the feasibility to treat-all criterion: the budget impact and lack of implementation of existing cost-saving interventions. Primary and secondary prevention were ranked as more interesting, as treating all, or large parts of the target population would not be feasible due to budget impact reasons. Within tertiary prevention, the microvascular complication alternative was ranked lowest as cost-saving interventions are readily available there, but not yet fully implemented.[13] The barriers to implement such interventions must therefore first be overcome before the improved risk stratification possibilities can be implemented. Considering the performance of existing competing technologies, secondary prevention was ranked lowest. There, the diagnosis of diabetes itself cannot be improved as the disease is defined on measurements with the gold standard (glucose measurements). Additionally, there are numerous pre-screening tools available that perform well and cost little (risk questionnaires).[16] As a result of the latter, primary prevention was ranked second lowest. On the contrary, such risk stratification tools are hardly available, and perform less well, for microvascular complications, and to a lesser extend macrovascular

complications. Lastly, the primary and secondary prevention settings of diabetes would necessitate some form of screening. Such a public health program could take years before realized. This entails a serious problem for the implementation of any biomarker technology. As diagnosed diabetes patients regularly consult a physician, access to the patient is less problematic in the case of tertiary prevention.

Preference modeling

Methods

The partial value functions for the numerical criteria were obtained by linearly rescaling the criteria measurements to the interval [0,1], with the values of 0 and 1 assigned to the worst and best levels of performance on these criteria, respectively. The rankings of the alternatives on the ordinal criteria were randomly mapped to partial values between 0 and 1 consistent with these rankings by using the SMAA-O method. With respect to the weights, we specified three scenarios. First, we considered a base case scenario in which no additional constraints on the values of the weights were incorporated. The results of such a preference-free analysis can be used to eliminate alternatives that always fall short to at least one other alternative, irrespective of the decision maker's preferences. Second, we considered a scenario where a large commercial headroom was considered more important than avoiding barriers to realize potential, implying that $w_1 + w_2 + w_3 > w_4 + w_5 + w_6$. Lastly, we considered a scenario where the previous preference statement was reverted, implying that $w_1 + w_2 + w_3 < w_4 + w_5 + w_6$. All analyses were conducted in R (version 3.0.1) using the smaa (version 0.1.1) and hitandrun (version 0.2.2) packages that are available from CRAN.

Results

For the preference free analysis (Figure 3), we found that secondary prevention has a very low (<0.05) first rank acceptability index, making it unlikely to be optimal for any decision maker. The optimality of the three remaining strategies was however strongly dependent on the decision maker's

preferences. Primary prevention was very likely to be the best alternative when maximizing the commercial headroom available is considered more important than minimizing the barriers and constraints to utilize this headroom (Figure 4). This is confirmed when looking at the pairwise winning indices, which show that the probability that primary prevention is preferred over tertiary prevention of microvascular complications, the second best alternative when improvement of commercial headroom is favored, is 61% (Table 2). Contrarily, tertiary prevention of microvascular complications and tertiary prevention of macrovascular complications were clearly the preferred strategies when having to deal with lesser obstacles is preferred over potential higher gains in terms of the objectives stated by the stakeholders (Figure 5). However, as is shown by the pairwise winning indices for this scenario (Table 3), the provided preference information with respect to the values of the weights was not precise enough to further discriminate between these two remaining strategies.

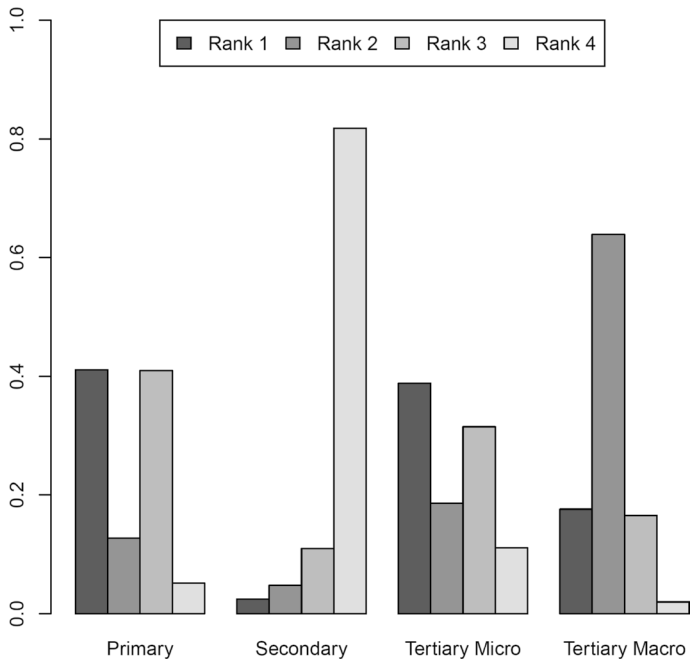


Figure 3: Rank acceptability indices for the base case scenario.

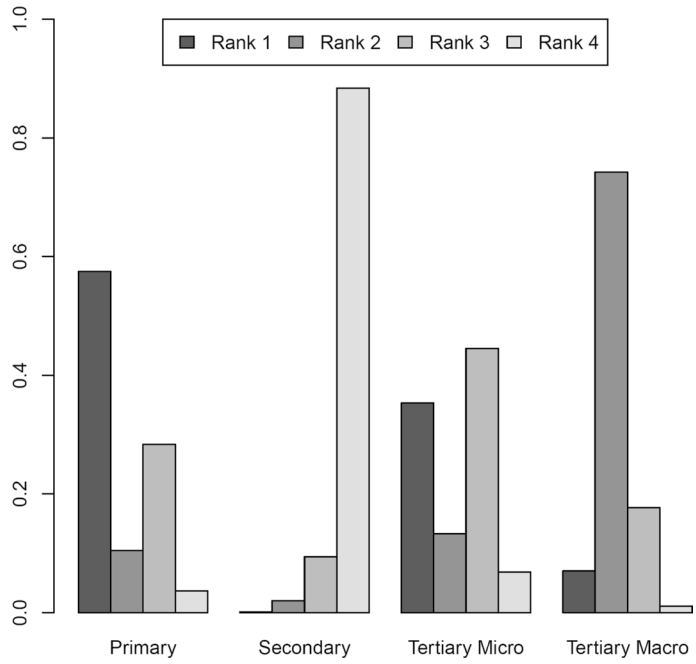


Figure 4: Rank acceptability indices when improvement of commercial headroom is favored.

Table 2: Pairwise winning indices when improvement of commercial headroom is favored

	Primary prevention	**Secondary prevention**	**Tertiary prevention microvascular**	**Tertiary prevention macrovascular**
Primary prevention		0.96	0.61	0.65
Secondary prevention	0.04		0.07	0.02
Tertiary prevention microvascular	0.39	0.93		0.45
Tertiary prevention macrovascular	0.35	0.98	0.55	

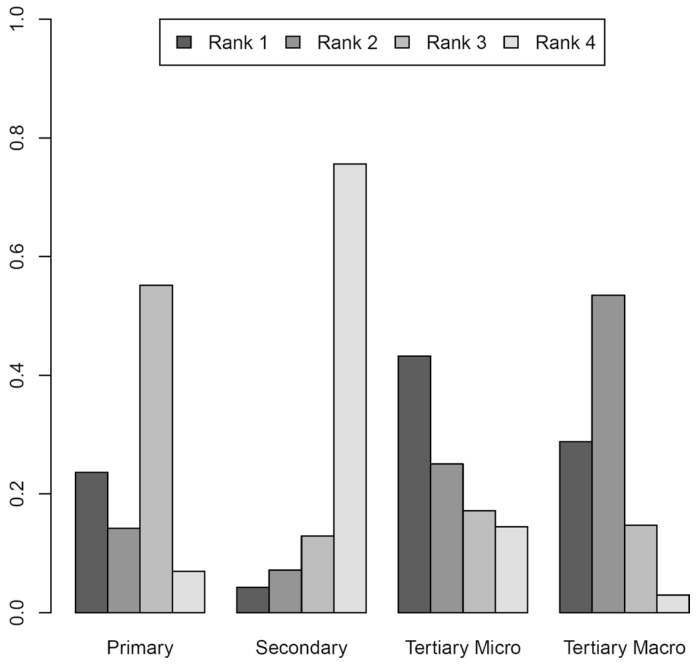


Figure 5: Rank acceptability indices when avoidance of barriers is favored.

Table 3: Pairwise winning indices when avoidance of barriers is favored

	Primary prevention	**Secondary prevention**	**Tertiary prevention microvascular**	**Tertiary prevention macrovascular**
Primary prevention		0.88	0.35	0.31
Secondary prevention	0.12		0.18	0.12
Tertiary prevention microvascular	0.65	0.82		0.48
Tertiary prevention macrovascular	0.69	0.88	0.52	

DISCUSSION

Priority setting for translational research is a complex problem that requires decision makers to gather and synthesize expertise from different fields.

In this paper, we have shown through a case study how this process can be supported in a formal way by applying MCDA.

The complete value chain in biomedical innovation poses a complex and multifaceted problem for priority setting. Additionally, ethics, public opinion, and politics come into play when dealing with a healthcare setting. Under these conditions, informal decision-making will lead to the use of intuitive and heuristic approaches as a decision maker is unable to grasp the full complexity and trade-offs in a decision.[17] Informal decision-making will therefore depend to a large extend on who is appointed to make the decision, and what the background expertise of the decision maker (or group of decision makers) is, which would be undesirable in case of large investments or investments of public funds. The problem structuring phase of MCDA helps to overcome this by encouraging the incorporation of expertise exogenous to the decision makers. In our case study, this led to the integration of two different perspectives on the decision problem: that of the commercial headroom (based on the improvement in diagnostic power of new technologies over existing ones), and that of the barriers that new technologies would face to access the market. After the scoring phase it became apparent that the development of novel methods to measure biomarkers that can be used in secondary prevention of DM2 was certainly an unattractive research objective. If decision makers were willing to invest in all remaining three alternatives, the priority setting process could be stopped after this phase. However, in order to explore under which preferences the remaining alternatives would be most attractive, we proceeded with the preference modeling phase. A preference of decision makers for the maximization of commercial headroom made the development of novel methods to measure biomarkers used in primary prevention the most attractive strategy. Alternatively, investing in novel methods to measure biomarkers for tertiary prevention of microvascular and macrovascular complications was optimal in case a safer strategy with fewer obstacles, but less gain, would be preferred.

Early health economic modeling—the process of performing an initial assessment of the costs and health effects associated with a new medical technology before the technology has been fully developed—has recently been suggested as a tool to inform new product development within translational research projects.[18–20] However, given that such calculations require very strict assumptions about how a new technology performs in a specific clinical setting, this approach cannot yet be applied when specific biological targets still need to be identified. Other, softer approaches such as SMAA-O are therefore required to support research priority setting at the start of a translational research project, where outcomes are generally too uncertain to make a full quantitative assessment of the expected return-on-investment meaningful. Using MCDA for priority setting at the beginning of a research project can facilitate decision-making further on in the research and development process. For example, the data during the scoring phase can serve as input for quantitative approaches such as headroom analysis for product investment decision-making[11] and value-based pricing for market access.[21] We therefore see SMAA-O or similar MCDA methods as a new instrument in the early health technology assessment toolbox, being one to be used at the very start of translational research projects.

A strength of the SMAA-O methodology that we employed in our case study is the possibility to combine ordinal and numerical scoring of the alternatives. This allowed us to make full use of the large amount of data available in the scientific literature on costs and health burden related to DM2, while still being able to incorporate expert judgment on aspects for which no data was available. A limitation of our study is that, apart from the scenarios considered, we did not elicit any preference information on the weights from the decision makers. Ordinal and ratio constraints on the weights can however easily be incorporated in a SMAA analysis by utilizing efficient weight generation techniques such as hit-and-run sampling.[22]

We have demonstrated in this paper how the priority setting in translational research may be approached by applying MCDA. Future research is needed to fully assess the applicability of this method at the very start of a translational research project. Nonetheless, we are confident that we have already made a convincing case for formal decision-making in priority setting in translational research. Our report may serve as a guide for future decision makers, ultimately making the approach common practice.

REFERENCES

1. Adams, J. Building the bridge from bench to bedside. *Nature Reviews Drug Discovery* **7**, 463–464 (2008).
2. Butler, D. Crossing the valley of death. *Nature* **453**, 840–842 (2008).
3. Belton, V & Stewart, T. *Multiple criteria decision analysis: an integrated approach*. (Kluwer Academic Publishers, 2002).
4. Henriksen, a. D & Traynor, a. J. A practical R&D project-selection scoring tool. *IEEE Transactions on Engineering Management* **46**, 158–170 (1999).
5. Huang, C-C, Chu, P-Y & Chiang, Y-H. A fuzzy AHP application in government-sponsored R&D project selection. *Omega* **36**, 1038–1052 (2008).
6. Lahdelma, R, Miettinen, K & Salminen, P. Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). *European Journal of Operational Research* **147**, 117–127 (2003).
7. Lahdelma, R, Hokkanen, J & Salminen, P. SMAA - Stochastic multiobjective acceptability analysis. *European Journal of Operational Research* **106**, 137–143 (1998).
8. Lahdelma, R & Salminen, P. SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making. *Operations Research* **49**, 444–454 (2001).
9. van Dieren, S, Beulens, JWJ, van der Schouw, YT, Grobbee, DE & Neal, B. The global burden of diabetes and its complications: an emerging pandemic. *European Journal of Cardiovascular Prevention & Rehabilitation* **17 Suppl 1**, S3-8 (2010).
10. Fowler, MJ. Microvascular and Macrovascular Complications of Diabetes. *Clinical Diabetes* **26**, 77–82 (2008).
11. Cosh, E, Girling, A, Lilford, R, McAteer, H & Young, T. Investing in new medical technologies: A decision framework. *Journal of Commercial Biotechnology* **13**, 263–271 (2007).
12. Jacobs-van der Bruggen, M, Engelfriet, P, Bos, G, *et al. Opportunities for preventing diabetes and its cardiovascular complications: a modelling approach. RIVM Report 260801004/2007* (2007).
13. van Os, N, Niessen, LW, Bilo, HJ, Casparie, a F & van Hout, B a. Diabetes nephropathy in the Netherlands: a cost effectiveness analysis of national clinical guidelines. *Health Policy* **51**, 135–147 (2000).
14. Sawicki, PT. Screening for diabetes: hope and despair. *Diabetologia* **55**, 1568–1571 (2012).
15. Redekop, WK, Koopmanschap, MA, Rutten, GEHM, *et al.* Resource consumption and costs in Dutch patients with type 2 diabetes mellitus. Results from 29 general practices. *Diabetic Medicine* **19**, 246–253 (2002).

16. Alssema, M, Feskens, EJM, Bakker, SJL, *et al*. Finse vragenlijst redelijk goede voorspeller van het optreden van diabetes in Nederland. *Nederlands Tijdschrift voor Geneeskunde* **152**, 2418–2424 (2008).

17. Baltussen, R & Niessen, L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation* **4**, 14 (2006).

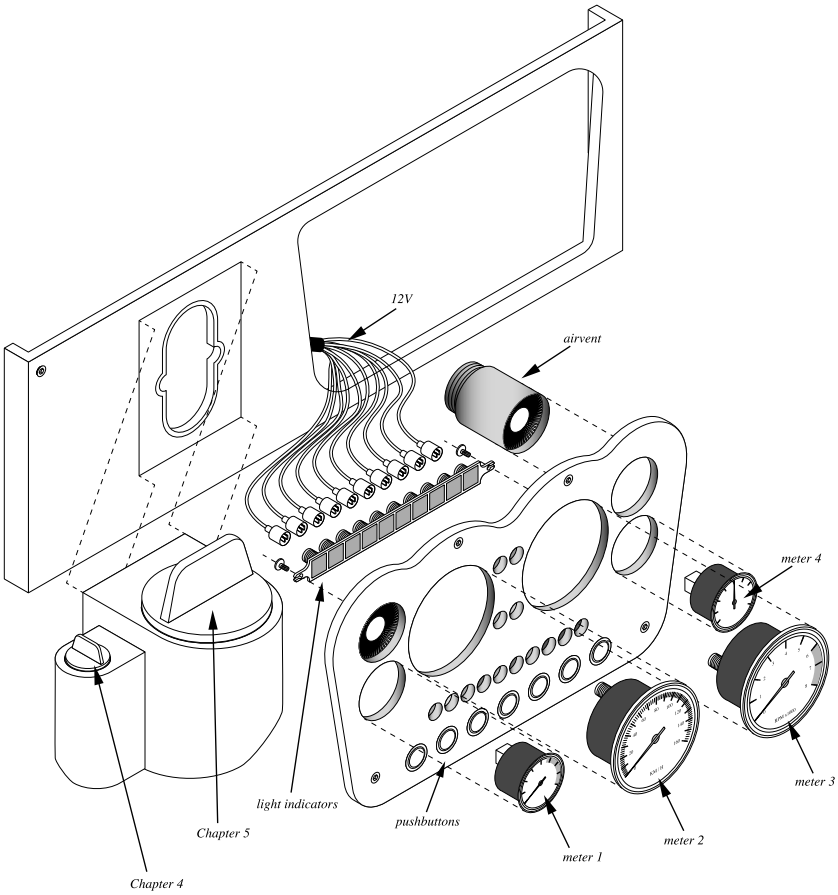
18. Postmus, D, De Graaf, G, Hillege, HL, Steyerberg, EW & Buskens, E. A method for the early health technology assessment of novel biomarker measurement in primary prevention programs. *Statistics in Medicine* **31**, 2733–2744 (2012).

19. Cao, Q, Postmus, D, Hillege, HL & Buskens, E. Probability elicitation to inform early health economic evaluations of new medical technologies: A case study in heart failure disease management. *Value in Health* **16**, 529–535 (2013).

20. Pietzsch, JB & Paté-Cornell, ME. Early technology assessment of new medical devices. *International Journal of Technology Assessment in Health Care* **24**, 36–44 (2008).

21. IJzerman, M & Steuten, LM. Early Assessment of Medical Technologies to Inform Product Development and Market Access: A Review of Methods and Applications. *Applied Health Economics and Health Policy* **9**, 331–347 (2011).

22. Tervonen, T, van Valkenhoef, G, Baştürk, N & Postmus, D. Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research* **224**, 552–559 (2013).


12V
airvent
meter 4
meter 3
meter 2
meter 1
light indicators
pushbuttons
Chapter 5
Chapter 4

Chapter 3

The early economic evaluation of novel biomarkers to accelerate their translation into clinical applications

Published as:
de Graaf, G., Postmus, D., Westerink, J., & Buskens, E.
The early economic evaluation of novel biomarkers to accelerate their translation into clinical applications. *Cost Effectiveness and Resource Allocation* **16**, 23 (2018)

ABSTRACT

BACKGROUND Translating prognostic and diagnostic biomarker candidates into clinical applications takes time, is very costly, and many candidates fail. It is therefore crucial to be able to select those biomarker candidates that have the highest chance of successfully being adopted in the clinic. This requires an early estimate of the potential clinical impact and commercial value. In this paper, we aim to demonstratively evaluate a set of novel biomarkers in terms of clinical impact and commercial value, using occurrence of cardiovascular disease (CVD) in type-2 diabetes (DM2) patients as a case study.

METHODS We defined a clinical application for the novel biomarkers, and subsequently used data from a large cohort study in the Netherlands in a modeling exercise to assess the potential clinical impact and headroom for the biomarkers.

RESULTS The most likely application of the biomarkers would be to identify DM2 patients with a low CVD risk and subsequently withhold statin treatment. As a result, one additional CVD event in every 75 patients may be expected. The expected downstream savings resulted in a headroom for a point-of-care device ranging from €119.09 at a willingness to accept of €0 for one additional CVD event, to €0 at a willingness to accept of €15,614 or more.

CONCLUSION It is feasible to evaluate novel biomarkers on outcomes directly relevant to technological development and clinical adoption. Importantly, this may be attained at the same point in time and using the same data as used for the evaluation of association with disease and predictive power.

INTRODUCTION

Within the current paradigm of personalized medicine or precision medicine, many research efforts are aimed at identifying novel biomarkers.[1–3] Although the expectations of improved clinical practice through better patient characterization remain high, it has long been recognized that the vast amount of biomarker research fails to live up to these expectations.[4–10] The fact that so few biomarkers are successfully translated from scientific discovery to clinical application entails a loss in health potential for patients and society. Moreover, resources from public and private investors allocated to research, development, and evaluation with the aim to improve patient outcomes appear wasted.

Biomarker discovery research has produced a vast body of literature on the association between biomarker and disease or outcome, and their diagnostic or prognostic performance (i.e. discrimination or reclassification).[3,11] While this is often regarded as the end-point of discovery research, it is only the start of the translational research phase. Herein, a candidate biomarker is developed into a diagnostic or prognostic technology and evidence required for its adoption in the clinic is generated.[8,12–15] Akin to clinical trials for pharmaceuticals, translational research is a long and complex trajectory requiring large financial investments, and will result in the rejection of a number of biomarker candidates.[13] Expert estimates of the costs of developing and commercializing a new biomarker based diagnostic technology exceed $100M.[16] As a result, the large number of candidate biomarkers that could be developed into clinical applications far exceeds the resources available to do so. It is therefore of great importance to identify those candidate biomarkers that have the highest chance to succeed as a commercial product. This requires an estimate of their potential clinical value and cost-effectiveness.[11,15,17] Unfortunately, currently employed methods for early biomarker evaluation provide little insight into clinical value.[8,9] On the other hand, proposed

methods for the assessment of clinical value are to extensive to be applied for the selection of biomarker candidates.[13–15,17]

The PREdiction and early diagnosis of DIabetes and diabetes-related Cardiovascular Complications (PREDICCt) project of the Center for Translational Molecular Medicine (CTMM) was initiated to enhance the possibilities for prevention of DM2 and associated complications through the development of molecular diagnostics and molecular imaging of novel biomarkers.[18] Its research efforts identified three novel biomarkers that were associated with incident CVD in DM2 patients: NT-proBNP, MMP-3, and Osteopontin. The association of these biomarkers with CVD incidence, as well as their predictive power within a prediction model have been described previously.[19] Whether further investments in translational research to develop diagnostic technologies based on these biomarkers is warranted has yet to be determined.

In this paper, we aim to demonstrate an evaluation framework for the assessment of novel biomarkers on clinical impact and commercial value (headroom). Such an assessment can be used to support the selection of biomarker candidates for further development and R&D investment decisions during development. We claim that this may be achieved at the same point in time and using the same data as used for the evaluation of predictive power or technical accuracy (i.e. data often available from discovery research). The CTMM PREDICCt project is used as a real-life case study to illustrate our framework. In our framework, we first define the application of the PREDICCt biomarkers in the clinical pathway and subsequently estimate the headroom of the markers in this application.

CLINICAL APPLICATION DEFINITION

Numerous publications on the translation of biomarkers stress the importance of defining a clinical application early in the discovery and development

process.[10,14,20] This is because the value of any diagnostic or prognostic test depends on the setting in which it is applied and the decision it is used to support. For many published biomarkers no clinical application has been specified, or this has been defined so broadly that it cannot possibly be used to determine their potential (cost-)effectiveness or commercial value. In our case study project, two very broad possible applications of the discovered biomarkers have been proposed. The first is to identify low-risk DM2 patients for whom treatment could be postponed, the second is to identify high risk DM2 patients for whom treatment could be initiated or intensified.[19] With respect to the economic value of the biomarkers it has been proposed that an individual patient risk-based approach has the apparent potential to allocate treatment resources more efficiently and effectively.[19]

To define a sufficiently detailed clinical application for the biomarkers, we gathered input from two clinical experts: an internist specialized in vascular medicine (third author on this publication), and the resident cardiologist that authored the publication describing the predictive power and possible clinical application of the biomarkers.[19] Under current international guidelines, DM2 patients are regarded as a high risk group for which the prescription of statins is advised.[21–24] In terms of risk, the so called high risk-group is defined by a 10-year risk of 10% or higher. Recent studies indicated that there is a wide distribution of CVD risk in the DM2 patient population.[25,26] Consequently, for part of the DM2 patient population the 10-year risk will likely fall below 10%, in which case these patients could be considered to be over-treated under current guidelines. This could potentially be remedied by using a more accurate risk prediction based on the newly discovered biomarkers. The second application of the PREDICCt biomarkers – to identify high risk patients and initiate or intensify treatment – is less likely to have a substantial clinical impact, due to the current clinical practice of CVD risk management in DM2 patients. As DM2 patients already fall in the highest risk category according to most guidelines, and given the limited options available for more intensive treatment, using the biomarkers as a risk stratification tool

to select very high-risk patients for intensified treatment is not a viable option. Apart from intensifying preventive treatment, high risk patients could also be screened for prevalent asymptomatic CVD. However, current guidelines also clearly recommend against this practice, as it does not improve outcomes in patients that already receive preventive treatment.[24]

HEADROOM ANALYSIS

In this section, we aim to evaluate the clinical impact and headroom of a risk stratification tool based on the three biomarkers identified in the PREDICCt project that is used to identify patients at low risk for CVD (10 year risk <10%) and subsequently withholding statin treatment in these patients. The headroom of a new technology is the maximum net incremental cost for which its intended clinical application is still likely to be cost effective.[27] We conducted a model-based evaluation using data from a large cohort study in the Netherlands. First, we developed a prediction model comprising the risk factors of the UKPDS risk engine[28] and the three novel biomarkers. Then, we estimated the impact of withholding treatment in those that fell below the risk cut-off using published data on the effectiveness of statins. Clinical impact was defined as the number of treatments withheld per additional CVD case. The headroom of the risk stratification tool was calculated for different levels of willingness to accept for one additional CVD event in the target population. The willingness to accept is the minimum monetary amount that the healthcare payer must save or receive in order to be willing to forgo a certain health benefit. As the current status quo is to provide the intervention to all patients, the new technology leads to reduced health benefits at lower costs. Thus, willingness to accept is an appropriate measure of preference, rather than the more ubiquitous willingness to pay, which applies when an additional benefit can be obtained at an additional cost.

Study population

We used patient level data from the Secondary Manifestations of ARTerial disease (SMART) study, a prospective cohort from the Netherlands. This study included patients that were referred to hospital with either manifest artherosclerotic disease or for the management of cardiovascular risk factors, such as hypertension, hyperlipidaemia, and DM2. A detailed description of the study design has been published previously.[29]

For the purpose of the current study, we selected patients with DM2 that had at least 5 years of follow-up and no prior history of CVD at the time of inclusion (n = 389). DM2 was defined as a referral diagnosis of DM2, self-reported DM2, the use of glucose-lowering agents, or a plasma glucose concentration of ≥7.0 mmol/L at baseline combined with the initiation of glucose-lowering treatment within 1 year after inclusion. Patients were considered to have a prior history of CVD when their medical records stated cerebrovascular disease (transient ischemic attack, cerebral infarction, cerebrovascular ischemia, amaurosis fugax, or retinal arterial occlusion), peripheral vascular disease, coronary artery disease, or an abdominal aortic aneurysm. The characteristics of the study population included in our analysis is shown in Table 1.

Risk assessment

The 10-year CVD risk (defined as the occurrence of myocardial infarction, stroke or vascular death) for each patient in the study population was calculated using an internally developed risk prediction model based on the Fine and Gray methodology.[30] This model consisted of the risk factors in the UKPDS risk engine (age at diagnosis of DM2, sex, current smoking, HbA1c, systolic blood pressure, and the total cholesterol/ HDL cholesterol ratio), and the three novel biomarkers. Missing values on these predictor variables in our dataset were dealt with using multiple imputation using the R-library MICE.[31]

CVD risk was then computed by taking the average of the risk values predicted from each of the imputed datasets.

Table 1: Study population characteristics of the 389 patients without prior cardiovascular disease history in the SMART cohort.

Parameter	Baseline value
Age (years, mean (SD))	54.8 (11.0)
Female sex (%)	39.8
Age at diagnosis of type-2 diabetes (years, mean (SD))	49.8 (11.6)
Currently smoking (%)	24.9
HbA1c (%, median (IQR))	7.4 (6.6 – 8.6)
Systolic blood pressure (mmHg, mean(SD))	145 (21)
Total cholesterol/ HDL cholesterol ratio (median (IQR))	4.6 (3.7 – 6.1)
NT-proBNP (pg/mL, median (IQR))	92 (44 – 216)
MMP-3 (ng/mL, median (IQR))	12.4 (8.1 – 17.3)
Osteopontin (ng/ml, median (IQR))	17.0 (13.3 – 21.9)

Effectiveness gap

We assumed that withholding statin treatment only has an impact on the incidence of CVD events and not on the non-CVD death rate. To estimate the clinical impact of this change in treatment policy, we fitted a competing risks model predicting the 10-year incidence of CVD events to the low-risk group. The model estimated cause-specific hazards for having a CVD event and for non-CVD death. These hazards were assumed to have a proportional hazard structure described by a Weibull distribution, and are described as follows:

$$h_{CVD}(t) = (\alpha_c/\beta_c)(t/\beta_c)^{(\alpha_c\text{-}1)} HR_{notreatment}$$

and

$$h_{nonCVDdeath}(t) = (\alpha_d/\beta_d)(t/\beta_d)^{(\alpha_d\text{-}1)}$$

where α_c (0.098) and β_c (4.879) are the shape and scale parameter of the Weibull distribution for CVD events, respectively, and α_d (0.362) and β_d (4.348) the shape and scale parameter of the Weibull distribution for non-

CVD death, respectively. Lastly, $HR_{notreatment}$ is the hazard ratio for the effect of withholding treatment. A large trial on the effects of statins in DM2 patients reported a hazard ratio of 0.76,[32] and in a meta-analysis of 14 randomized trials a relative risk of 0.79 per mmol/L reduction in LDL cholesterol was found[33]. We therefore assumed that the effect of withholding statin treatment in our target population would lead to a hazard ratio of 1.25 for CVD events.

The effectiveness gap was defined as the increase in 10-year CVD incidence resulting from withholding statin treatment in the low-risk group. For each treatment strategy (prescribing statins and withholding statins), these cumulative incidences were calculated as

$$I_{CVD}(t) = \int_0^t h_{CVD}(s)S(s)ds$$

where

$$S(t) = \exp[-\int_0^t h_{CVD}(s)ds - \int_0^t h_{nonCVDdeath}(s)ds]$$

is the overall survival function.

Headroom

The costs of statin treatment were based on the average cost of simvastatin 40 mg in The Netherlands and were estimated to be €0.06 per day.[34] As DM2 patients will have periodic checks with their general practitioner, as well as other prescription medication, costs for physician visits and prescription filling by pharmacies were assumed not to change when withholding statin treatment. The headroom of the point-of-care device was expressed as a function of the willingness to accept for one additional CVD event:

$$H(WTA) = f_{LR}(C_T\text{-}\Delta I_{CVD} \cdot WTA)$$

in which f_{LR} is the fraction of patients in the DM2 population with a CVD risk below 10%, ΔI_{CVD} is the change in CVD incidence as a result of withholding statin treatment, WTA is the willingness to accept for one additional CVD event, and C_T is the average per patient cost of statin treatment over the study horizon of 10 years. This was based on the average time patients DM2 patients are alive and did not experience a CVD event in our competing risk model, and defined as:

$$C_T = 365.25 \cdot €0.06(\int_0^{10} t[h_{CVD}(t) + h_{nonCVDdeath}(t)]\, S(t)dt + 10 \cdot S(10))$$

This willingness to accept was varied between €0 and the level at which the resulting headroom would be €0.

Sensitivity analysis

Apart from the willingness to accept, which was varied in the base case analysis, the headroom is to a large extend determined by the cost of treatment C_T and the effects of withholding statin treatment on CVD incidence ΔI_{CVD}. We therefore assessed the impact on the headroom of changes in the per diem cost of statin treatment and the hazard ratio for the effect of withholding statin treatment in the low-risk group. Per diem costs of statin treatment were €0.06 in the base case and were varied by 25% in the sensitivity analysis (€0.045 and €0.075). We assessed two alternative scenarios for the effects of withholding stating treatment in the low-risk group. First, we assumed that the relative effectiveness of statin treatment is related to baseline CVD risk, meaning that low-risk patients have a lower relative risk reduction as a result of statin treatment. This was implemented by using a hazard ratio for the effect of withholding statin treatment of 1.10, as opposed to 1.25 in the base case. In the second scenario we based the effects of statin treatment on a different study, which found a hazard ratio of 0.63 for the effect on CVD

incidence in DM2 patients.[35] This was implemented by using a hazard ratio for withholding treatment in the low-risk group of 1.58.

RESULTS

The low-risk group (10-year CVD risk <10%) thus identified consisted of 57.1% of the study population (Figure 1). A large difference in the observed 10-year incidence was found between the two risk groups, indicating that the risk assessment model had a high predictive power (Figure 1).

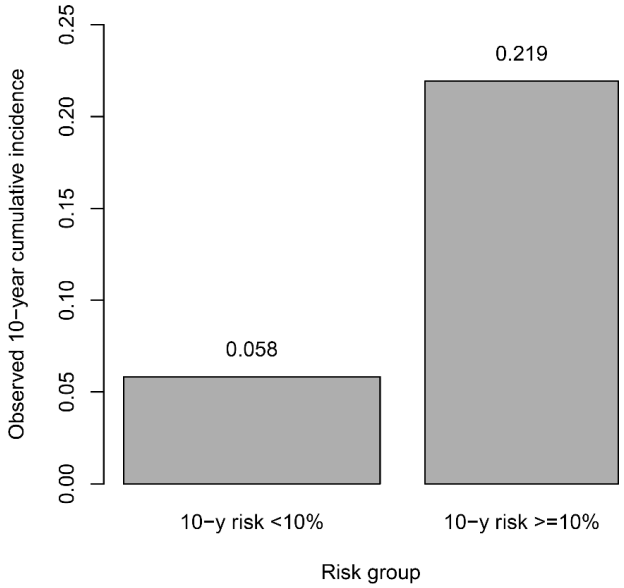


Figure 1: Difference in observed 10-year cumulative cardiovascular disease (CVD) incidence (bar height) and group size (bar width) between the low risk group (estimated 10-year CVD incidence < 10%) and the high risk group (estimated 10-year CVD incidence > = 10%) based on the risk prediction model consisting of age at diagnosis of DM2, sex, current smoking, HbA1c, systolic blood pressure, and the total cholesterol/HDL cholesterol ratio, NT-proBNP, MMP-3, and Osteopontin

The predicted and observed 10-year CVD incidences are shown in Figure 2. Withholding treatment in the low-risk group increased the predicted

cumulative CVD incidence at 10 years by approximately 0.0133. This means that withholding treatment will lead to one additional CVD event in every 75 patients.

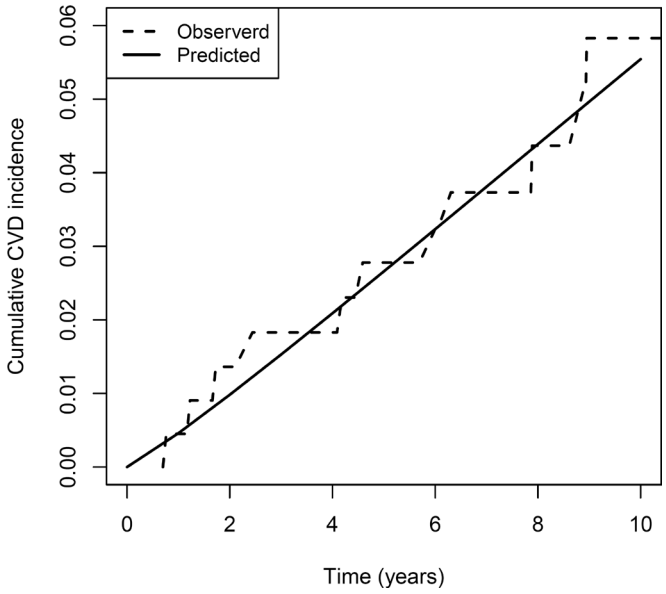


Figure 2: Comparison between the cumulative incidence of cardiovascular disease as predicted by the competing risk model and as observed in the SMART cohort

The average duration of treatment in the high risk group was estimated to be 9.52 years. This lead to an estimated total average treatment cost over 10 years of €208.67. The headroom of a point-of-care device using the novel biomarkers was found to be €119.09 at a willingness to accept of €0 (that is, no savings or monetary gain would be required to accept an additional CVD event). The headroom became less than €0 when the willingness to accept for one additional CVD event exceeded €15,614 (that is, an additional CVD event is accepted when a cost saving of more than €15,614 is realized).

The results of the sensitivity analysis are shown in Table 2. Varying the treatment effect of statins did not have an impact on the maximum headroom

but did impact the willingness to accept level at which the headroom becomes €0 (which increased when statin effects were less). Changes in the cost of statin treatment were reflected in the total cost of treatment and thereby had an impact on the maximum headroom (higher statin costs led to a higher headroom).

Table 2: Results of the sensitivity analysis.

Outcome	Base case	Lesser effect of statins (HR 1.10)	Larger effect of statins (HR 1.58)	Statin cost +25%	Statin cost -25%
Additional CVD incidence	0.0134	0.0054	0.0307	0.0134	0.0134
Number needed to withhold	75	186	33	75	75
Total average cost of treatment	€208.67	€208.67	€208.67	€260.83	€156.50
Headroom at WTA = €0	€119.09	€119.09	€119.09	€148.86	€89.31
WTA at which headroom = €0	€15,614	€38,867	€6,795	€19,518	€11,711

DISCUSSION

In this study, we demonstrated that an early assessment of the clinical impact and commercial value (headroom) of novel biomarkers can be performed at the same time and using the same data as used to determine predictive power and accuracy. We used a case study of biomarkers for additional CVD risk stratification in DM2 patients, more specifically a setting where such biomarkers would be used as a prognostic test to inform the decision on withholding statin treatment from low-risk patients. We found that withholding statin treatment in DM2 patients with a 10 year CVD risk of <10% lead to an additional CVD event in every 75 patients for which treatment would be withheld. Furthermore, we found the headroom to be €119.09 in the optimal scenario from the industry perspective (that is, when no savings

would be required in order to accept an additional CVD event). The headroom reduced to €0 when the willingness to accept would be €15,614 or more. When a larger cost saving is demanded for an additional CVD case (that is, there is a higher willingness to accept), a smaller part of the costs saved by withholding treatment is available to pay for the biomarker test. Headroom thus decreases as the willingness to accept increases. The willingness to accept at which the headroom is reduced to €0 was sensitive to changes in both the effect of statin treatment in the low-risk group, as well as the cost of statin treatment (lesser treatment effect and higher statin cost led to a higher willingness to accept at which headroom is €0). The maximum headroom was only sensitive to the cost of statin treatment (increased cost of statins led to a higher maximum headroom).

Our study is the first that estimates the clinical impact and commercial value of biomarkers for the estimation of CVD risk in DM2 patients, and one of the first to perform such an analysis for a biomarker technology before it is actually developed. A large body of literature exists demonstrating the predictive power and strength of association between biomarker and disease for many different types of biomarkers. Based on such results, there is often a positive and hopeful attitude towards novel biomarkers. These outcome measures, however, have little relation to the clinical, commercial, or economic value of a biomarker technology.[11,17] Notably, it is not uncommon for a biomarker to be developed without a clear clinical implementation in mind. Without a clinical application definition, any assessment of clinical value or cost-effectiveness is impossible. Such evidence is crucial for the adoption of a new biomarker technology in the clinic and by extension thereof its commercial success. As a result, many novel biomarkers fail to deliver on the high hopes that have been placed on them, and represent a waste of public and private research funds. Existing methods for the economic evaluation of biomarkers (and other healthcare innovations) such as early health economic modeling require more data, are computationally more complex, and as a result demand more time and financial resources to implement.[15,36] Assessing multiple biomarker

candidates, each with multiple possible applications, is often not feasible using those methods. Our less resource-demanding method employing data from biomarker discovery research and published literature in a computationally uncomplicated approach can provide relevant support in decision making.

The methods we employ are not completely novel. A number of methodological studies have dealt with the issue of biomarker assessment, some of which focus on the statistical aspects of such an assessment,[37–40] while others describe assessment in a broader scope, including decisions on area of application and current care comparators.[27,41,42] Our main goal was to demonstrate the applicability of such methods in a real-life setting of biomedical development. Likewise, a few recent studies demonstrated the potential for using health economic modeling as an alternative for RCTs to generate evidence on the cost-effectiveness of diagnostic tests.[43,44] In several ways these studies have used an approach similar to ours. The main difference being that our method is aimed at an earlier stage of development – immediately after discovery – where most biomarkers are falling out of the translational process. It thereby aims to primarily inform decisions on the direction of development and investment, rather than adoption in the clinic.

Our case-study outcomes are difficult to compare to outcomes of other studies. Most economic evaluations use cost per Quality Adjusted Life-Year (QALY) as their primary outcome and determine cost-effectiveness by specifying a willingness to pay for an additional QALY. Accurately estimating the loss of QALYs as a result of withholding treatment would require a disease progression model, which is beyond the scope of this showcase research. Moreover, the applicability of QALYs as an outcome measure in modeling studies for diagnostic test has previously been questioned.[17] A further issue regarding comparability with previous research is the fact that the willingness to accept is a concept not often encountered in health economic evaluations. A threshold for willingness to accept an additional CVD event has never been specified. However, even in the absence of a relevant threshold the outcomes

of our method can be informative for R&D and investment decisions. When a large headroom exists even when extremely unfavorable (i.e. low in the case of willingness to pay, high in the case of willingness to accept) threshold values are used in the analysis, further investments in the development of the new technology are certainly warranted from an economic perspective. When no or a very small headroom exists when favorable threshold levels are used, it is unlikely that the new technology will ever be cost-effective when used in the evaluated application, and therefore it would not be wise to invest in further development. By this token, due to the high costs and burden associated with cardiovascular events such as myocardial infarction and stroke, it would appear unlikely that the willingness to accept for an additional CVD case will be sufficiently low to ever make a risk stratification tool in DM2 patients like the one analyzed in our case study a viable strategy. A threshold defined in willingness to accept is rare because most new interventions provide increased health outcomes at an additional cost. However, as many societies are increasingly concerned by the sustainability of healthcare expenditures, we believe that it will become increasingly important to be able to express the willingness to forgo health benefits in return for cost reductions. These limitations notwithstanding, we believe that we have demonstrated that without using other evidence than datasets used for biomarker discovery and published literature, it is possible to go beyond the usual evaluation of biomarkers on association with disease and predictive power and additionally give an insight in potential clinical impact and commercial value.

REFERENCES

1. Ludwig, J & Weinstein, J. Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer* **5**, 845–856 (2005).
2. Nordström, A & Lewensohn, R. Metabolomics: Moving to the clinic. *Journal of Neuroimmune Pharmacology* **5**, 4–17 (2010).
3. Moons, KGM. Criteria for scientific evaluation of novel markers: A perspective. *Clinical Chemistry* **56**, 537–541 (2010).
4. Lumbreras, B, Parker, LA, Porta, M, *et al*. Overinterpretation of clinical applicability in molecular diagnostic research. *Clinical Chemistry* **55**, 786–794 (2009).
5. Zolg, W. The proteomic search for diagnostic biomarkers: lost in translation? *Molecular & Cellular Proteomics* **5**, 1720–1726 (2006).
6. Ioannidis, JPA & Panagiotou, OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA: The Journal of the American Medical Association* **305**, 2200–2210 (2011).
7. Ioannidis, JPA. Expectations, validity, and reality in omics. *Journal of Clinical Epidemiology* **63**, 945–949 (2010).
8. Ioannidis, JPA. Is molecular profiling ready for use in clinical decision making? *The Oncologist* **12**, 301–311 (2007).
9. Vickers, AJ, Jang, K, Sargent, D, Lilja, H & Kattan, MW. Systenatic Review of Statistical Methods Used in Molecular Marker Studies in Cancer. *Cancer* **112**, 1862–1868 (2008).
10. Williams, PM, Lively, TG, Jessup, JM & Conley, BA. Bridging the Gap: Moving Predictive and Prognostic Assays from Research to Clinical Use. *Clinical Cancer Research* **18**, 1531–1539 (2012).
11. Bossuyt, PMM, Reitsma, JB, Linnet, K & Moons, KGM. Beyond diagnostic accuracy: The clinical utility of diagnostic tests. *Clinical Chemistry* **58**, 1636–1643 (2012).
12. Pepe, MS, Etzioni, R, Feng, Z, *et al*. Phases of Biomarker Development for Early Detection of Cancer. *Cancer* **93**, 1054–1061 (2001).
13. Sackett, D & Haynes, R. The architecture of diagnostic research. *BMJ* **324**, 539–541 (2002).
14. Van den Bruel, A, Cleemput, I, Aertgeerts, B, Ramaekers, D & Buntinx, F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *Journal of Clinical Epidemiology* **60**, 1116–1122 (2007).
15. Horvath, AR, Lord, SJ, StJohn, A, *et al*. From biomarkers to medical tests: The changing landscape of test evaluation. *Clinica Chimica Acta* **427**, 49–57 (2014).

16. Dolginow, D, Tynan, K, Doheney, N & Keeling, P. *Mystery Solved! What is the cost to develop and launch a Diagnostic?* (2017).

17. Durtschi, A & Jülicher, P. Assessing the value of cardiac biomarkers: going beyond diagnostic accuracy ? *Future Cardiology* **10**, 367–380 (2014).

18. CTMM Predicct. *Biomarkers for the Prediction and Early Diagnosis of Diabetes and Diabetes-related Cardiovascular Complications: Output Report*. (2015).

19. van der Leeuw, J, Beulens, JWJ, van Dieren, S, *et al*. Novel Biomarkers to Improve the Prediction of Cardiovascular Event Risk in Type 2 Diabetes Mellitus. *Journal of the American Heart Association* **5**, e003048 (2016).

20. Moynihan, R, Henry, D & Moons, KGM. Using Evidence to Combat Overdiagnosis and Overtreatment: Evaluating Treatments, Tests, and Disease Definitions in the Time of Too Much. *PLoS Medicine* **11**, 11–13 (2014).

21. Piepoli, MF, Hoes, AW, Agewall, S, *et al*. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *European Heart Journal* **37**, 2315–2381 (2016).

22. Stone, NJ, Robinson, JG, Lichtenstein, AH, *et al*. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the american college of cardiology/american heart association task force on practice guidelines. *Circulation* **129**, (2014).

23. Fox, CS, Golden, SH, Anderson, C, *et al*. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: A scientific statement from the American Heart Association and the American diabetes association. *Diabetes Care* **38**, 1777–1803 (2015).

24. American Diabetes Association (ADA). Standard of medical care in diabetes - 2017. *Diabetes Care* **40** (**sup 1**), s4–s128 (2017).

25. Paynter, N, Mazer, N, Pradhan, A, *et al*. Cardiovascular risk prediction in diabetic men and women using hemoglobin A1c vs diabetes as a high-risk equivalent. *Archives of Internal Medicine* **171**, 1712–1718 (2011).

26. Kaasenbrood, L, Boekholdt, SM, Van Der Graaf, Y, *et al*. Distribution of Estimated 10-Year Risk of Recurrent Vascular Events and Residual Risk in a Secondary Prevention Population. *Circulation* **134**, 1419–1429 (2016).

27. Cosh, E, Girling, A, Lilford, R, McAteer, H & Young, T. Investing in new medical technologies: A decision framework. *Journal of Commercial Biotechnology* **13**, 263–271 (2007).

28. Stevens, RJ, Kothari, V, Adler, a I & Stratton, IM. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). *Clinical Science* **101**, 671–679 (2001).

29. Simons, PCG, Algra, A, Van De Laak, MF, Grobbee, DE & Van Der Graaf, Y. Second manifestations of ARTerial disease (SMART) study: Rationale and design. *European Journal of Epidemiology* **15**, 773–781 (1999).

30. Fine, JP & Gray, RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* **94**, 496–509 (1999).

31. Van Buuren, S & Groothuis-Oudshoorn, K. MICE: Multivariate Imputation by Chained Equations in R. *Journal Of Statistical Software* **45**, 1–67 (2011).

32. Protection, H & Collaborative, S. MRC/BHF Heart Protection Study of cholesterol lowering simvastatin in 5963 people with diabetes: a randomized controlled trial. Heart Protection Study Collaborative Group. *The Lancet* **361**, 2005–2016 (2003).

33. Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy of cholesterol-lowering therapy in 18 686 people with diabetes in 14 randomised trials of statins: a meta-analysis. *The Lancet* **371**, 117–125 (2008).

34. Zorginstituut Nederland. Medicijnkosten. (2016). Available at: http://www.medicijnkosten.nl. (Accessed: 28th November 2016)

35. Colhoun, HM, Betteridge, DJ, Durrington, PN, *et al*. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. *Lancet* **364**, 685–696 (2004).

36. Annemans, L, Genesté, B & Jolain, B. Early modelling for assessing health and economic outcomes of drug therapy. *Value in Health* **3**, 427–434 (2000).

37. Pencina, M, D'Agostino Sr, R, D'Agostino Jr, R & Vasan, R. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine* **27**, 157–172 (2008).

38. Pencina, MJ, Steyerberg, EW & D'Agostino, RB. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine* **30**, 11–21 (2011).

39. Postmus, D, De Graaf, G, Hillege, HL, Steyerberg, EW & Buskens, E. A method for the early health technology assessment of novel biomarker measurement in primary prevention programs. *Statistics in Medicine* **31**, 2733–2744 (2012).

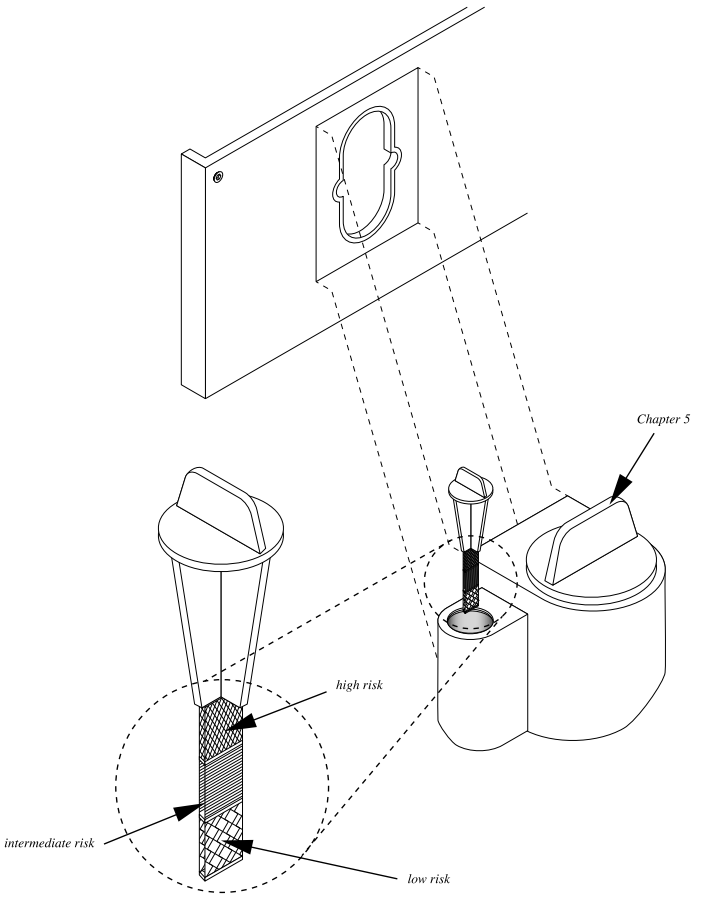
40. Steyerberg, E, Vickers, A, Cook, NR, *et al*. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology* **21**, 128–138 (2010).

41. D'Agostino, RB. Risk prediction and finding new independent prognostic factors. *Journal of Hypertension* **24**, 643–645 (2006).

42. IJzerman, M & Steuten, LM. Early Assessment of Medical Technologies to Inform Product Development and Market Access: A Review of Methods and Applications. *Applied Health Economics and Health Policy* **9**, 331–347 (2011).

43. Koffijberg, H, van Zaane, B & Moons, KGM. From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Medical Research Methodology* **13**, 12 (2013).

44. Gerke, O, Poulsen, MH & Hoilund-Carlsen, PF. Added value of cost-utility analysis in simple diagnostic studies of accuracy:18F-fluoromethylcholine PET/CT in prostate cancer staging. *American Journal of Nuclear Medicine and Molecular Imaging* **5**, 183–194 (2015).


Chapter 5
high risk
intermediate risk
low risk

Chapter 4

A method for the early health technology assessment of novel biomarker measurement in primary prevention programs

Published as:
Postmus, D., de Graaf, G., Hillege, H. L., Steyerberg, E. W., & Buskens, E.
A method for the early health technology assessment of novel biomarker measurement in primary prevention programs. *Statistics in Medicine* **31**, 2733-2744 (2012).

ABSTRACT

Many promising biomarkers for stratifying individuals at risk of developing a chronic disease or subsequent complications have been identified. Research into the potential cost-effectiveness of applying these biomarkers in actual clinical settings has however been lacking. Investors and analysts may improve their venture decision making should they have indicative estimates of the potential costs and effects associated with a new biomarker technology already at the early stages of its development. To assist in obtaining such estimates, this paper presents a general method for the early health technology assessment of a novel biomarker technology. The setting considered is that of primary prevention programs where initial screening to select high-risk individuals eligible for a subsequent intervention occurs, e.g., prevention of type 2 diabetes. The method is based on quantifying the health outcomes and downstream health-care consumption of all individuals who get reclassified as a result of moving from a screening variant based on traditional risk factors to a screening variant based on traditional risk factors plus a novel biomarker. As these individuals form well-defined subpopulations, a combination of disease progression modeling and sensitivity analysis can be used to perform an initial assessment of the maximum increase in screening cost for which the use of the new biomarker technology is still likely to be cost-effective.

INTRODUCTION

Much research effort is currently directed at discovering novel biomarkers for identifying individuals at risk of developing a chronic disease (primary prevention) or subsequent complications (tertiary prevention). As these biomarkers provide additional information beyond standard clinical risk factors, applying them in actual clinical settings is expected to result in improved risk stratification. This, in turn, may help to optimize the selection of individuals eligible for a focused intervention, such as behavioral counseling or chemoprevention. Ultimately, this should improve the population's health outcomes at affordable (possibly lower) costs.

After a promising biomarker has been identified and a (prototype) technology has been developed to measure this biomarker in actual clinical settings, its performance needs to be critically evaluated before the new biomarker technology will eventually be adopted in clinical practice. According to Hlatky et al.[1], such a critical assessment involves six phases, ranging from showing that the levels of the novel biomarker differ between individuals with and without the outcome of interest (proof of concept) to assessing whether using the biomarker improves health outcomes at an affordable cost (cost-effectiveness). In this traditional framework, cost-effectiveness analysis is conducted at the end of the product development process. The results are intended to assist health policy makers in deciding whether the new biomarker technology should be routinely adopted in clinical practice. This form of health technology assessment (HTA) is what Pietzsch and Paté-Cornell[2] refer to as classical HTA, to be distinguished from an initial assessment of the likely costs and effects associated with a new medical technology at the early stages of its development process. This so-called early HTA is conducted before the technology has been fully developed and serves to support health technology firms in making appropriate product investment decisions.

Although appropriate quantification of the added predicted value of a novel biomarker over conventional risk factors is a problem of active research and debate[3–9], research into the cost-effectiveness of applying such a biomarker in actual clinical settings has so far been limited to two recent case studies in the context of prioritizing patients waiting for coronary artery surgery.[10,11] A more in-depth discussion on how the cost-effectiveness of using prognostic biomarkers could be established seems therefore desirable. To that end, this paper presents a general method for the early HTA of a novel biomarker technology that is used, in combination with a set of conventional risk factors, as an initial screening test to select high-risk individuals eligible for a subsequent preventive intervention. The use of the method is illustrated in a case study related to the prevention of type 2 diabetes.

Added predictive value and cost-effectiveness of novel biomarker measurement in primary prevention programs

IJzerman and Steuten have recently provided a conceptual model of the medical technology development process.[12] According to their model, this process consists of four main stages: (i) basic research, (ii) translational research, (iii) clinical research, and (iv) market access. Preceding each of these stages is a decision gate where it has to be decided whether to proceed with the next stage, and if so in what direction. In this paper, we assume that the basic research on biological mechanisms is completed and that this has resulted in the identification of several candidate biomarkers. We are therefore at the decision gate preceding the translational research phase, where it has to be decided which of these biomarkers should be selected for further development, if any. The purpose of performing early HTA at this stage of the product development process is to assist health technology firms in making realistic commercial valuations of the conceived new products by providing for each potential new biomarker technology a rough estimate of the maximum additional cost for which its intended clinical application is still likely to be cost-effective. In this section, we will describe how this upper

bound on the technology's cost, also known as the commercial headroom available[13], can be estimated for an improved, biomarker-based screening test.

Consider N individuals who participate in a primary prevention program. Based on the results of an initial screening, individuals are classified into m ordinal risk categories, such as low, intermediate, and high risk in case of three categories. Those who are considered to be at risk are offered a subsequent intervention, which may be tailored to the risk category an individual is classified into (e.g., no intervention in individuals who are being classified as low risk, a non-invasive and relatively safe intervention in individuals who are being classified as intermediate risk, and an invasive and more risky intervention in individuals who are being classified as high risk). Suppose that a decision maker can choose between two variants of the risk stratification model: screening variant s^1 in which the risk stratification is based on a vector of cut-off points $\gamma^1 = (\gamma_1^1, \dots, \gamma_{m\text{-}1}^1)'$ on a risk score consisting of conventional risk factors, and screening variant s^2 in which the risk stratification is based on a vector of cut-off points $\gamma^2 = (\gamma_1^2, \dots, \gamma_{m\text{-}1}^2)'$ on a risk score comprising the same conventional risk factors as well as a novel biomarker. For clinically meaningful values of the cut-off points γ^1 and γ^2, consider the m by m reclassification table that results from combining the risk classifications obtained under s^1 and s^2 (Table 1). Let N_{kl} denote the number of individuals within the klth entry of the reclassification table (i.e., all individuals who get classified in the kth risk category under s^1 and in the lth risk category under s^2). It then follows that a fraction of $\frac{\sum_{k=1}^{m} \sum_{l \neq k} N_{kl}}{N}$ of the individuals are reclassified when applying s^2 instead of s^1.

Table 1: The reclassification table that results from combining the risk classifications under s^1 and s^2.

		Classification under s^2			
		risk category 1	risk category 2	⋯	risk category m
Classification under s^1	risk category 1	N_{11}	N_{12}	⋯	N_{1m}
	risk category 2	N_{21}	N_{22}	⋯	N_{2m}
	⋮	⋮	⋮	⋱	⋮
	risk category m	N_{m1}	N_{m2}	⋯	N_{mm}

The extent to which this reclassification can be considered an improvement can be determined in several ways.[14] A measure of reclassification that has nowadays gained wide-spread acceptance is the net reclassification index (NRI).[7] The main idea behind this measure is to consider the reclassification of individuals who develop and who do not develop the event of interest separately. Moving from s^1 to s^2 can then be considered an improvement when the proportion of subjects who move upward towards a higher risk category is larger than the proportion of subjects who move downward towards a lower risk category for individuals with the event and when the opposite holds for individuals without the event. To assess this in a formal way, consider a random sample of size n from the screening population, and let n_{kl}^{event} and $n_{kl}^{\text{no event}}$ denote the number of events and non-events within the klth cell of the reclassification table corresponding to this sample, respectively. The NRI is then computed as

$$\frac{\sum_{k=1}^{m}\sum_{l>k} n_{kl}^{\text{event}} - \sum_{k=1}^{m}\sum_{l<k} n_{kl}^{\text{event}}}{n^{\text{event}}} + \frac{\sum_{k=1}^{m}\sum_{l<k} n_{kl}^{\text{no event}} - \sum_{k=1}^{m}\sum_{l>k} n_{kl}^{\text{no event}}}{n^{\text{no event}}},$$

where n^{event} and $n^{\text{no event}}$ denote the total number of events and non-events in the total sample, respectively. The novel biomarker is then considered to have incremental predictive value over the conventional risk factors if the null hypothesis of NRI = 0 is rejected.

Although the NRI and other proposed measures of reclassification are useful for establishing the added predicted ability of a novel biomarker, they do not directly provide insight into which of the risk stratification models would be preferable from a societal perspective. To address this latter aspect, we have to determine whether the increase in added predictive value is sufficient to make changing from s^1 to s^2 an efficient allocation of scarce health care resources, and this is the domain of health economic analysis.[15] In this type of analysis, it is common practice to assume that all relevant health effects can be aggregated into a single measure of effectiveness. The net monetary benefit (NMB) of an intervention can then be calculated by assuming a threshold value of the decision maker's willingness-to-pay for one unit of health gain.[15,16] The most common measure of effectiveness is the quality-adjusted life year (QALY), and this is also the one that will be used in the case study. In the remainder of this paper, we will therefore assume that the effectiveness of the two screening variants is evaluated in terms of QALYs. The equations derived in this section are however applicable in all situations where the health consequences are captured in terms of a single measure of effectiveness.

To determine the relative merits of the two screening variants, let c^i and e^i denote the average cost and QALYs associated with screening variant s^i, and let λ denote the willingness-to-pay (in terms of monetary units) for a QALY. Screening variant s^2 is then preferred over screening variant s^1 if

$$\text{NMB}^2\text{-NMB}^1 \geq 0, \tag{1}$$

where $\text{NMB}^i = \lambda e^i\text{-}c^i$ represents the average NMB associated with screening variant s^i. To use Equation (1) to compute the headroom available to the improved, biomarker-based screening test, let the treatment assignment indicator t^i_{kl} return the treatment that is assigned to the individuals in the klth entry of the reclassification table under screening variant s^i. For example, if under s^1 treatment A is offered to all individuals who are classified into risk

category 2, t_{2l}^1 = treatment A, $\forall l \in \{1, \ldots, m\}$. The average cost and QALYs associated with screening variant s^i can then be written as

$$c^i = c_{\text{scr}}^i + \frac{1}{N}\sum_{k=1}^{m}\sum_{l=1}^{m} N_{kl} c_{kl}^{t_{kl}^i}, \tag{2}$$

$$e^i = \frac{1}{N}\sum_{k=1}^{m}\sum_{l=1}^{m} N_{kl} e_{kl}^{t_{kl}^i}, \tag{3}$$

where $c_{kl}^{t_{kl}^i}$ and $e_{kl}^{t_{kl}^i}$ denote the average cost and QALYs associated with applying treatment t_{kl}^i to the individuals in the klth entry of the reclassification table, and where c_{scr}^i denotes the average screening cost under screening variant s^i. If we now define $f_{kl} = N_{kl}/N$ and $\Delta_{\text{scr}} = c_{\text{scr}}^2 \text{-} c_{\text{scr}}^1$, it follows by substituting (2) and (3) into (1) that screening variant s^2 is preferred over screening variant s^1 if

$$\Delta_{\text{scr}} \leq \sum_{k=1}^{m}\sum_{j\neq k} f_{kl}\left[\lambda(e_{kl}^{t_{kl}^2} \text{-} e_{kl}^{t_{kl}^1}) \text{-} (c_{kl}^{t_{kl}^2} \text{-} c_{kl}^{t_{kl}^1})\right]. \tag{4}$$

Individuals who end up at one of the diagonal entries of the reclassification table are assigned to the same treatment under both s^1 and s^2. Consequently, it is not required to consider these individuals' QALYs and downstream health-care consumption when choosing between the two screening variants. In Equation (4), this is reflected by the fact that the average beneficial and/or harmful consequences (in terms of downstream NMB) associated with switching from s^1 to s^2, i.e. $\lambda(e_{kl}^{t_{kl}^2} \text{-} e_{kl}^{t_{kl}^1}) \text{-} (c_{kl}^{t_{kl}^2} \text{-} c_{kl}^{t_{kl}^1})$, are only computed for those individuals who move upwards or downwards in risk classification. For the biomarker-based screening variant to be cost-effective, the overall increase in downstream NMB (i.e., the right-hand side of Equation (4)) needs to be sufficiently large to offset the upfront increase in screening cost, which are incurred by all individuals, irrespective of whether they are reclassified.

PARAMETER ESTIMATION

In this section, we will describe how the parameters at the right-hand side of Equation (4) can be estimated at the decision gate preceding the translational research stage of the medical technology development process. As the amount of clinical data available for estimating these parameters very much depends on whether the biomarker in question has already been measured in a prospective cohort study, we will make a distinction between technologies that aim at providing an alternative way of measuring an existing biomarker and technologies that aim at measuring a completely new biomarker.

General considerations

As the initiation of a preventive intervention is expected to have cost and effect implications on the remainder of a patient's live, the appropriate time horizon for the economic evaluation of such interventions is the patient's lifetime. [15] In such situations, health economic analysts generally rely on disease progression modeling to extrapolate from the event rates and treatment effects observed in clinical trials and observational studies to what would be expected to happen over a lifetime period.[16] The quantitative models used for this purpose typically consist of several discrete health states reflecting the occurrence of the events of interest and a set of transition intensities (or transition probabilities in case of a discrete-time model) that govern the movement between these health states. The expected long-term cost and effect consequences of an intervention can then be estimated by multiplying the average sojourn time in each of the model's health states by a cost and utility weight attached to these health states. To include patient heterogeneity into the model, the logarithms of the transition intensities are sometimes expressed as linear functions of a set of explanatory covariates, resulting in a so-called patient-level model. Disease progression models that do not take into account patient heterogeneity are generally referred to as cohort models.[17]

For the individuals in the klth cell of the reclassification table, the expected cost and QALY consequences of moving from s^1 to s^2 will depend on three main aspects: (i) the cumulative disease incidence $I_{kl}(\tau)$ as a function of the time since screening τ that would be observed in this population in the absence of screening, (ii) the reduction in cumulative disease incidence due to t^1_{kl}, and (iii) the relative risk of t^2_{kl} relative to t^1_{kl}. Our strategy is to derive the health economic consequences that result from these changes in the cumulative disease incidence through disease progression modeling. As the individuals from the different cells of the reclassification table form well-defined subpopulations, we propose to fit separate disease progression models to each of these subpopulations. In particular, let the vector α_{kl} denote the model parameters that apply to subpopulation kl. The disease progression models can then be expressed as functions $h(t^i_{kl}|\alpha_{kl})$ that map the administered treatment to the expected values of $c^{t^i_{kl}}_{kl}$ and $e^{t^i_{kl}}_{kl}$.

Prospective data available

Many companies in the medical device industry do not only focus on developing novel equipment for measuring promising new biomarkers but also on finding alternative (e.g., more efficient, less invasive, or less risky) ways of measuring an existing biomarker, such as a multiplex ELISA that can simultaneously measure a whole panel of biomarkers. In such situations, it may already be possible to evaluate the added predictive value of the selected (panel of) biomarker(s) over a set of conventional risk factors by applying the currently available measurement techniques to blood, urine, or tissue samples collected in an existing cohort study. Base-case values of the fractions f_{kl} can then readily be derived from a reclassification table that is constructed from the data collected in this study. The same applies to all parameters in α_{kl} that directly depend on the incremental predicted value of the considered biomarker(s).

Prospective data not available

When dealing with a completely new biomarker, nothing will yet be known about the performance of this biomarker in actual clinical settings. Initial values of f_{kl} must then be derived from surrogate data, such as early bench and animal testing, the performance of related but already clinically validated biomarkers, or expert judgment. A similar problem is encountered when specifying the parameters of the disease progression model: although it may be possible to obtain some of the parameter values from previously conducted economic evaluations, such as the costs and utilities attached to the model's transient health states, others depend on the incremental predictive value of the new biomarker and must therefore also be derived from indirect sources. Probability Aggregation for Medical Device Assessment is particularly suited for synthesizing evidence from multiple indirect sources, such as the results from several pilot studies in different types of animal model.[2] For a thorough discussion on how expert knowledge can be elicited and incorporated in a probabilistic way, the reader is referred to Garthwaite et al.[18]

INITIAL ECONOMIC EVALUATION

After the base-case values of all parameters have been specified, the commercial headroom available to the new biomarker technology, denoted by $\Delta_{\text{scr}}^{\max}$, can be estimated by applying the algorithm depicted in Figure 1. As the values of most parameters are still uncertain at the early stages of the medical technology development process, the base-case analysis should be followed by an extensive amount of sensitivity analysis to determine the robustness of the obtained results with respect to changes in the parameter values. How the sensitivity analysis can best be conducted depends on the amount of clinical data available.[19] If the added predictive value of the considered biomarker has already been evaluated in a prospective cohort study, the use of probabilistic sensitivity analysis (PSA) seems most appropriate as probability distributions of the parameters of interest can then directly be derived from the data

collected within this study. On the other hand, unless expert knowledge has been elicited in a probabilistic way, the use of PSA is generally not feasible when the novel biomarker has not yet been measured in a prospective cohort study. The use of a deterministic approach, such as 1-way sensitivity analysis, would then be more appropriate.

Input: $\lambda, f_{kl}, \mathbf{x}_{kl}, \alpha_{kl}, t_{kl}^1, t_{kl}^2, \;\; k,l = 1,\ldots,m, \; k \neq l$
Output: $\Delta_{\text{scr}}^{\max}$

1: **for** $k \leftarrow 1$ **to** m **do**
2: **for** $l \leftarrow 1$ **to** m **do**
3: **if** $k \neq l$ **then**
4: $(c_{kl}^{t_{kl}^1}, e_{kl}^{t_{kl}^1}) \leftarrow h(\mathbf{x}_{kl}, t_{kl}^1 | \alpha_{kl})$
5: $(c_{kl}^{t_{kl}^2}, e_{kl}^{t_{kl}^2}) \leftarrow h(\mathbf{x}_{kl}, t_{kl}^2 | \alpha_{kl})$
6: **end if**
7: **end for**
8: **end for**
9: $\Delta_{\text{scr}}^{\max} \leftarrow \sum_{k=1}^{m} \sum_{j \neq k} f_{kl} \left[\lambda (e_{kl}^{t_{kl}^2} - e_{kl}^{t_{kl}^1}) - (c_{kl}^{t_{kl}^2} - c_{kl}^{t_{kl}^1}) \right]$

Figure 1: Algorithm for estimating $\Delta_{\text{scr}}^{\max}$ for a specific set of parameter values.

In a PSA, the uncertainty in one or more input parameters is propagated to uncertainty in the outcome variable by repeatedly calculating $\Delta_{\text{scr}}^{\max}$ for different samples from the (joint) distribution of the input parameters. As the parameters of interest are treated as random variables, it seems logical to adopt a Bayesian approach in estimating the probability distributions of these parameters. In particular, for given values of the cut-off points, consider a random sample $\psi_j = (y_j, \mathbf{z}_j), j = 1, \ldots, n$, from the screening population, where y_j denotes the cell of the reclassification table where individual j is classified into and $\mathbf{z}_j$ all other measurements taken on individual j required to derive the joint distribution of the vector $\alpha = (\alpha_{11}, \ldots, \alpha_{mm})'$. The ψ_j's may assumed to be independent across individuals, but y_j and $\mathbf{z}_j$ are likely to be correlated within individuals. Let $g(\psi_j | f_{11}, \ldots, f_{mm}, \alpha)$ denote the joint density of ψ_j, and consider the factorization

$$g(\psi_j|f_{11},\dots,f_{mm},\alpha) = g(\mathbf{z}_j|y_j,\alpha)g(y_j|f_{11},\dots,f_{mm}) \\ = g(\mathbf{z}_j|y_j,\alpha_{y_j})g(y_j|f_{11},\dots,f_{mm}). \quad (5)$$

Marginally, $g(y_j|f_{11},\dots,f_{mm})$ is the frequency function of a discrete random variable with probability f_{kl} that individual j is classified into the klth cell of the reclassification table. Given that $y_j \perp\!\!\!\perp y_k$, the observed number of individuals $n_{11},\dots,n_{mm}$ in the different cells of the reclassification table is multinomially distributed with probability vector $(f_{11},\dots,f_{mm})'$. Using a Dirch$(a_{11},\dots,a_{mm})$ conjugate prior, the posterior distribution of $(f_{11},\dots,f_{mm})'$ can be modeled as $(f_{11},\dots,f_{mm})' \sim \text{Dirch}(a_{11}+n_{11},\dots,a_{mm}+n_{mm})$.[20] To estimate the joint distribution of α_{kl}, $k,l = 1,\dots,m$, $k \neq l$, it makes sense to condition the observations on the observed value of y_j. The data $\mathbf{z}_{kl,1},\dots,\mathbf{z}_{kl,n_{kl}}$ can then be treated as independent samples from each level of y_j, such that separate multivariate Bayesian models can be fitted to each of these samples to obtain the posterior distributions of α_{kl}.

In a deterministic sensitivity analysis, no attempt is made to specify parameter uncertainty through the use of probability distributions. Instead, reasonable lower and upper bounds are identified for each of the parameters of interest, after which the actual sensitivity analysis is conducted by exploring in a deterministic way how different combinations of the parameter values affect the value of $\Delta_{scr}^{\max}$. To explore which of the input parameters have the highest impact on the outcome variable, it is common practice to change one parameter value at a time, resulting in a so-called 1-way sensitivity analysis. It is also possible to perform a multi-way sensitivity analysis by changing two or more parameter values simultaneously. However, the parameter values are still allowed to vary independently from each other as nothing is known about the correlation between these parameters.

ILLUSTRATIVE CASE STUDY

To demonstrate how our proposed method can assist in quantifying the headroom available to an improved, biomarker-based screening test, we applied the method in a case study related to the prevention of type 2 diabetes mellitus (DM2) and its associated micro- and macro-vascular complications.

Clinical context

Lifestyle interventions have previously shown to be a cost-effective strategy to reduce the incidence of DM2 in patients with pre-diabetes.[21] The primary prevention program considered in this case study therefore consists of providing a lifestyle intervention to individuals who are at increased risk of developing DM2. Screening variant s^1 is based on existing clinical risk factors that have previously been shown to have a strong predictive power for the risk of developing DM2. Screening variant s^2 comprises of the same risk factors as well as a hypothetical new biomarker for predicting the onset of DM2, such as a genetic marker related to metabolic programming by perinatal nutrition or a blood-based marker related to lipotoxicity and its metabolic consequences. As a result of the initial screening, individuals are classified into three risk categories. Individuals who are considered to be at high risk are offered an intensive, three-year lifestyle-intervention program consisting of both a dietary part and a physical activity part. Individuals who are considered to be at intermediate risk are offered a more basic variant consisting of a dietary component only. No intervention is offered to individuals who are considered to be at low risk.

Structure of the disease progression model

To estimate the expected lifetime cost and QALY consequences of applying t^i_{kl} to subpopulation kl, we constructed a discrete-time Markov model with three health states (see Figure 2 for a schematic representation): *no diabetes*,

diabetes, and *death*. The transition probabilities $p^{13}(\text{sex, age})$ and $p^{23}(\text{sex, age})$ were assumed to depend on sex and age, whereas the probability $p_{kl}^{12}(t_{kl}^{i})$ of making a transition from the *no diabetes* to the *diabetes* state was assumed to depend on the subpopulation kl and on the treatment t_{kl}^{i} that is assigned to these individuals under screening variant s^{i}. The applied cycle length was one year.

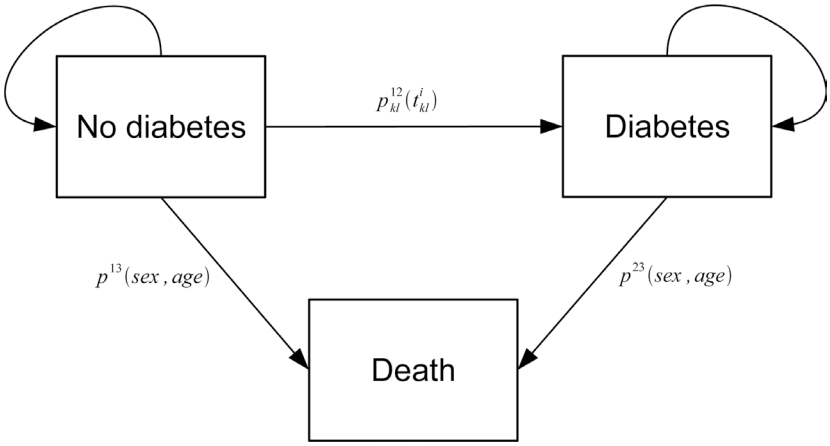


Figure 2: A discrete-time Markov model with three health states.

Table 2: Model parameters, their initial values, and the sources used to obtain these values.

Parameter	Symbol in text (if defined)	Value	Source
Willingness-to-pay for a QALY	λ	20,000	Reference value
Reclassification table			
Size of the study population	n	4977	[22]
Size of subpopulation *kl*	n_{kl}	Table 3	[22]
Observed number of DM2 cases in subpopulation *kl*	n_{kl}^{event}	Table 3	[22]
Fraction of individuals in subpopulation *kl*	f_{kl}	n_{kl}/n	
Seven-year incidence of DM2 in subpopulation *kl* in the absence of screening	$I_{kl}(7)$	$n_{kl}^{\text{event}}/n_{kl}$	

Table 2: Model parameters, their initial values, and the sources used to obtain these values. *Continued*

Parameter	Symbol in text (if defined)	Value	Source
Transition probabilities			
Reduced risk of developing DM2 for the intensive variant of the lifestyle-intervention program (hazard ratio)	$\beta^{\text{intensive}}$	0.70	[23]
Reduced risk of developing DM2 for the basic variant of the lifestyle-intervention program (hazard ratio)	β^{basic}	0.85	Expert judgment
Increased risk of death with diabetes (hazard ratio)	β^{diabetes}	2.13	[24]
Instantaneous rate of transiting from the no diabetes to the diabetes state in subpopulation *kl*	μ_{kl}^{12}	$\frac{-\log(1-I_{kl}(7))}{7}$	
One-year probability of transiting from the no diabetes to the diabetes state in subpopulation *kl*	$p_{kl}^{12}(t_{kl}^{i})$	$1 - \exp(-\beta^{t_{kl}^{i}}\mu_{kl})$	
One-year probability of transiting from the no diabetes to the death state (sex and age dependent)	$p^{13}(\text{sex}, \text{age})$	various	National life tables
One-year probability of transiting from the diabetes to the death state (sex and age dependent)	$p^{23}(\text{sex}, \text{age})$	$1-\exp(\log(1-p^{13}(\text{sex}, \text{age}))\beta^{\text{diabetes}})$	
Costs and utilities attached to the Markov model's health states			
Cost attached to the no diabetes state		0	
Utility attached to the no diabetes state		0.84	[25]
Cost attached to the diabetes state		1805	[24]
Utility attached to the diabetes state		0.65	[24]
Costs of the lifestyle-intervention program			
Cost of the intensive variant in the first year		800	[23]
Cost of the intensive variant in the second and third year		520	[23]
Cost of the intensive variant in all subsequent years		0	
Cost of the basic variant in the first year		320	Expert judgment
Cost of the basic variant in the second and third year		160	Expert judgment
Cost of the basic variant in all subsequent years		0	
Patient characteristics			
Mean age in subpopulations 12 and 21		60	Expert judgment

Table 2: Model parameters, their initial values, and the sources used to obtain these values. *Continued*

Parameter	Symbol in text (if defined)	Value	Source
Mean age in subpopulations 13 and 31		63	Expert judgment
Mean age in subpopulations 23 and 32		67	Expert judgment
Fraction of female subjects in sub-population kl		0.543	[22]

Parameter estimation

A summary of all parameters of interest, their initial values, and the sources used to obtain these values is provided in Table 2. As the novel biomarker included in s^2 had not yet been evaluated in a prospective cohort study, we had to rely on surrogate data to obtain initial values of some of these parameters. How this was done exactly is briefly described in the subsections below.

Specification of the fractions f_{kl}

Salomaa et al. evaluated whether a combined score of four novel biomarkers (adiponectin, apolipoprotein B, C-reactive protein, and ferritin) could improve the prediction of clinically incident diabetes over and above eleven classical risk factors, including blood glucose.[22] For the purpose of this case study, we assumed that the performance of this biomarker score could serve as a proxy for the performance of the novel biomarker included under s^2. This allowed us to derive initial values of the fractions f_{kl} from the reclassification table that the authors produced for performing their NRI calculations (Table 3) by setting $\gamma^1 = \gamma^2 = (0.03, 0.15)'$ and $f_{kl} = n_{kl}/n$.

Table 3: Observed number of diabetes cases (n_{kl}^{event}) / total number of subjects (n_{kl}) after seven years of follow-up in the HEALTH 2000 cohort. 22

		Predicted risk with classical risk factors plus biomarker score		
		low risk	**intermediate risk**	**high risk**
Predicted risk with	low risk	29/3029	9/141	-
classical risk factors	intermediate risk	8/337	89/1228	15/68
	high risk	0/1	5/27	33/146

Specification of the transition probabilities of the Markov model

The sex- and age-dependent transition probabilities from the *no diabetes* to the *death* state were taken from national life tables. The transition probabilities from the *diabetes* to the *death* state were derived from the transition probabilities from the *no diabetes* to the *death* state by first converting them into instantaneous death rates and then multiplying these death rates by a relative risk increase (hazard ratio) of $\beta^{\text{diabetes}} = 2.13$, which was obtained from a previously conducted economic evaluation. [24] To estimate $p_{kl}^{12}(t_{kl}^i)$, we assumed that the cumulative incidence functions $I_{kl}(\tau)$ were exponentially distributed, which allowed us to express the underlying instantaneous transition rates μ_{kl} as [16]

$$\mu_{kl} = \frac{\text{-log}(1\text{-}I_{kl}(\tau))}{\tau}. \quad (6)$$

The corresponding one-year transition probabilities $p_{kl}^{12}(t_{kl}^i)$ can then be expressed as

$$p_{kl}^{12}(t_{kl}^i) = 1\text{-exp}(\text{-}\beta^{t_{kl}^i}\mu_{kl})), \quad (7)$$

where $\beta^{t_{kl}^i}$ denotes the hazard ratio comparing individuals receiving the preventive intervention t_{kl}^i to individuals not receiving a preventive

intervention. For this case study, we set $\beta^{\text{intensive}} = 0.7$, which corresponds to the reduction in DM2 risk observed in the SLIM study. [23] As lifestyle interventions consisting of a dietary component only are less effective than lifestyle interventions consisting of both a dietary and a physical component, the value of $\beta^{\text{intermediate}}$ was assumed to be slightly higher and set equal to 0.85. Equations (6) and (7) were subsequently applied to transform the 7-year diabetes incidences $I_{kl}(7) = n_{kl}^{\text{event}}/n_{kl}$ as observed in the HEALTH 2000 cohort (Table 3) into initial values of the one-year transition probabilities $p_{kl}^{12}(t_{kl}^{i})$.

Cost and utility estimates

The cost and utility estimates attached to the Markov model's transient health states were taken from previously conducted economic evaluations. The costs associated with the intensive variant of the lifestyle-intervention program were derived from the SLIM study and were set equal to EUR 800 for the first year and EUR 520 for the second and third year. For the basic variant of the lifestyle-intervention program, these costs were set equal to EUR 320 and EUR 160, respectively.

Specification of the patient characteristics

As we did not have access to the original data used to construct Table 3, we had to rely on expert judgment to obtain initial values of the mean age and male/female ratio in the different subpopulations kl. Age is generally considered to be a strong predictor for the development of DM2, and this was taken into account when specifying the base-case values of the mean age in each of the subpopulations kl. In particular, the mean age in individuals who were classified as low risk under one of the screening variants and high risk under the other was set equal to 63, which corresponds to the third quartile of the age distribution observed in the HEALTH 2000 cohort. The mean age in individuals who are classified as low risk under one screening variant and intermediate risk under the other was subsequently assumed to

be slightly lower and set equal to 60, whereas the mean age in individuals who are classified as intermediate risk under one screening variant and high risk under the other was assumed to be slightly higher and set equal to 67. Sex is usually not associated with the development of DM2. The fraction of female subjects in each of the subpopulations kl was therefore assumed to be equal to 0.543, which corresponds to the fraction of female subjects observed in the HEALTH 2000 cohort.

Results of the initial economic evaluation

For the base-case values of the model parameters, the commercial headroom available to the biomarker-based screening variant was found to be equal to EUR 75. To determine the robustness of this result with respect to small changes in the parameter values, we performed a one-way sensitivity analysis (Figure 3). The amount of headroom available was most sensitive to changes in the effect of the two lifestyle interventions and to changes in the mean age in each of the subpopulations kl. Changing the 7-year disease incidences $I_{kl}(7)$ had a less profound impact on $\Delta_{\mathrm{scr}}^{\mathrm{max}}$. We also varied the male/female ratios in each of the subpopulations kl as well as the different cost components of the two lifestyle interventions, but these changes only had a marginal impact on the amount of headroom available (results not shown).

Implications

Based on the results of our initial economic evaluation, we can conclude that if the cost of measuring the novel biomarker is expected to be relatively low, there seems still sufficient room for improving the predictive performance of the existing risk classification models to warrant further research on novel biomarkers that are independently associated with the onset of DM2. On the other hand, if the unit cost of measuring the novel biomarker is likely to exceed EUR 100, it may be more fruitful to focus the research effort on identifying prognostic biomarkers that have a strong correlation with one

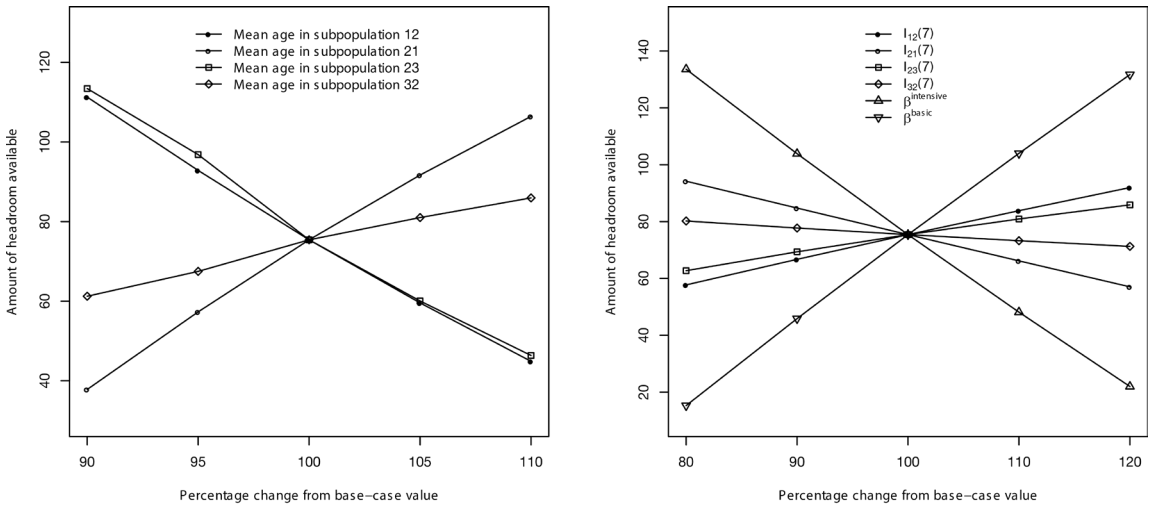


Figure 3: Results of the 1-way sensitivity analysis.

of the established risk factors but are less expensive and/or invasive to measure. However, whether such a biomarker would actually be suitable as a substitute for an established risk factor not only depends on the effect size of this new biomarker compared to the established risk factor but also on its whole correlation structure with all the other variables included in the risk stratification model. This should be considered as well when making a go/no-go decision on the search for such a biomarker.

When performing our analysis, we implicitly assumed that the hypothetical new biomarker will have similar predictive capabilities as the biomarker risk score considered by Salomaa et al.[22] It should therefore be noted that if such a biomarker is expected to have better (worse) capabilities in reclassifying subjects at risk of developing DM2, the cost of measuring the biomarker may be higher (should be lower) than the suggested upper bound of EUR 100. However, care should be taken not to raise the amount of headroom available too easily as the results of previous studies suggest that the initial expectations of a new biomarker have often been too optimistic, with disappointments in later phases of analyses.[1]

DISCUSSION

Moving from a screening variant based on traditional risk factors to a screening variant based on traditional risk factors plus a novel biomarker results in a reclassification of some of the individuals. To determine the maximum increase in screening cost for which this reclassification is still likely to be cost-effective, we first restructured the decision problem in such a way that part of the parameters of interest could be estimated through disease progression modeling. We then described how these models could be combined with estimated values of the degree of reclassification to obtain initial estimates of the amount of commercial headroom available. Our method was illustrated in a case study related to the prevention of DM2, were a Markov model with three health states was used to perform an initial economic evaluation of a potential new biomarker technology by using published data on the NRI of related but already clinically validated biomarkers.

A general method for the early HTA of new medical devices has previously been suggested by Pietzsch and Paté-Cornell.[2] Their approach requires an analyst to represent the dependency between the decision variable and the outcome of interest through a sequence of primary and intermediate effect variables, thereby allowing the analyst to obtain concrete realizations of the outcome variable by sampling from a series of conditional probability distributions. Our method is similar in the sense that we also determine the effect of the decision variable (biomarker-based screening or screening without using the biomarker) on the outcome of interest (the amount of headroom available) by sampling from a series of conditional probability distributions. However, instead of requiring the analyst to provide exact functional forms for each of these probability distributions, we have restructured the decision problem in such a way that some of these distributions can be approximated through disease progression modeling. This makes our method easier to apply in situations were there are no clear functional relationships between the variables of interest, like in our case study related to the prevention of DM2.

A limitation of performing early HTA is that there is generally only a limited amount of clinical data available with which to populate the decision models. This implies that to be able to compute the amount of headroom available, it is sometimes required to make strong assumptions on some of the unknown parameters, and in this respect our method is no exception. In our case study, this became especially apparent when specifying the values of the cut-off points used to differentiate between low-, intermediate-, and high-risk individuals, the changes in the distribution of individuals across these three risk categories as a result of moving from the traditional risk classification model to the biomarker-based risk classification model, and the patient characteristics in each of the six subpopulations. The specification of the cut-off points should ideally be based on the ratio of the costs associated with a false positive and the benefits foregone due to a false negative, and the NMB framework provides a means of formally quantifying this trade-off.[14] However, even with such a formal framework in place, it remains difficult to determine the values of the cut-off points in an 'objective' way: (i) analysts are still required to make a value judgment about the willingness-to-pay per unit of health gain, and (ii) the effectiveness of the administered treatments is likely to depend on the selected cut-off points, but the clinical data required to estimate this dependency may not be available.

Athough the use of our method provides insight into the amount of headroom available to a novel biomarker, it does not directly provide an answer to the question of whether a medical technology firm should proceed with developing a technology that can measure this biomarker in actual clinical settings. To address this latter aspect, the results of the initial economic evaluation must first be translated into an estimate of the technology's maximum sales price by applying the principle of value-base pricing.[26] This estimate can then be fed into an appropriate product investment evaluation method, such as the one proposed by Girling et al.[27], to determine whether the expected post-market cash flows are sufficiently large to warrant further investments to transform the current concept into a fully developed end

product. When performing such a return-on-investment (ROI) analysis for a specific biomarker technology, one should be aware that the technology can potentially be used for multiple purposes, e.g., not only as a screening test for selecting individuals eligible for a subsequent preventive intervention, but also as a test for monitoring treatment response once the disease has been clinically established. All these potential uses of a new biomarker technology should ideally be taken into account when determining the technology's maximum sales price and estimating the subsequent expected post-market cash flows. However, performing such a comprehensive ROI analysis is not straightforward, and future research on this problem area seems desirable.

To conclude, we presented a method for the early HTA of novel biomarker measurement in primary prevention programs and applied this method in a case study related to the prevention of DM2. Although we have focused on the use of the biomarker as a screening test for identifying individuals at risk of developing a chronic disease, our approach of first identifying the parameters of interest and then restructuring the decision problem in such a way that part of these parameters can be estimated through disease progression modeling seems more generally applicable. Future research effort may therefore be directed at exploring whether it is possible to quantify the clinical value of other potential applications of a new biomarker technology, such as a diagnostic test or a disease monitoring test, in a similar way.

REFERENCES

1. Hlatky, MA, Greenland, P, Arnett, DK, *et al*. Criteria for Evaluation of Novel Markers of Cardiovascular Risk. *Circulation* **119**, 2408–2416 (2009).
2. Pietzsch, JB & Paté-Cornell, ME. Early technology assessment of new medical devices. *International Journal of Technology Assessment in Health Care* **24**, 36–44 (2008).
3. Chambless, LE, Cummiskey, CP & Cui, G. Several methods to assess improvement in risk prediction models: Extension to survival analysis. *Statistics in Medicine* **30**, 22–38 (2011).
4. Chambless, LE & Diao, G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine* **25**, 3474–3486 (2006).
5. Cook, NR & Ridker, PM. Advances in measuring the effect of individual predictors of cardiovascular risk: The role of reclassification measures. *Annals of Internal Medicine* **150**, 795–802 (2009).
6. Heagerty, PJ & Zheng, Y. Survival Model Predictive Accuracy and ROC Curves. *Biometrics* **61**, 92–105 (2005).
7. Pencina, M, D'Agostino Sr, R, D'Agostino Jr, R & Vasan, R. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine* **27**, 157–172 (2008).
8. Pencina, MJ, Steyerberg, EW & D'Agostino, RB. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine* **30**, 11–21 (2011).
9. Vickers, AJ & Elkin, EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making* **26**, 565–574 (2006).
10. Hemingway, H, Henriksson, M, Chen, R, *et al*. The effectiveness and costeffectiveness of biomarkers for the prioritisation of patients awaiting coronary revascularisation: A systematic review and decision model. *Health Technology Assessment* **14**, 1–178 (2010).
11. Henriksson, M, Palmer, S, Chen, R, *et al*. Assessing the cost effectiveness of using prognostic biomarkers with decision models: Case study in prioritising patients waiting for coronary artery surgery. *BMJ* **340**, 194–205 (2010).
12. IJzerman, M & Steuten, LM. Early Assessment of Medical Technologies to Inform Product Development and Market Access: A Review of Methods and Applications. *Applied Health Economics and Health Policy* **9**, 331–347 (2011).
13. Cosh, E, Girling, A, Lilford, R, McAteer, H & Young, T. Investing in new medical technologies: A decision framework. *Journal of Commercial Biotechnology* **13**, 263–271 (2007).

14. Steyerberg, E, Vickers, A, Cook, NR, *et al.* Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology* **21**, 128–138 (2010).

15. Drummond, MF, Sculpher, MJ, Torrance, GW, O'Brien, BJ & Stoddart, GL. *Methods for the Economic Evaluation of Health Care Programmes.* (Oxford University Press, 2005).

16. Briggs, A, Sculpher, M & Claxton, K. *Decision Modelling for Health Economic Evaluation.* (Oxford University Press, 2006).

17. Groot Koerkamp, B, Stijnen, T, Weinstein, MC & Hunink, MGM. The combined analysis of uncertainty and patient heterogeneity in medical decision models. *Medical Decision Making* **31**, 650–661 (2011).

18. Garthwaite, PH, Kadane, JB & O'Hagan, A. Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association* **100**, 680–701 (2005).

19. Vallejo-Torres, L, Steuten, LMG, Buxton, MJ, *et al.* Integrating health economics modeling in the product development cycle of medical devices: A Bayesian approach. *International Journal of Technology Assessment in Health Care* **24**, 459–464 (2008).

20. Briggs, AH, Ades, AE & Price, MJ. Probabilistic Sensitivity Analysis for Decision Trees with Multiple Branches: Use of the Dirichlet Distribution in a Bayesian Framework. *Medical Decision Making* **23**, 341–350 (2003).

21. Gillies, CL, Abrams, KR, Lambert, PC, *et al.* Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *BMJ (Clinical research ed.)* **334**, 299–308 (2007).

22. Salomaa, V, Havulinna, A, Saarela, O, *et al.* Thirty-one novel biomarkers as predictors for clinically incident diabetes. *PLOS ONE* **5**, 1–8 (2010).

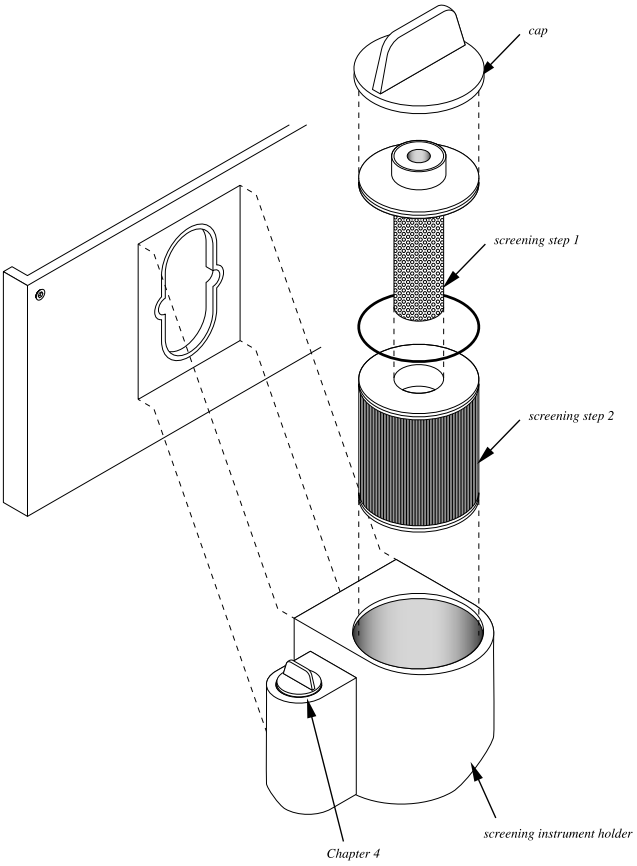
23. Roumen, C, Corpeleijn, E, Feskens, EJM, *et al.* Impact of 3-year lifestyle intervention on postprandial glucose metabolism: the SLIM study. *Diabetic Medicine* **25**, 597–605 (2008).

24. Gillies, CL, Lambert, PC, Abrams, KR, *et al.* Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *BMJ (Clinical research ed.)* **336**, 1180–1191 (2008).

25. Sullivan, PW & Ghushchyan, V. Preference-based EQ-5D index scores for chronic conditions in the United States. *Medical Decision Making* **26**, 410–420 (2006).

26. Claxton, K, Briggs, A, Buxton, MJ, *et al.* Value based pricing for NHS drugs: an opportunity not to be missed? *BMJ* **336**, 251–254 (2008).

27. Girling, A, Young, T, Brown, C & Lilford, R. Early-stage valuation of medical devices: the role of developmental uncertainty. *Value in Health* **13**, 585–591 (2010).


cap
screening step 1
screening step 2
screening instrument holder
Chapter 4

Chapter 5

Design of stepwise screening for prediabetes and type 2 diabetes based on costs and cases detected

Published as:
de Graaf, G., Postmus, D., Bakker, S. J. L., & Buskens, E.
Design of stepwise screening for prediabetes and type 2 diabetes based on costs and cases detected. *Journal of Clinical Epidemiology*, 68, 1010-1018 (2015)

ABSTRACT

OBJECTIVE To provide insight in the trade-off between cost per case detected and the detection rate in questionnaire based stepwise screening for impaired fasting glucose and undiagnosed type 2 diabetes.

STUDY DESIGN We considered a stepwise screening in which individuals whose risk score exceeds a predetermined cut-off value are invited for further blood glucose testing. Using individual patient data to determine questionnaire sensitivity and specificity and external sources to determine screening costs and patient response rates, we rolled-back a decision tree to estimate the cost per case detected and the detection rate for all possible cut-offs on the questionnaire.

RESULTS We found a U-shaped relation between cost per case detected and detection rate, with high costs per case detected at very low and very high detection rates. Changes in patient response rates had a large impact on both the detection rate and the cost per case detected, whereas screening costs and questionnaire accuracy mainly impacted the cost per case detected.

CONCLUSION Our applied method makes it possible to identify a range of efficient cut-offs where higher detection rates can be achieved at an additional cost per detected patient. This enables decision makers to choose an optimal cut-off based on their willingness to pay for additional detected patients.

INTRODUCTION

Type 2 diabetes mellitus (DM2) is a disease associated with a large burden at both patient and societal level. In Europe, an estimated 56.3 million people aged 20 – 79 have diabetes, of which 90% have DM2. The associated direct healthcare costs amounted to €111 billion in 2013.[1] DM2 is therefore widely considered to be a major public health problem. There are two main strategies to address this issue.[2] First is the reduction in the incidence of DM2 related complications through the early detection and treatment of asymptomatic DM2 patients (secondary prevention). Second is the provision of interventions aimed at slowing down the progression to DM2 in patients considered to be at high risk of developing DM2 (primary prevention), which is usually defined in terms of the presence or absence of prediabetes (i.e. impaired glucose tolerance or impaired fasting glucose (IFG)). As both strategies rely on blood glucose testing to either diagnose DM2 (secondary prevention) or to diagnose prediabetes and rule-out undiagnosed DM2 (primary prevention), a practical implementation of the second strategy results in finding undiagnosed DM2 patients as well. As a result, the combined screening for prediabetes and previously undiagnosed DM2 is more efficient and has gained widespread interest in the past years.[3–5]

The target population for prediabetes and DM2 screening includes a large part of the entire population, but prevalences are low. Consequently, economic and logistic aspects of screening tools are an important consideration. To that end, blood glucose testing is generally considered too burdensome and costly to be applied in all individuals eligible for screening.[6–9] Instead, consensus has been reached that screening should proceed in a stepwise manner by first making a preselection of high-risk individuals and then inviting those exceeding a predetermined threshold for further blood glucose testing. Risk questionnaires based on a small set of bio-characteristics have shown to be accurate predictors of DM2 risk, while being a relatively inexpensive form of

testing.[10,11] Stepwise screening using a risk questionnaire has therefore found its way into several guidelines.[12–14]

Although the strategy of stepwise screening is more feasible and practical, it inevitably also leads to a number of undiagnosed DM2 and prediabetes patients remaining undetected. Stepwise screening thus presents a tradeoff between feasibility, often measured in terms of the cost per case detected[15–17], and the detection rate (percentage of patients with disease in the target population that are detected through screening[18]). In current guidelines, the selection of the cut-offs was based on an arbitrary value of absolute risk[12,13] or was not supported at all [14]. It therefore seems that the economic aspects were not explicitly considered during the formation of these guidelines, which may have been caused by the lack of insight in the trade-off between the cost per case detected and the detection rate.

In this paper, we seek to provide insight in the trade-off between cost per case detected and detection rate that comes with choosing a cut-off on a risk questionnaire. Furthermore, we want to estimate the effects of changes in patient response rates, screening costs, and questionnaire accuracy within the strategy of questionnaire based stepwise screening on this trade-off.

METHODS

Structure of the stepwise screening program

The stepwise screening program evaluated in this study was based on the Dutch guideline 'Preventieconsult'.[13] This guideline was developed to identify individuals at an increased risk for developing cardiovascular diseases, DM2, and kidney damage. We adapted this strategy to focus solely on IFG and DM2 by assuming the use of a dedicated DM2 questionnaire based on a version of the FINDRISC validated in the Dutch population.[19]

The protocol for the screening program is as follows. Screening is initiated through the GP office by sending a questionnaire to all registered individuals of between the ages of 40 and 75 that have not been diagnosed with DM2 before. This questionnaire is returned to the GP office and assessed by the GP or a nurse. All individuals with a score equal to or above a predetermined cut-off value are invited for a consult and instructed to follow an 8-hour fasting protocol. At the consult, the answers of the questionnaire are verified and discussed and a fasting plasma glucose test is performed using a plasma calibrated capillary blood glucose meter. All patients with fasting plasma glucose levels of 6.1 mmol/L or higher are invited for a second consult and instructed to follow the fasting protocol again. During the second consult another fasting plasma glucose test is performed. The final diagnosis is based on the lower of the two test results. Thus, patients are diagnosed with DM2 if their fasting plasma glucose levels on both tests are 7.0 mmol/L or higher, with IFG if their fasting glucose level for the second test is between 6.1 mmol/L and 7.0 mmol/L, and with normal fasting glucose if their fasting plasma glucose level on the second test is below 6.1 mmol/L.

Risk questionnaire

The Dutch version of the FINDRISC questionnaire used in our screening design calculates a risk score based on five patient characteristics. These characteristics and the maximum number of points that can be acquired for each are: age (4 points), body mass index (3 points), waist circumference (4 points), the use of antihypertensives (2 points), and the occurrence of parental diabetes (5 points). This means that a patient can score between 0 points (lowest risk) and 18 points (highest risk).[19] The original version of the Dutch FINDIRSC questionnaire includes an item on previously diagnosed DM2. As this was an exclusion criterion for the screening protocol as defined in the guideline, we removed this item from the questionnaire in our study.

Study population

The assessment of the stepwise screening program was performed using data from the PREVEND Groningen study, a cohort drawn from the general population in the city of Groningen in the Netherlands. Details of the study design have been published elsewhere.[20] In short, a total of 40,856 individuals provided a urine sample and completed a questionnaire on demographics, history of cardiovascular and metabolic risk factors, known diabetes, medication use, and pregnancy. Pregnant females and patients using insulin were excluded. All participants with urinary albumin concentration of at least 10 mg/L willing to participate were enrolled in the study (n = 6,000). A random sample of those with urinary albumin concentration less than 10 mg/L were added to form a total study cohort of 8,592 participants. At baseline, participants underwent outpatient visits to assess demographics, anthropometric measurements, cardiovascular and metabolic risk factors, health behavior, and family history. Additionally, a blood sample was collected after an overnight fasting, from which fasting plasma glucose measurements were taken.

We selected all patients from the cohort that fulfilled the age criterion in the guideline (40-75 years, n = 6,244). For the purpose of this paper, participants in the cohort who were known to have diabetes (n = 149), who did not adhere to the fasting protocol (n = 857), or who did not have a baseline fasting plasma glucose measurement (n = 56) were excluded from the analysis. Known diabetes cases were identified either through the registered use of anti-diabetics in the pharmacy registry or the indication of being diabetic on the baseline questionnaire. Adherence to the fasting protocol was based on self-indication of consumption of any food or drinks other than water since midnight on the day of the glucose test. The use of antihypertensive medication was based on data in the pharmacy registry or self-reported use for those participants not present in the pharmacy registry. The PREVEND baseline questionnaire contained separate questions on the presence of DM2

in both parents. These variables were combined into one variable indicating the presence of parental diabetes. Finally, all patients with missing data on one of the variables required to calculate the FINDRISC (i.e. age, sex, body mass index, waist circumference, use of hypertension medication, and parental diabetes) were excluded (n = 333).

Outcome measures

The performance of the questionnaire-based stepwise screening program was assessed using two outcome measures: 1) the number of IFG and DM2 patients identified as a result of screening as a percentage of the total number of IFG and DM2 patients in the target population (detection rate)[18], and 2) the screening cost per case detected (CPCD). To compute these two outcome measures for a given cut-off on the FINDRISC, we first determined the expected cost and probability of being detected for each individual in our study population. These were calculated by rolling back the decision tree depicted in Figure 1. The individuals' scores on the FINDRISC were computed based on the risk factor data taken from the baseline questionnaire of the PREVEND study. The detection rate was then obtained by summing all individual detection probabilities and dividing it by the total number of IFG and DM2 patients in the study population. The CPCD was calculated by summing all expected costs and dividing it by the sum of all individual detection probabilities. Confidence intervals (CI) for the outcome measures were obtained by repeating the analyses in 10,000 bootstrap samples from the study population and taking the 2.5th and 97.5th percentile of the outcomes.

Base case analysis

All cost and response parameters for the base case analysis are listed in Table 1. Probabilities of non-response at each step of the stepwise screening program were taken from the Preventieconsult trial[21] and were assumed to be the same for all individuals. The Preventieconsult trial did not report screening

costs. Instead, we based these on a screening program for chlamydia that also send out invitations via mail.[22] This program did however not involve the assessment of returned questionnaires. The total costs of questionnaire assessment were based on the costs for return postage (€0.50) and an estimation of the labor costs of evaluating the questionnaire (€0.50). Costs for a GP consult were taken from the Dutch reference price list.[23] Lastly, a cost estimate for the glucose test was based on a commercial quotation.

Table 1: Parameter data

Patient response rate	**Base case scenario**	**Aphrodite scenario**	**Increased awareness scenario**	**Full response scenario**
Return of questionnaire (%)	75	55	86	100
Attend 1st consult (%)	72	73	83	100
Attend 2nd consult (%)	84	84	84	100

Screening costs	**Base case scenario**	**Double consult scenario**	**Nurse consult scenario**	**Email invitation scenario**
Invitation and questionnaire (€)	2.50	2.50	2.50	0.00
Questionnaire assessment (€)	1.00	1.00	1.00	0.50
Invitation for consult (€)	2.50	2.50	2.50	0.00
Consult (€)	28.00	56.00	18.67	28.00
Fasting plasma glucose test (€)	0.95	0.95	0.95	0.95

Impact of selected parameters

The impact of response rates, screening costs, and questionnaire accuracy on the outcome measures were calculated by changing those parameters in the decision tree model, rolling back the tree, and calculating the outcome measures using the same method described previously. Parameter values used in the impact scenarios for response rates and screening costs are shown in Table 1.

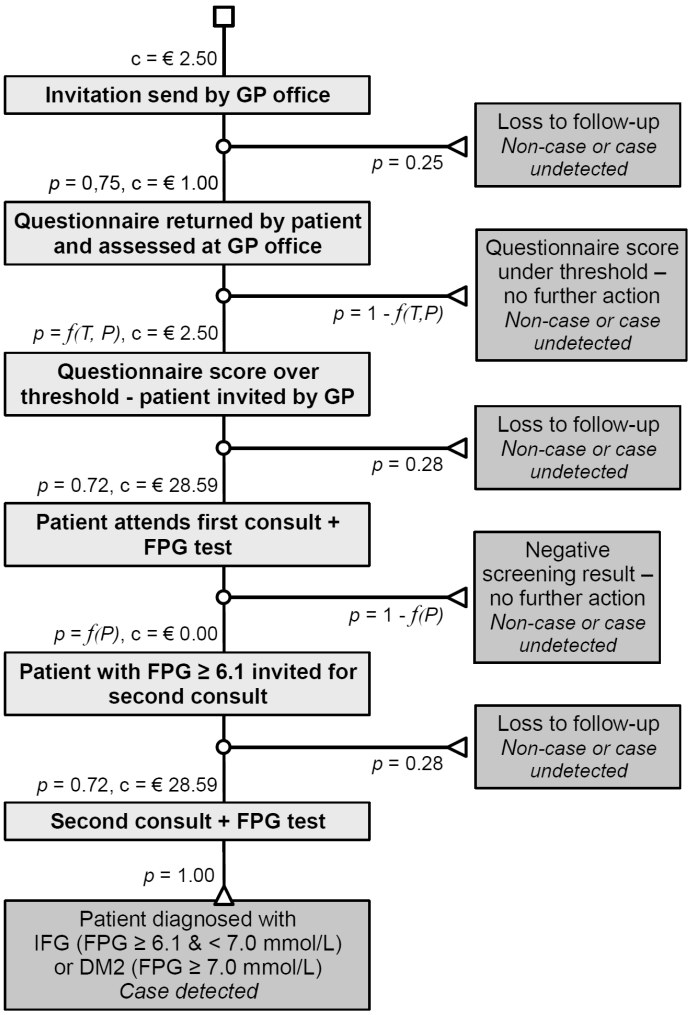


Figure 1: Model of the questionnaire based stepwise screening program. GP = General Practitioner; FPG = fasting plasma glucose; IFG = Impaired Fasting Glucose; DM2 = type 2 diabetes; TN = True Negative; FN = False Negative; TP = True Positive. Squares indicate decision notes, circles indicate chance nodes, and triangles indicate end notes of the decision tree. *p* indicates the probability a patient enters that branch of the decision tree, and c indicates the costs associated with entering that branch. *f(T,P)* indicates that the probability for that branch is a function of the chosen threshold value as well patient characteristics, whereas *f(P)* indicates that the probability is a function of patient characteristics only

Response rates

The Dutch APHRODITE study, which also included questionnaires send out by GPs, found a response rate of 55% for the questionnaire and 73% for the first consult.[24] A second response scenario was one where investment in public awareness would lead to an increase of 15% in response to both the questionnaire and the first consult, leading to response rates of 86% and 83%, respectively. To fully appreciate the effects of non-response on the outcomes of the screening, we also incorporated a hypothetical scenario where there is no non-response on all three invitations.

Screening costs

In the pilot study of the Preventieconsult[21], about half of the GPs indicated that they were unable to perform the consults within the standard duration of 10 minutes. This would mean that a 20-minute consult is required, which would double the consult costs. An alternative approach would be to have a nurse practitioner perform the consults. This would lead to an estimated reduction of the cost of the consults by one third. In a third and last scenario, the invitations and questionnaire would be sent out by email, reducing these costs to €0.00. The €0.50 for assessment of the questionnaire remains.

Questionnaire accuracy

A number of alternative questionnaires exist that could be applied in our study population, such as the Danish Diabetes Risk Score[25] and the PM1 score[26]. However, the difference in the accuracy of these screening tools with the FINDRISC is very small. An analysis of their impact would therefore not be very informative. Instead, we constructed three hypothetical questionnaires, with an area under the receiver operating characteristic (ROC) curve of approximately 5%, 10%, and 15% larger than the original FINDRISC questionnaire in our study population. For the questionnaires

with a 5% and 10% increased area under the curve, the FINDRISC score of all patients with IFG and DM2 was increased by 1 point, whereas the scores of all those with normal fasting glucose levels were lowered with one point. For the questionnaire with a 15% larger area under the curve the same approach was taken, but original scores were altered with 2 points. When this change led to questionnaire scores below 0 points or over the maximum of 18 points, these scores were set at 0 and 18, respectively. Subsequently, the ROC curves of these alternative questionnaires were inspected visually and additional changes were made to individual patient scores to improve the shape of the ROC and to approximate the intended area under the curve as closely as possible.

Table 2: Study population characteristics

	Normal fasting glucose	Impaired fasting glucose	Type 2 diabetes	All patients
Number of patients	4832	219	150	5201
Age, years (median (IQR))	52 (46 – 63)	60 (52 – 67)	63 (55 – 68)	53 (46 – 63)
Female sex	50.0%	38.4%	40.7%	49.2%
Fasting plasma glucose, mmol/L (median (IQR))	4.8 (4.4 – 5.1)	6.3 (6.2 – 6.6)	8.2 (7.3 – 10.9)	4.8 (4.5 – 5.3)
BMI kg/m2 (median (IQR))	25.9 (23.6 – 28.7)	28.7 (26.4 – 31.6)	28.9 (26.2 – 32.0)	26.1 (23.8 – 28.9)
Waist circumference, cm (mean ± SD)	89.9 ± 12.5	99.3 ± 11.4	100.4 ± 12.9	90.6 ± 12.7
Use of antihypertensive medication	16.2%	32.4%	38.7%	17.5%
Diabetes in family	15.8%	31.1%	25.3%	16.7%
FINDRISC questionnaire score (median (IQR))	7 (3 – 9)	10 (7 – 12)	10 (8 – 12)	7 (3 – 9)

RESULTS

Study population and questionnaire accuracy

The final sample of the study population consisted of 5,201 individuals, of which 219 (4.21%) had IFG and 150 (2.88%) had previously undetected DM2, yielding in a total number of 369 cases (Table 2). Compared to those with normal fasting glucose, patients with IFG and DM2 are on average older, more often male, have a higher BMI, larger waist circumference, use antihypertensive medication more often, and more frequently have family members with diabetes (Table 2). Within this study population, the FINDRISC based questionnaire had an area under the ROC curve of 74.3% (95% CI 71.9%-76.7%) (Figure 2 left panel). The area under the ROC curve for the IFG patients alone was 72.5% (95% CI 69.2%-75.7%), whereas that for the DM2 patients alone was 74.8% (95% CI 71.3%-78.2%) (results not shown). Sensitivity and specificity for each cut of point of the FINDRISC for the IFG and DM2 patients separately are shown in Table 3. The three hypothetical questionnaires used in the sensitivity analysis of questionnaire accuracy had areas under the curve of 80%, 85% and 90% (Figure 2 right panel).

Base case analysis

A U-shaped relation was found between detection rate and CPCD, with a high CPCD at both very low and high detection rates (Figure 3 top left panel). Lowest CPCD was achieved at cut-off 10, being €445 (95% CI €398-€507) with a detection rate of 24.6% (95% CI 21.3%-28.0%). Increasing the cut-off to 11 led to unfavorable effects on both outcomes, as CPCD increased to €481 (95% CI €423-€555) and the detection rate decreased to 19.3% (95% CI 16.3%-22.4%). On the other hand, decreasing the cut-off below 10 had a favorable effect on the detection rate, but an unfavorable effect on the CPCD. These cut-offs therefore present a trade-off between higher detection rates and higher CPCD. Cut-offs below 6 resulted in very little additional detection, but did lead to marked increases in CPCD. Due to the non-response, the detection

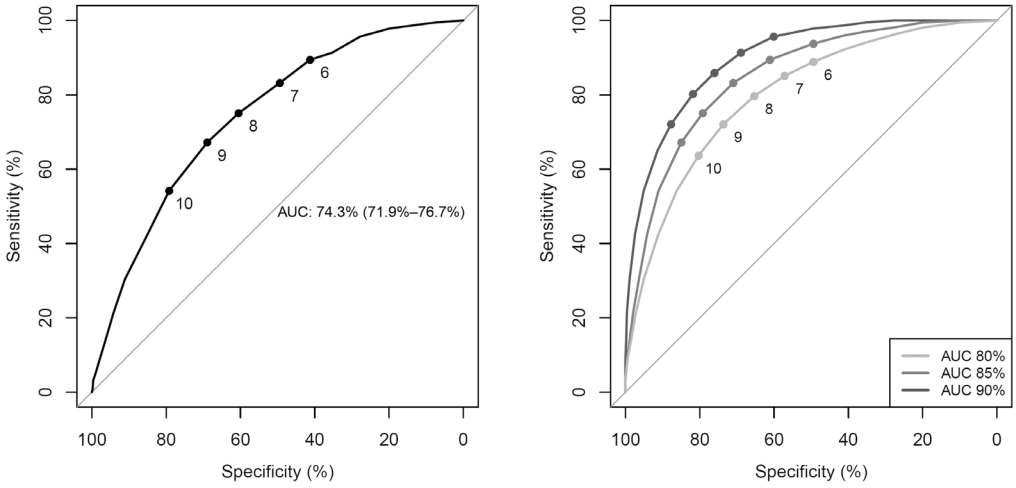


Figure 2: Receiver operator characteristic curves of the base case (left) and impact analysis (right). Numbers accompanying points on the curve indicate the cut-off score on the questionnaire. Numbers between parentheses are the 95% confidence interval of the area under the curve.

rate never exceeded 45.4% (detection rate at cut-off 0). The subgroup analysis demonstrated that the detection rate at each cut-off point is very similar in both IFG and DM2 patients (Table 3). This means that the proportion of IFG and DM2 patients identified at each cut-off is similar to proportion of prevalences in the target population.

Impact of selected parameters

The two outcome measures were impacted differently by the three parameters in the impact analysis (Figure 3). Changes in response levels had a large impact on both outcome measures. An increase in response led to a higher detection rate and lower CPCD and vice versa (Figure 3 top right panel). In the scenario with full response on all three invitations, the detection rate reached 100% (cut-off 0) and the detection rate at each cut-off corresponded to the sensitivity of that cut-off as shown on the ROC curve (Figure 2 left panel). In contrast, changes in the screening cost only influenced the CPCD. Changes

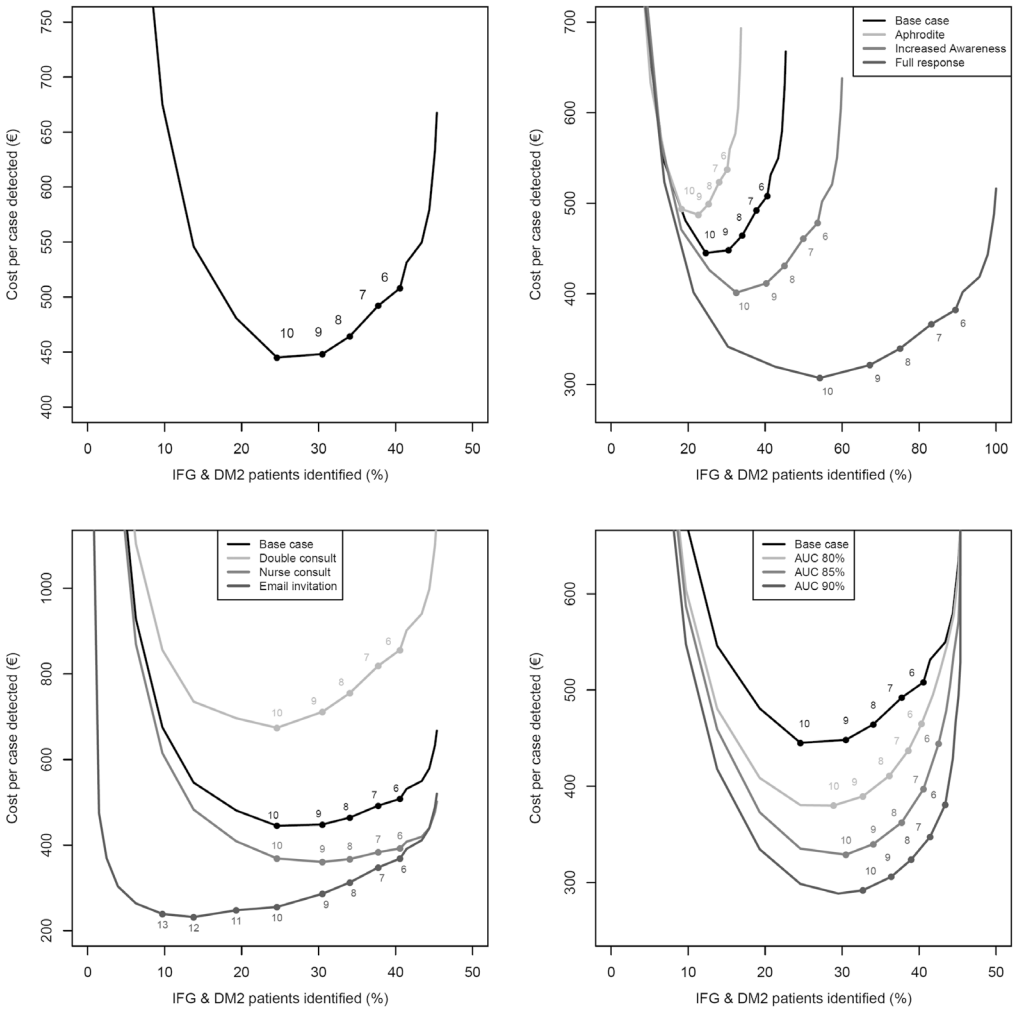


Figure 3: Outcomes of the base case (top left), impact analysis of patient response rates (top right), impact analysis of the screening costs (bottom left), and impact of the questionnaire accuracy (bottom right). Numbers indicate the cut-off score on the questionnaire. Parameter values of the impact analysis scenarios (gray curves) are shown in Table 1.

in costs later in the screening process (consult costs) had a larger impact on the lower cut-offs, whereas changes in costs early in the screening process (invitations and questionnaire assessment) had a larger impact on the higher cut-offs (Figure 3 bottom left panel). A change in accuracy did change both the CPCD and detection rate at each individual cut-off, but did not change the maximum achievable uptake. Finally, a larger area under the curve led to both a higher detection rate and a lower CPCD (Figure 3 bottom right panel). The effects of increased accuracy were most prominent in the cut-offs that are closest to the upper left corner in the ROC, whereas the position of the extremes remained more or less the same.

Table 3: Subgroup analysis for questionnaire accuracy and detection rate

	Impaired fasting glucose patients			Type 2 diabetes patients		
Cut-off	Sensitivity (%)	Specificity (%)	Detection rate (%)	Sensitivity (%)	Specificity (%)	Detection rate (%)
0	100.0	0.0	45.4	100.0	0.0	45.4
1	99.1	7.2	44.9	100.0	7.1	45.4
2	98.6	9.4	44.7	100.0	9.3	45.4
3	96.8	19.3	43.9	99.3	19.2	45.1
4	94.5	27.0	42.9	97.3	26.8	44.2
5	90.4	34.5	41.0	92.7	34.2	42.0
6	87.7	40.2	39.8	92.0	40.0	41.7
7	80.4	48.3	36.5	87.3	48.1	39.6
8	72.1	59.3	32.7	79.3	59.1	36.0
9	66.7	67.8	30.2	68.0	67.4	30.8
10	54.3	78.2	24.6	54.0	77.7	24.5
11	43.4	84.2	19.7	41.3	83.7	18.7
12	29.2	90.4	13.3	32.0	90.2	14.5
13	20.5	93.6	9.3	22.7	93.5	10.3
14	14.2	96.1	6.4	13.3	95.9	6.0
15	8.2	97.7	3.7	9.3	97.7	4.2
16	5.0	98.7	2.3	6.0	98.7	2.7
17	3.7	99.5	1.7	2.7	99.5	1.2
18	0.9	99.8	0.4	0.7	99.8	0.3

All three parameters had an impact on the trade-off between detection rate and CPCD. This trade-off became more favorable (i.e., a higher detection rate can be achieved for a smaller increase in CPCD) when response rates increased, late-stage screening costs decreased (nurse consult), or accuracy of the questionnaire was reduced. However, the latter also resulted in lower absolute detection rates for any given cut-off.

The cut-off with the lowest CPCD changed in some of the impact scenarios. The largest shift occurred in the email invitation scenario, where it shifted to 12. Contrarily, in the scenario with 33% lower consult costs the cut-off with lowest CPCD decreased to 9. This was the same in the scenario with reduced patient response rates. Lastly, in the scenario with the largest increase in accuracy (area under the curve 90%), the cut-off with lowest CPCD increased to 11.

DISCUSSION

In this study, we set out to provide insight in the trade-off between CPCD and detection rate when choosing a cut-off on a risk questionnaire used in stepwise screening for IFG and DM2. At low cut-off scores, unnecessary GP consults and glucose tests are provided to a large number of false positive patients. Contrarily, at high cut-off scores the initial costs of sending out invitations and questionnaires are shared by a small group of detected cases due to the large number of false negative patients. Combined, these aspects resulted in a U-shaped relation between CPCD and detection rate with the lowest CPCD attained at cut-off 10. Additionally, we investigated the impact of possible changes within the framework of questionnaire based stepwise screening on the trade-off between CPCD and detection rate. Changes in patient response rates had the largest impact on the results of screening as these had a strong effect on both detection rate and CPCD. Changes in screening costs or accuracy mainly affected CPCD. In terms of CPCD, the effect of 15% more response was very similar to a reduction of consult costs by 33% for the cut-off

with lowest CPCD. However, in the increased response scenario the detection rate additionally increased by one third at this cut-off.

Increasing the cut-off from the score with the lowest CPCD would result in a lower detection rate and higher CPCD. These cut-offs can therefore be discarded as sub-optimal. All remaining cut-offs, from the lowest up to and including the one with the lowest CPCD, present a trade-off where additional detection can be gained for an increase in CPCD. However, because positive patients with a very low score are very rare, decreasing the cut-off in the lower range of cut-offs results in very little additional detection but a large increase in CPCD. It is therefore possible to identify a range of efficient cut-offs, in our case from 6 up to and including 10. Within this set, decision makers would have to determine their willingness to pay for the additional detection in order to find the optimal cut-off.

One approach to find the optimal cut-off would be to define a maximum CPCD based on an investment perspective on screening. Taking such a perspective, the maximum CPCD is determined by the average gain in net monetary benefit that can be achieved by treating the detected cases. If all cases were to have the same level of utility in being detected, the maximum CPCD would form a horizontal line in the plot of CPCD versus detection rate. In reality, however, the impact of treatment on the downstream costs and health effects differs between patients depending on their age, disease status (IFG or DM2), whether they have comorbidities, and on other patient characteristics, meaning that the maximum CPCD is likely to vary with the cut-off selected. If the entire U-curve is above the maximum CPCD curve, stepwise screening is not viable from a health economic perspective. In contrast, if the maximum CPCD curve crosses the U-curve or if the entire U-curve is below the maximum CPCD curve, the rightmost cut-off under the maximum CPCD curve leads to the largest feasible uptake. In these situations, there are two ways to approach an optimal cut-off. A cost-dominated approach would start at the cut-off with the lowest CPCD and require the decision

maker to determine whether he is willing to pay the additional CPCD for the additional cases detected when decreasing the cut-off by one point. This is then repeated until the willingness to pay is not large enough to decrease the cut-off more. Alternatively, a detection-dominated approach would start at the rightmost cut-off below the maximum CPCD curve. The cut-off is then increased as long as the reduction in CPCD is large enough to offset the reduced detection (willingness to forgo). A different perspective a decision maker could have is one of a budget constraint, in which there is a maximum amount of resources available to be allocated to screening. Taking such a perspective, the optimal cut-off can be found by calculating the total costs for screening, which, although not displayed in our analyses, can easily be obtained from the presented model. The decision maker can then simply look for the lowest cut-off for which the total costs of screening is still within his budget.

Economic considerations are important when initiating a large screening program as there will always be a considerable budget impact. Despite of this, CPCD is not yet routinely considered when evaluating different screening designs. One recent study compared a large number of different single step and stepwise screening strategies on their detection rate, false positive rate, and CPCD.[17] The authors concluded that stepwise screening using available data in GP records or self-administered questionnaires were the most feasible strategies in terms of CPCD. However, they only considered a limited number of cut-offs as alternative scenarios, and attendance rates were fully based on assumptions. The only other studies known to us to use CPCD as an outcome measure did not consider stepwise screening using a questionnaire, but rather two steps of blood glucose testing.[15,16] In one of these[16], the authors applied a similar approach to ours by assessing the detection rate and CPCD for all possible cut-off points. However, rather than summarizing the performance on these two outcomes in a single graphical presentation as we did in this study, they presented the results separately for the detection rate and CPCD, and declared the cut-off with the lowest CPCD as the optimal one. As

detecting patients in the target population is the primary aim of screening, it is highly unlikely that policy makers would indeed consider the point with lowest CPCD as optimal.

Although we presented results of a specific Dutch scenario in terms of costs and prevalences, the main findings hold for any type of stepwise screening in any target population. The results of our study therefore have a broad set of implications going beyond the realm of stepwise screening for prediabetes and DM2. First and foremost is the conclusion that selecting a cut-off without evaluating its implications in terms of CPCD and detection rate could lead to a waste of resources and missed opportunities of patients who could be treated (when a cut-off outside the efficient range is selected). Alternatively, it could lead to a possibly undesirable implicit statement about the willingness to pay for additional cases detected. Second, innovations in information technology infrastructure make it possible to do more with patient data in GP records. There is a move towards automated risk profiling within GP administration systems.[27] This changes the costs associated with the first steps of screening, similar to that in our 'email invite' cost scenario. We have demonstrated that this change has a large impact on the range of efficient cut-offs, and it is therefore advisable to revisit the cut-off decision when such changes take effect. Finally, our impact analysis provides insight in where further improvements within the strategy of stepwise screening would have the most beneficial effects on the trade-off between CPCD and detection rates. In the past, research has mainly focused on improving the accuracy of risk questionnaires through the identification of novel biomarkers that would further improve risk stratification[28], while there has been little focus on implementation aspects such as response rates. However, our analysis indicates that improving the latter has a much stronger effect on the outcomes of stepwise screening than improving the former. This suggests that funds available for research would have a higher impact when they are invested in implementation research, rather than in laboratory research.

In conclusion, policy makers should be aware that the choice of a cut-off on the first test in a stepwise screening program has direct implications on the feasibility, effectiveness, and budget impact of the program. This choice should therefore be based on an integral assessment of all these aspects rather than solely on test accuracy. The methods presented in this paper can assist policy makers to do so.

REFERENCES

1. International Diabetes Federation. *IDF Diabetes Atlas, 6th edition*. (International Diabetes Federation, 2013).
2. Paulweber, B, Valensi, P, Lindström, J, *et al.* A European evidence-based guideline for the prevention of type 2 diabetes. *Hormone and Metabolic Research* **42 Suppl 1**, S3-36 (2010).
3. Waugh, N, Scotland, G, McNamee, P, *et al.* Screening for type 2 diabetes: literature review and economic modelling. *Health Technology Assessment* **11**, (2007).
4. Simmons, RK, Echouffo-Tcheugui, JB & Griffin, SJ. Screening for type 2 diabetes: an update of the evidence. *Diabetes, Obesity and Metabolism* **12**, 838–844 (2010).
5. Waugh, N, Shyangdan, D, Taylor-Phillips, S, Suri, G & Hall, B. Screening for type 2 diabetes: a short report for the National Screening Committee. *Health Technology Assessment* **17**, 1–90 (2013).
6. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* **26**, 5–20 (2003).
7. Echouffo-Tcheugui, JB, Ali, MK, Griffin, SJ & Narayan, KMV. Screening for type 2 diabetes and dysglycemia. *Epidemiologic Reviews* **33**, 63–87 (2011).
8. Rahman, M, Simmons, RK, Harding, A-H, Wareham, NJ & Griffin, SJ. A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study. *Family Practice* **25**, 191–196 (2008).
9. Schwarz, PEH & Peltonen, M. Prevention of type 2 diabetes - lessons we have learnt for implementation. *Hormone and Metabolic Research* **39**, 636–641 (2007).
10. Buijsse, B, Simmons, RK, Griffin, SJ & Schulze, MB. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. *Epidemiologic Reviews* **33**, 46–62 (2011).
11. Schwarz, PEH, Li, J, Lindstrom, J & Tuomilehto, J. Tools for predicting the risk of type 2 diabetes in daily practice. *Hormone and Metabolic Research* **41**, 86–97 (2009).
12. Pottie, K, Jaramillo, A, Lewin, G, *et al.* Recommendations on screening for type 2 diabetes in adults. *Canadian Medical Association Journal* **184**, 1687–1696 (2012).
13. Dekker, J, Alssema, M, Janssen, P, *et al.* NHG-Standaard Het PreventieConsult module Cardiometabool. *Huisarts en Wetenschap* **54**, 138–155 (2011).
14. Schwarz, PEH, Lindström, J, Kissimova-Scarbeck, K, *et al.* The European perspective of type 2 diabetes prevention: diabetes in Europe--prevention using lifestyle, physical activity and nutritional intervention (DE-PLAN) project. *Experimental and Clinical Endocrinology & Diabetes* **116**, 167–172 (2008).

15. Zhang, P, Engelgau, MM, Valdez, R, *et al*. Costs of screening for pre-diabetes among US adults: a comparison of different screening strategies. *Diabetes Care* **26**, 2536–2542 (2003).

16. Zhang, P, Engelgau, MM, Valdez, R, *et al*. Efficient cutoff points for three screening tests for detecting undiagnosed diabetes and pre-diabetes: an economic analysis. *Diabetes Care* **28**, 1321–1325 (2005).

17. Khunti, K, Gillies, CL, Taub, NA, *et al*. A comparison of cost per case detected of screening strategies for Type 2 diabetes and impaired glucose regulation: modelling study. *Diabetes Research and Clinical Practice* **97**, 505–513 (2012).

18. Wald, NJ. Guidance on terminology. *Journal of Medical Screening* **15**, 50 (2008).

19. Alssema, M, Feskens, EJM, Bakker, SJL, *et al*. Finse vragenlijst redelijk goede voorspeller van het optreden van diabetes in Nederland. *Nederlands Tijdschrift voor Geneeskunde* **152**, 2418–2424 (2008).

20. Lambers Heerspink, HJ, Brantsma, AH, de Zeeuw, D, *et al*. Albuminuria assessed from first-morning-void urine samples versus 24-hour urine collections as a predictor of cardiovascular morbidity and mortality. *American Journal of Epidemiology* **168**, 897–905 (2008).

21. Van de Kerkhof, RM, Godefrooij, MB, Wouda, PJ, *et al*. Cardiometabole risicofactoren opgespoord met preventieconsult. *Nederlands Tijdschrift voor Geneeskunde* **154**, 1891–1897 (2010).

22. RIVM. *Chlamydia Screening Implementation Netherlands: Impact evaluation and cost effectiveness*. (2010).

23. Hakkaart- van Roijen, L., Tan, S.S., Bouwmans, C a. M. Handleiding voor kostenonderzoek Methoden en standaard kostprijzen voor economische evaluaties. *College voor Zorgverzekeringen (CVZ)* **Januari**, 1–127 (2011).

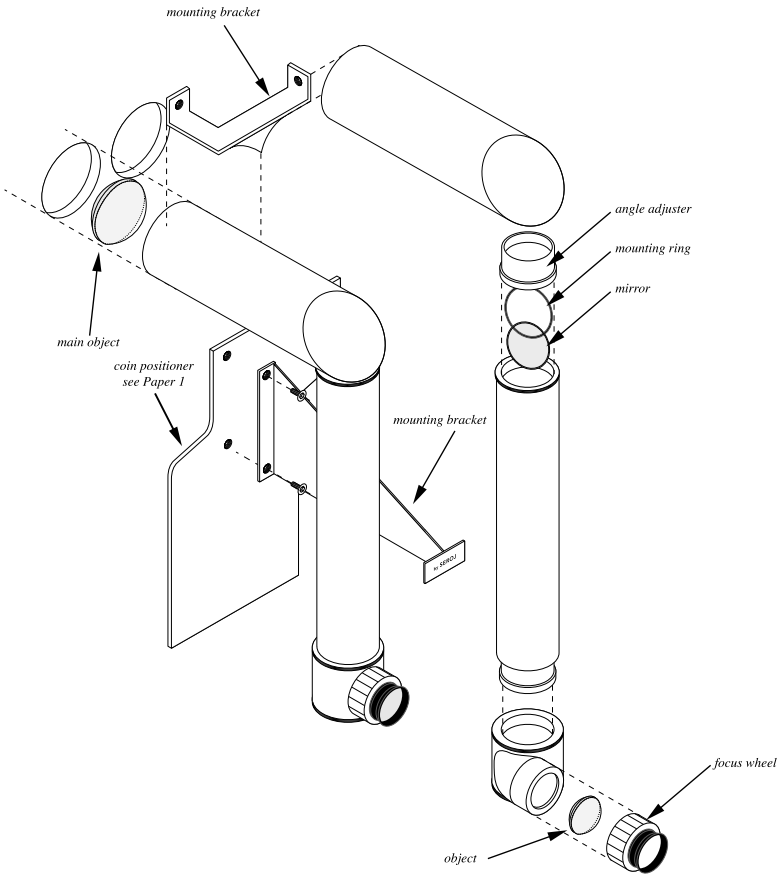
24. Vermunt, PWA, Milder, IEJ, Wielaard, F, van Oers, JAM & Westert, GP. An active strategy to identify individuals eligible for type 2 diabetes prevention by lifestyle intervention in Dutch primary care: the APHRODITE study. *Family Practice* **27**, 312–319 (2010).

25. Glümer, C, Carstensen, B, Sandbaek, A, *et al*. A Danish diabetes risk score for targeted screening: the Inter99 study. *Diabetes Care* **27**, 727–733 (2004).

26. Baan, CA, Ruige, JB, Stolk, RP, *et al*. Performance of a Predictive Model to Identify Undiagnosed Diabetes in a Health Care Setting. *Diabetes Care* **22**, 213–219 (1999).

27. Gray, LJ, Khunti, K, Edwardson, C, *et al*. Implementation of the automated Leicester Practice Risk Score in two diabetes prevention trials provides a high yield of people with abnormal glucose tolerance. *Diabetologia* **55**, 3238–3244 (2012).

28. Echouffo-Tcheugui, JB, Dieffenbach, SD & Kengne, AP. Added value of novel circulating and genetic biomarkers in type 2 diabetes prediction: a systematic review. *Diabetes Research and Clinical Practice* **101**, 255–269 (2013).


mounting bracket
angle adjuster
mounting ring
mirror
main object
coin positioner
see Paper 1
mounting bracket
focus wheel
object

Chapter 6

The impact of short lead time for diabetes on the cost-effectiveness of treating patients Identified in prediabetes screening

Under review for publication

ABSTRACT

BACKGROUND: Providing lifestyle interventions to patients with impaired glucose regulation (IGR) has been shown to be an effective strategy to reduce the large societal burden of disease caused by type 2 diabetes (DM2). As IGR is asymptomatic this requires a screening effort, which will inadvertently also identify undiagnosed DM2 patients. Given current treatment protocols, early treatment of the latter category of patients is however not cost-effective. Our aim is to investigate to what extent the costs of treating newly detected DM2 patients can be offset by providing lifestyle interventions to IGR patients.

METHODS: A model-based evaluation was conducted to estimate the difference in downstream costs and health effects as a result of the provision of lifestyle intervention to IGR patients and standard glycemic control to DM2 patients that would be identified as a result of a structured screening program for IGR in the general population. The characteristics of the treatment population identified through such a screening were taken from a large general population cohort in The Netherlands.

RESULTS: In our base-case analysis, treating patients identified in a population wide IGR screening resulted in an incremental downstream costs of €827.34 and 0.03509 Quality Adjusted Life Years (QALYs), resulting in an incremental cost-effectiveness ratio (ICER) of €23,576 per QALY. Screening started to become cost-effective (i.e., ICER < 20k per QALY) when at least 2 IGR patients are enrolled in an intervention for every DM2 patient identified.

CONCLUSION: The favorable effects of providing lifestyle intervention in IGR patients are insufficient to offset the additional costs of early treatment in DM2 patients. The treatment of patients identified in a screening for IGR may therefore not be considered cost-effective.

INTRODUCTION

Reducing the growing burden of disease and societal costs of type 2 diabetes (DM2) has increasingly received attention over the past decade. Numerous studies have indicated that the prevention or delay of onset of DM2 in patients with impaired glucose regulation (IGR) could be achieved at acceptable costs by means of lifestyle intervention programs.[1–3] IGR is a condition in which fasting and/or postprandial blood glucose levels are abnormally elevated, but are still below levels defined as established DM2. As IGR is an asymptomatic condition, identifying individuals with IGR would require population wide screening. Since both conditions are defined by the same diagnostic test, screening would also identify previously undiagnosed DM2 patients. Contrary to the case for IGR, there is considerable doubt that the identification and treatment of screen detected DM2: in their recent report, patients under current protocols is cost-effective. It has been found that the lead time of screening (the period between diagnosis as a result of screening and otherwise clinical diagnosis) is relatively short.[4,5] This short period of additional treatment has no significant effect on mortality or cardiovascular complications.[6–8] These insights caused the UK national screening committee to take a more negative stance on screening for DM2 compared to their previous report in 2007.[9] Although the standard treatment for DM2 is not particularly costly, the lack of effects are likely to lead to an unfavorable cost-effectiveness ratio.

A policy dilemma thus arises, as it is currently not clear whether the unfavorable cost-effectiveness of treating screen detected DM2 patients can be offset by the favorable cost-effectiveness of treating patients with IGR. Nevertheless, screening for IGR has already found its way into several guidelines.[10–13] Prior to broad implementation, however, insight in the cost-effectiveness of IGR screening is urgently needed. Accordingly, we set out to examine the combined cost-effectiveness of providing lifestyle intervention to patients with IGR and standard care to screen detected DM2 patients,

while incorporating the recent evidence of very limited effects of early treatment in DM2 patients. Additionally, we examine the effects of various combinations of lead time and effects of early treatment in DM2 patients on the cost-effectiveness for different proportions of IGR and DM2 patients in the treatment population (that is, the population of patients enrolled in a treatment as a result of the screening program).

METHODS

Decision context

We conducted a model-based evaluation to estimate the difference in downstream costs and health effects as a result of the provision of lifestyle intervention to IGR patients and standard glycemic control to DM2 patients that would be identified as a result of a structured screening program for IGR in the general population. The cost perspective taken was that of the healthcare system. We assumed that the screening would target the general population between ages 45 and 75, as suggested in a number of existing guidelines.[10,13] The treatment population in our analysis was based on data from the PREVEND Groningen study, a cohort drawn from the general population in the city of Groningen in the Netherlands.[14] In this cohort, IGR and DM2: were defined on fasting plasma glucose levels (IGR: >=6.1 and <7.0 mmol/L, DM2 >=7.0 mmol/L).

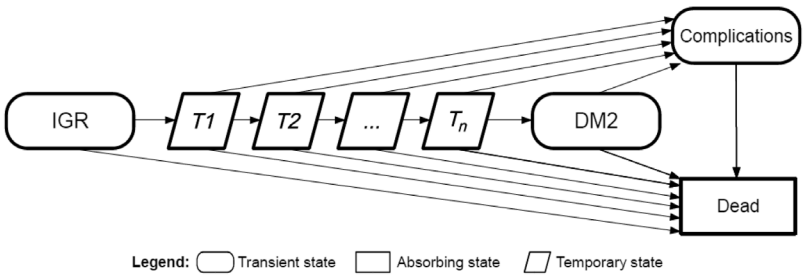


Figure 1: Disease progression model

Model structure

A patient level non-homogeneous discrete time Markov model[15] (i.e., a discrete time Markov model whose transition probability matrix depends on patient characteristics and is updated each model cycle) of the disease progression of patients was used to estimate the change in costs and quality adjusted life years (QALYs) as a result of lifestyle intervention or treatment after screening. The model has three alive states (IGR, DM2, and complications) and a dead state (Figure 1). The alive states reflect the blood glucose level of the patient (IGR or DM2) and whether diabetic complications have occurred. Only progressive transitions (from IGR to DM2 and from DM2 to complications) are allowed. The rationale behind this is that, in practice, once a patient reaches a more severe disease state, he or she will incur the additional healthcare costs, reduced health related quality of life, and increased risk of death even if an improvement in the glucose metabolism, i.e. a reversion to the IGR state, would be achieved.[16] From each of the alive states, patients can die.

The DM2 state was subdivided in a treated DM2 state and several temporary states indicating untreated DM2. Temporary states are states in which a patient can only reside for one cycle that we used to model the predefined time spend with undiagnosed DM2. A recent study estimated the time between onset and diagnosis of DM2 at 4 to 6 years.[5] We therefore used 4 temporary states for untreated DM2 in the base case scenario, to represent 4 years between onset and clinical diagnosis. We assumed that during the first round of screening, patients would on average be identified halfway the duration of untreated DM2, thus 2 years after onset. In an analysis, IGR patients always started in the IGR state. DM2 patients either started in the DM2 state (treatment scenario) or in the 3rd temporary state (no treatment scenario).

Transition probabilities

All parameters of the disease progression model are listed in Table 1. The transition probabilities from the transient health states to the dead state were dependent on the age and gender of the patient and were updated after each model cycle, whereas the transition probabilities between the transient health states were assumed to be time-independent and constant across all patients. In particular, we used Dutch gender specific life-table data per 5-year age group for the transition of both the IGR and DM2 state to the death state to reflect the recent findings that well controlled DM2 patients have a death rate similar to the general population.[17,18] For the complication state, we increased this death rate with a relative risk of 2.13.[19] The transition from the IGR to the DM2 state was based on the 7-year incidence of DM2 in impaired fasting glucose patients in the PREVEND cohort. Lastly, the transition from DM2 to the complication state was estimated by fitting a competing risks model on the event rates in the ROSSO cohort.[20]

Lifestyle intervention costs and effects

We modeled a lifestyle intervention as implemented in the Finnish Diabetes Prevention Study (DPS), both for costs and effects.[21] We assumed that there was a steady effect of the intervention during the period the intervention was provided, and a linear decline of its effect once the intervention was stopped. Based on the median intervention duration in the DPS, we modeled an intervention period of 4 years. The DPS had a total follow up of 8 years, at the end of which an intervention effect was still observable.[21] We therefore assumed an 8-year linear decline of effects after stopping the intervention, resulting in a total effect duration of 12 years. We used the risk reduction found in the DPS at 8 years and fitted a steady intervention effect of 4 years and a linear decline of an additional 8 years on this figure. Costs of the intervention were based on the breakdown of intervention activities in the DPS[25], for which Dutch unit costs were taken from the SLIM study.[22] These costs were counted

Table 1: Parameter data

Parameter	Value		Source
Transition rates alive states			
IFG to DM2	0.0350		PREVEND cohort data
DM2 to complications	0.0196		ROSSO4[20]
Death rates	**Male**	**Female**	
45 – 49	0.0018	0.0014	
50 – 54	0.0031	0.0024	
55 – 59	0.0050	0.0039	
60 – 64	0.0085	0.0060	
65 – 69	0.0134	0.0090	
70 – 74	0.0221	0.0143	Dutch life table data[17]
75 – 79	0.0381	0.0242	
80 – 84	0.0708	0.0467	
85 – 89	0.1273	0.0918	
90 – 94	0.2177	0.1756	
95 >	0.3582	0.3219	
Relative risk of dying when in Complication state	2.13		Gillies (2008)[19]
Intervention effects			
Relative risk of developing DM2	0.53		Lindstrom (2006)[21]
Duration steady intervention effect	4 years		
Duration of linear decline intervention effects	8 years		
Costs			
Cost of intervention 1st year	€ 387.32		Lindstrom (2006)[21]
Cost of intervention 2nd to 4th year	€ 226.38		Lindstrom (2006)[21] + Jacobs-van der Bruggen (2007)[22]
Costs DM2 state	€ 1,643.83		Mortaz (2011)[16]
Costs complications state	€ 4,230.08		
Utilities			
IFG	0.80		Mortaz (2011)[16], Tapp (2010)[23]
Undiagnosed DM2	0.79		Mortaz (2011)[23]
DM2	0.79		
Complications	0.70		UKPDS[24]

GP = General Practitioner; IFG = Impaired Fasting Glucose; DM2 = type 2 diabetes

in both the IGR state and the temporary states, should a patient progress from IGR to undiagnosed DM2 during the 4 year intervention period. Once DM2 was diagnosed, these patients were again assumed to be put on standard glycemic control like all other patients residing in the treated DM2 state.

State costs and utilities

The cost and utility estimates attached to the Markov model's transient health states were taken from previously conducted economic evaluations (see Table 1). The model estimates the difference in direct healthcare costs between the two scenarios. Costs were therefore taken relative to the IGR state. Patients with undiagnosed DM2 (temporary states) were assumed to have the same costs as IGR patients. All costs were converted and inflated to 2015 Euros. We used the utilities presented by Mortaz et al.[16] for the treated and untreated DM2 states and those presented in UKPDS[24] for the complication state. The largest assessment of health related quality of life in IGR patients was conducted within the Diabetes Prevention Program, though this included only patients with impaired glucose tolerance.[26] We used the average of the gender specific utility values in our model. By doing this, the difference between the utility weights of the IGR and DM2 state is in concordance with the finding that the health related quality of life in IGR patients is just slightly higher than in newly detected DM2 patients.[23]

Disease prevalence and patient heterogeneity

The proportion of IGR patients in the treatment population for the base case was taken from the PREVEND cohort, and was 60.4%. A joint distribution of age and gender for each disease group was also based on those patients in the PREVEND cohort. This joint distribution was factorized as the conditional distribution of age given gender and the marginal distribution of gender (i.e. the proportion of each gender in both disease groups). The conditional distributions of age given gender were estimated by fitting a

Weibull distribution to the data of the four groups and are shown in Figure 2. The percentage of males in the IGR and DM2 groups were 60.9% and 61.7%, respectively.

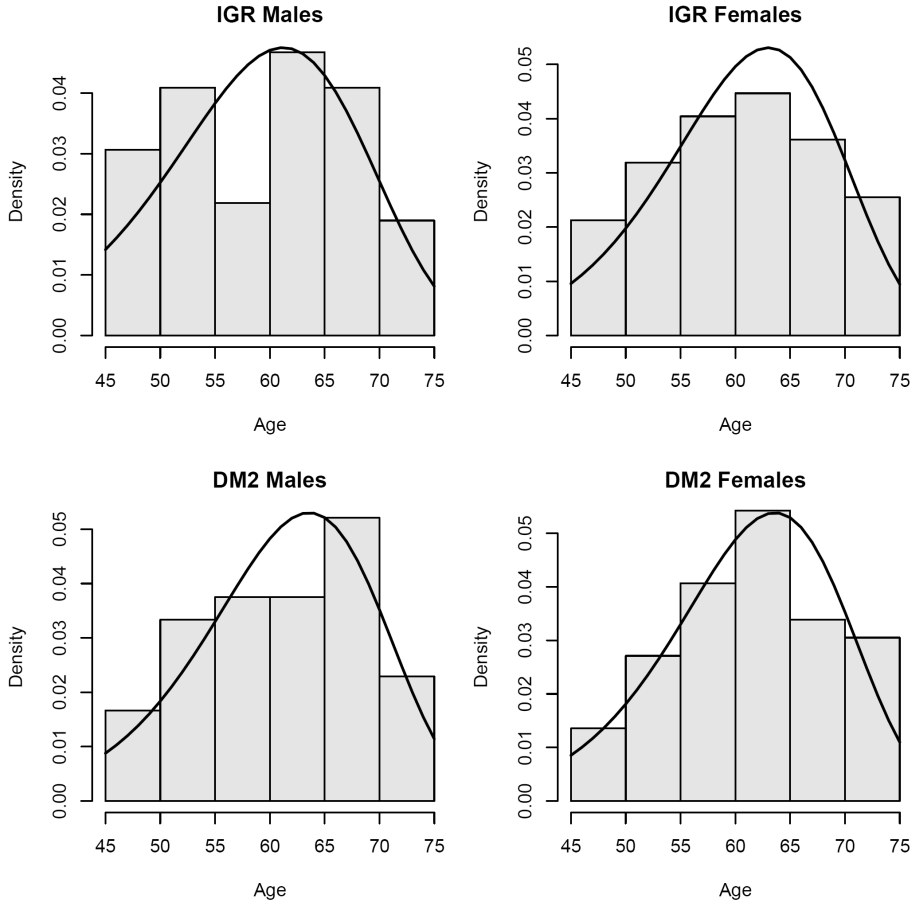


Figure 2: Conditional distributions of age given gender. Bars show observed density in PREVEND cohort; line shows Weibull distribution fitted to PREVEND data

Cost effectiveness

The difference in downstream costs and QALYs were calculated separately for IGR and DM2 patients. To obtain these outcomes for both groups, all possible

combinations of age and gender within the age interval of the treatment population (45 to 75) were entered into the model once (i.e. 45 year old male, 45 year old female, 46 year old male, etc.) and subsequently were weighed with the probability of that combination as determined from the joint distribution of age and gender. Lastly, the proportion of IGR and DM2 patients in the treatment cohort was used to weigh the outcomes for the separate groups to obtain the total cost-effectiveness.

For each patient, the model was evaluated in cycles of one year over a lifetime horizon. At the end of each cycle, the costs and effects attached to each state were multiplied with the probability of the patient being in that state. The model was advanced one cycle by multiplying a matrix containing the probabilities of the patient residing in each state with a matrix containing the transition probabilities between states. The evaluation for a patient ended when there was a 99.5% probability that the patient had died. Costs and QALYs were both discounted at 3%.

Sensitivity analysis

There is considerable uncertainty regarding the exact lead time of DM2 screening and the effect earlier treatment has on the disease progression of DM2 patients. Moreover, it is likely that there is a correlation between lead time and early treatment effects. Lead time represents the additional period of treatment in a screening setting compared to a non-screening setting. When this period of additional treatment is short, lesser effects are to be expected on the course of the disease than when it is long. Currently, no direct evidence on the relation between the length of lead time and treatment effects is available. Consequently, we performed a 2-way sensitivity analysis where these parameters are varied independently as well as simultaneously. Additionally, we examined how different proportions of IGR and DM2 patients in the treatment population impact the cost-effectiveness of treatment in a 3-way sensitivity analysis.

In the 2-way sensitivity analysis, lead time was varied between the base case value of 2 years and two additional scenarios of 3 and 4 years, in order to represent the upper estimate of the study from which we took the estimate for the base case, as well as the evidence of another study that estimated lead time at 3.3 years.[4,5] This increase in lead time was modeled by increasing the number of temporary states in the model to 6 and 8, respectively. Effects of early treatment on the incidence of complications in DM2 patients was absent in the base case. In the sensitivity analysis, we added two scenarios where early treatment led to a relative risk for complications of 0.9 and 0.8. These effects were only modeled for those DM2 patients for which treatment is commenced earlier as a result of screening and not for IGR patients that develop DM2 later.

The effect of different proportions of IGR and DM2 in the treatment population was analyzed by weighing the outcomes of the 2-way sensitivity analysis for each disease with the proportion of that disease in the total treatment population. The analysis ranged from 40% IGR – 60% DM2 to 100% IGR – 0% DM2. We specifically explored the assumption that lead time and effects of early treatment are related by comparing the three scenarios where both parameters simultaneously increased (lead time 2 years – relative risk 1.0, lead time 3 years – relative risk 0.9, lead time 4 year – relative risk 0.8).

RESULTS

Base case analysis

As such, enrolling screen detected IGR patients in a lifestyle intervention program appeared cost saving in our disease progression model. The difference in downstream costs of screening compared to no screening was -€664.79 while adding 0.05811 quality adjusted life year per patient. On the other hand, treating screen detected DM2 patients came at an additional cost of €3102.37 per patient without adding any QALYs. Weighing these results with the prevalence of each group in the treatment population led to an expected downstream costs and effects of screening compared with no screening of

€827.34 and 0.03509 QALY, respectively, resulting in an incremental cost-effectiveness ratio (ICER) of €23,576.26 per QALY.

Sensitivity analysis

The resulting ICERs of all lead time and treatment effect combinations are shown in Table 2. When considered in isolation, changes in the lead time did not change the effects of treatment in both groups. A longer lead time reduced the average cost-saving in the IGR group to -€507.08 at 3 years and -€365.92 at 4 years. At the same time, it increased the average costs associated with earlier treatment in the DM2 group to €4523.03 and €5861.50 at 3 and 4 years, respectively. As a result, the ICER for combined intervention and treatment increased drastically (Table 2, first column).

Table 2: Results of the 2-way sensitivity analysis (cost per quality adjusted life years gained in €)

Lead time (years)	Relative risk on developing complications		
	1.0	**0.9**	**0.8**
2	23,576.26	11,027.38	5,657.21
3	42,325.56	22,087.58	13,418.53
4	59,862.45	32,432.58	20,677.98

The reduced incidence of complications as a result of early treatment effects led to a reduction in costs and an increase in effects in the DM2 group. When lead time was kept at 2 years, a relative risk of 0.9 as a result of treatment led to an average cost of €2669.80 and a gain of 0.06159 QALY. A relative risk of 0.8 decreased costs further to €2224.39 and DM2 patients gained on average 0.12543 QALY. The ICERs of these scenarios were therefore lower than in the base case (Table 2, first row).

The proportion of IGR in the total treatment population had a very strong impact on the combined cost-effectiveness (Fig 3, left panel). In the extreme

case of a treatment population without DM2 patients, screening was dominant in all scenarios. The impact of the early treatment effect in DM2 patients was stronger when a larger proportion of the treatment population consisted of DM2 patients. For sake of clarity, the scenarios with 3 years lead time are not plotted in Fig 3. The ICER value in those scenarios was exactly in between the ICER value for the 2 and 4 years lead time scenarios.

The effect of simultaneously increasing lead time and effects of early treatment on the ICER depended on the proportion of IGR in the treatment population (Fig 3, right panel). When the proportion of IGR in the treatment population exceeded 65.5%, increased lead time and early treatment effect led to higher costs per QALY, as the negative effect of lead time on the cost saving of lifestyle intervention in IGR patients outweighed the benefits of less treatment costs and additional treatment effects in DM2 patients. This was exactly the other way around when the proportion of IGR in the treatment population was below 62.7%.

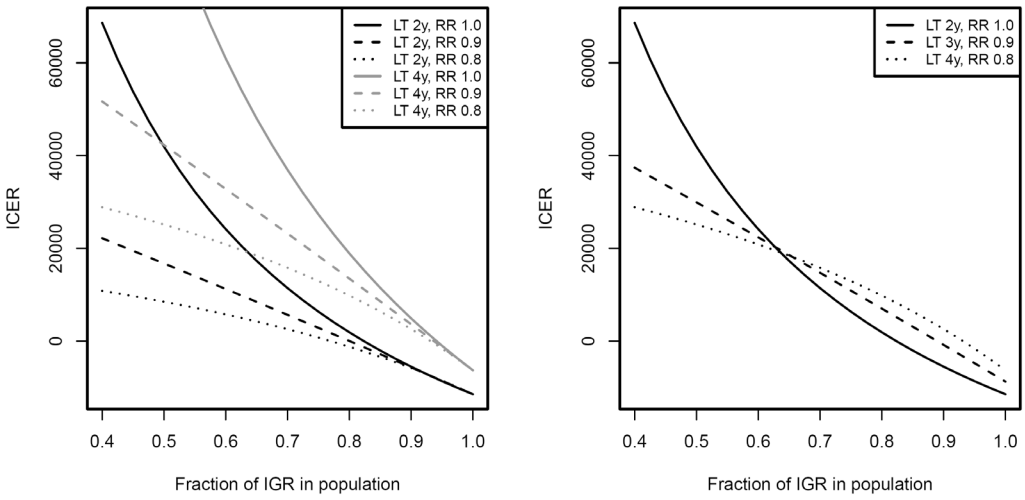


Figure 3: Outcomes of the 3-way sensitivity analysis . LT: lead time, RR: relative risk.

DISCUSSION

We estimated the cost-effectiveness of providing lifestyle interventions to IGR patients and standard glycemic control to newly detected DM2 patients identified through an IGR screening program. The resulting cost-effectiveness ratio of €23576.26 per QALY is considered not to be cost-effective by Dutch standards, which apply a €20,000 per QALY threshold for preventive interventions.[27] The results from our sensitivity analysis showed that treating patients identified in an IGR screening program becomes cost-effective when at least 2 IGR patients are enrolled in an intervention for every DM2 patient diagnosed in screening.

The results are very sensitive to the proportion of IGR patients in the treatment population. As lifestyle intervention in IGR patients is cost saving, a larger proportion of IGR patients makes the combined treatment more cost-effective. Furthermore, the results are especially sensitive to the effects of early treatment in DM2 patients when a larger proportion of the treatment population consists of DM2 patients. Lastly, longer lead time of DM2 led to a higher cost per QALY. As lead time increases, the average cost of treatment of DM2 decreases due to additional years without treatment. As a result, there are less costs to offset by preventing the onset of DM2 through lifestyle intervention in IGR patients, which therefore becomes less cost-saving.

Our finding that providing a lifestyle intervention to IGR patients is cost saving is in concordance with previous studies.[22,28,29] However, previous modeling studies assumed a positive effect on health related quality of life or a reduced complication incidence due to early treatment of screen detected DM2.[16,19,30–32] Instead of being optimistic about these effects, we took a reverse approach by assuming nil effects, and assessing the impact of possible effects in a sensitivity analysis. As a result, we conclude that lifestyle intervention in IGR and early treatment of DM2 patient has a far less favorable cost per QALY ratio than previously reported. Previous modeling studies predominantly

relied on data gathered between two and three decades ago. Since those studies were conducted, lead time of DM2 has shorted and treatment of clinically detected DM2 has improved. Using evidence of those early studies therefore leads to an overly optimistic estimate of additional treatment effects in screen detected DM2 that would not be obtainable if screening would be initiated today. Our sensitivity analyses confirm the results of other studies that if early treatment of DM2 through screening would result in a substantial reduction of DM2 related complications, screening would become cost-effective. This is both the case when these effects are persistent over the entire duration of DM2, as some studies assumed[16,19] (results shown), as well as when these effects mainly stem from an increased risk during the period of undiagnosed and untreated DM2, as others implemented[32] (results not shown).

Notably, our current study did not include the screening program that precedes the provision of lifestyle intervention and treatment. As the cost of screening must be offset by the benefits of subsequent treatment, a cost-effective treatment in screen positive patients is a prerequisite for the initiation of such a screening program. For this reason, the current study was conducted to assess whether this prerequisite is fulfilled. Aside from the cost aspect, the screening strategy has an influence on the case mix of the treatment population. Our results provide an estimation of the expected treatment benefits under the assumption of population wide screening with a blood glucose test. Most guidelines, however, consider a stepwise approach where a high-risk preselection is made using a risk questionnaire. Such a strategy has an impact on the characteristics of the treatment population through the selection of those characteristics that are included as a risk factor. In particular, as age is a frequently used risk factor in questionnaires, this is likely to result in a higher proportion of older patients in the treatment population. Because lifestyle intervention become less cost-effective as age increases, this has an impact on the overall cost-effectiveness of intervention and treatment. Additionally, when screening is performed repeatedly, it has an impact on the prevalence of IGR and DM2 in the population. The first rounds

of screening will identify a large number of patients. This will eventually reach a steady state where the incidence of new cases equals the uptake of screening. In this steady state, the average age of the treatment population will be younger and the proportion of IGR patients will be higher than in the current cross-sectional data. Both aspects lead to a better cost-effectiveness of screening. However, in our study we only explored the impact of the latter.

As with any modeling study, our study also has to deal with uncertainty regarding the parameter values used. The parameters essential to our research question were included in the sensitivity analysis, but uncertainty is not limited to these parameters. In general, parameter uncertainty is often addressed by conducting a probabilistic sensitivity analysis, in which the uncertainty surrounding the parameters is expressed as a probability distribution. By performing multiple iterations with different parameter values drawn from these distributions, such a sensitivity analysis provides an estimate of the total effect of the parameter uncertainty. As our aim was to specifically assess the impact of short lead time, early treatment effects in DM2, and the proportion of IGR and DM2 patients in the treatment population, an estimate of the overall uncertainty would provide limited additional insight to answer our research question.

To conclude, we have shown that the combined treatment of screen detected IGR and DM2 cannot be considered cost-effective for the Dutch population under current conditions. We are therefore of the opinion that the apparent enthusiasm for IGR screening should be tempered and guidelines.

REFERENCES

1. Gillies, CL, Abrams, KR, Lambert, PC, *et al*. Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *BMJ (Clinical research ed.)* **334**, 299–308 (2007).
2. Portero McLellan, KC, Wyne, K, Villagomez, ET & Hsueh, WA. Therapeutic interventions to reduce the risk of progression from prediabetes to type 2 diabetes mellitus. *Therapeutics and Clinical Risk Management* **10**, 173–188 (2014).
3. Li, R, Qu, S, Zhang, P, *et al*. Economic evaluation of combined diet and physical activity promotion programs to prevent type 2 diabetes among persons at increased risk: A systematic review for the community preventive services task force. *Annals of Internal Medicine* **163**, 452–460 (2015).
4. Rahman, M, Simmons, RK, Hennings, SH, Wareham, NJ & Griffin, SJ. How much does screening bring forward the diagnosis of type 2 diabetes and reduce complications? Twelve year follow-up of the Ely cohort. *Diabetologia* **55**, 1651–1659 (2012).
5. Porta, M, Curletto, G, Cipullo, D, *et al*. Estimating the delay between onset and diagnosis of type 2 diabetes from the time course of retinopathy prevalence. *Diabetes Care* **37**, 1668–1674 (2014).
6. Sawicki, PT. Screening for diabetes: hope and despair. *Diabetologia* **55**, 1568–1571 (2012).
7. Simmons, RK, Echouffo-Tcheugui, JB, Sharp, SJ, *et al*. Screening for type 2 diabetes and population mortality over 10 years (ADDITION-Cambridge): a cluster-randomised controlled trial. *Lancet* **380**, 1741–1758 (2012).
8. Echouffo-Tcheugui, JB, Simmons, RK, Prevost, AT, *et al*. Long-Term Effect of Population Screening for Diabetes on Cardiovascular Morbidity, Self-Rated Health, and Health Behavior. *The Annals of Family Medicine* **13**, 149–157 (2015).
9. Waugh, N, Shyangdan, D, Taylor-Phillips, S, Suri, G & Hall, B. Screening for type 2 diabetes: a short report for the National Screening Committee. *Health Technology Assessment* **17**, 1–90 (2013).
10. American Diabetes Association. Standards of Medical Care in Diabetes - 2016. *Diabetes Care* **39**, 1–112 (2016).
11. Ridley, J, Ischayek, A, Dubey, V & Iglar, K. Adult health checkup - Update on the Preventive Care Checklist Form. *Canadian Family Physician* **62**, 307–313 (2016).
12. Schwarz, PEH, Lindström, J, Kissimova-Scarbeck, K, *et al*. The European perspective of type 2 diabetes prevention: diabetes in Europe--prevention using lifestyle, physical activity and nutritional intervention (DE-PLAN) project. *Experimental and Clinical Endocrinology & Diabetes* **116**, 167–172 (2008).

13. Rutten, GEHM, De Grauw, WJC, Nijpels, G, *et al.* NHG-standaard Diabetes Mellitus type 2 (derde herziening). *Huisarts en Wetenschap* **56**, 512–525 (2013).

14. Lambers Heerspink, HJ, Brantsma, AH, de Zeeuw, D, *et al.* Albuminuria assessed from first-morning-void urine samples versus 24-hour urine collections as a predictor of cardiovascular morbidity and mortality. *American Journal of Epidemiology* **168**, 897–905 (2008).

15. Commenges, D. Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research* **11**, 167–182 (2002).

16. Mortaz, S, Wessman, C, Duncan, R, Gray, R & Badawi, A. Impact of screening and early detection of impaired fasting glucose tolerance and type 2 diabetes in Canada: a Markov model simulation. *ClinicoEconomics and Outcomes Research* **4**, 91–77 (2012).

17. Centraal Bureau voor de Statistiek. Sterfte; geslacht, leeftijd (op 31 december) en burgerlijke staat. (2015).

18. Lutgers, HL, Gerrits, EG, Sluiter, WJ, *et al.* Life expectancy in a large cohort of type 2 diabetes patients treated in primary care (ZODIAC-10). *PLOS ONE* **4**, 1–7 (2009).

19. Gillies, CL, Lambert, PC, Abrams, KR, *et al.* Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *BMJ (Clinical research ed.)* **336**, 1180–1191 (2008).

20. Martin, S, Schramm, W, Schneider, B, *et al.* Epidemiology of complications and total treatment costs from diagnosis of Type 2 diabetes in Germany (ROSSO 4). *Experimental and Clinical Endocrinology & Diabetes* **115**, 495–501 (2007).

21. Lindström, J, Ilanne-Parikka, P & Peltonen, M. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *The Lancet* **368**, 1673–1679 (2006).

22. Jacobs-van der Bruggen, M, Bos, G, Bemelmans, WJ, *et al.* Lifestyle interventions are cost-effective in people with different levels of diabetes risk: results from a modeling study. *Diabetes Care* **30**, 128–134 (2007).

23. Tapp, R, O'Neill, A, Shaw, J, Zimmet, PZ & Oldenburg, B. Is There a Link Between Components of Health-Related Functioning and Incident Impaired Glucose Metabolism and Type 2? *Diabetes Care* **33**, 757–762 (2010).

24. Clarke, P, Gray, A, Briggs, A & Farmer, A. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* **47**, 1747–1759 (2004).

25. Lindström, J, Louheranta, A, Mannelin, M, *et al.* The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care* **26**, 3230–3236 (2003).

26. Herman, WH, Hoerger, TJ, Brandle, M, *et al.* The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. *Annals of Internal Medicine* **142**, 323–332 (2005).

27. Smulders, Y & Thijs, A. Kostprijs per gewonnen levensjaar: trends en tegenstrijdigheden. *Nederlands Tijdschrift voor Geneeskunde* **150**, 2467–2470 (2006).

28. Neumann, A, Schwarz, P & Lindholm, L. Estimating the cost-effectiveness of lifestyle intervention programmes to prevent diabetes based on an example from Germany: Markov modelling. *Cost Effectiveness and Resource Allocation* **9**, 1–13 (2011).

29. Lindgren, P & Lindstrom, J. Lifestyle intervention to prevent diabetes in men and women with impaired glucose tolerance is cost-effective. *International Journal of Technology Assessment in Health Care* **23**, 177–183 (2007).

30. Kahn, R, Alperin, P, Eddy, D, *et al.* Age at initiation and frequency of screening to detect type 2 diabetes: a cost-effectiveness analysis. *Lancet* **375**, 1365–1374 (2010).

31. Chatterjee, R, Narayan, KMV, Lipscomb, J & Phillips, LS. Screening adults for pre-diabetes and diabetes may be cost-saving. *Diabetes Care* **33**, 1484–1490 (2010).

32. Herman, WH, Ye, W, Griffin, SJ, *et al.* Early detection and treatment of type 2 diabetes reduce cardiovascular morbidity and mortality: A simulation of the results of the Anglo-Danish-Dutch study of intensive treatment in people with screen-detected diabetes in primary care (ADDITION-Europe). *Diabetes Care* **38**, 1449–1455 (2015).

Chapter 7

General discussion

As a public-private translational research project, the aim of Center for Translational Molecular Medicine (CTMM) PREdiction and early diagnosis of DIabetes and diabetes-related Cardiovascular Complications (PREDICCt) project was not to merely generate new fundamental knowledge on molecular diagnostics for type-2 diabetes mellitus (DM2), but rather to develop clinical innovations that positively impact society by increasing length and quality of life.[1] This aim should be taken into account in the many decisions that have to be made from the initial investment until final market access of new technologies. Our project set out to conduct analyses to support this decision-making. We aimed to provide insight into the likely impact of the PREDICCt project on length and quality of life and costs, both of improved diagnostic and prognostic capabilities in general, as well as specific biomarker candidates. Our first finding was that, at the time of the start of this project, such research was rare and appropriate methods were lacking. As a result, we developed new methods for the early economic evaluation of biomarkers in translational research projects where we saw the need for them. In large parts, these methods built upon existing methods of health technology assessment and economic evaluation of medical technologies. By applying the newly developed methods to assess the work done within the PREDICCt project, we were able to provide insights on the potential clinical and economic impact of this work as the project went along. This chapter starts with a summary and discussion of the results of our work per research objective. Additionally, we discuss the general limitations related to our work. Finally, the chapter closes with an overview of the lessons learned during our research for several involved stakeholders.

POTENTIAL CLINICAL AND ECONOMIC IMPACT OF THE CTMM PREDICCT PROJECT

In its business plan, CTMM formulated ambitious objectives. In the area of health, it aimed to increase life expectancy and improve the quality of life of the Dutch population. More specifically, the aim was to reduce mortality rates

of the most prevalent cancer types by 10% to 30% by 2019.[1] It also formulated the specific objective to realize an annual 1 billion savings in Dutch health care costs by 2019.[1] Apart from objectives in the health and healthcare system areas, CTMM aimed to create a positive impact on the Dutch economy by strengthening the health technology sector. Lastly, it aimed to strengthen the Dutch academic institutions in the field of molecular medicine.

Our primary research objective was to assess the clinical and economic value of the biomarkers and biomarker-based technologies that were developed within the CTMM PREDICCt project. The first analysis performed was to support the investment decision of the remaining project funds. At the start of the project a strategic reserve was made which was to be allocated halfway the project, with the aim to further strengthen the most promising research lines. At the time of our first analysis, the project had not resulted in tangible research output in the form of candidate biomarkers or biomarker-based products. We therefore set out to prioritize research endeavors based on their overall potential to achieve the societal objectives of the PREDICCt project, as stated in the project's business plan. As the core aim of the PREDICCt project was to enhance the possibilities for prevention of DM2 and associated complications, we sought to prioritize between four different prevention strategies; primary prevention, secondary prevention, tertiary prevention of macrovascular complications, and tertiary prevention of microvascular complications (**Chapter 2**). Our analyses indicated that developing a biomarker technology applied to the general population to identify those with undiagnosed DM2 in order to initiate treatment sooner (i.e., secondary prevention) was the option that was least likely to contribute to achieving the goals set forth by CTMM. This was to a large extent due to the limited effects of earlier treatment of screen-detected diabetes on downstream healthcare costs and gain in quality-adjusted survival. Additionally, as DM2 is defined based on blood glucose tests, it is highly unlikely that a new biomarker-based technology will be used as the sole diagnostic test to identify patients for secondary prevention. A blood glucose test remains necessary

as long as DM2 is defined on that measure. As a result, novel biomarker technologies can only be used as a low cost or minimally invasive tool to select patients for blood glucose testing. Numerous very inexpensive and reasonably accurate anthropometric risk scores are already available for this purpose.[2] It is unlikely that a novel biomarker-based technology can present a cost-effective improvement over these risk scores. The attractiveness of the other strategies depends strongly on the decision makers preferences. If a large impact on downstream medical costs and quality-adjusted survival is prioritized, primary prevention is the best alternative. In case having an innovation developed and implemented in clinical care within the stated 10-year time span is prioritized, then the focus should be put on developing biomarker technologies for tertiary prevention. Primary prevention has the possibility to reduce the health and cost burden of both diabetes and all complications together, while tertiary prevention can only reduce the burden of complications. On the other hand, implementing a biomarker-based innovation in primary prevention requires the initiation of a new screening program in the general population, whereas the target population for tertiary prevention is already in contact with healthcare providers. No meaningful difference was found between the attractiveness of tertiary prevention of microvascular or macrovascular complications. At the time we conducted our analysis, secondary prevention was still pursued within the PREDICCt project. Based on our findings, the remaining resources were predominantly allocated to research efforts in the field of tertiary prevention.

The PREDICCt project delivered, amongst others, a set of biomarkers to increase the accuracy of macrovascular risk prediction in DM2 patients.[3] We sought to identify a clinical application for these biomarkers and estimate their commercial headroom (**Chapter 3**). We found that the most likely and quite possibly the only application of these biomarkers is to identify DM2 patients with a low cardiovascular risk, and, subsequently, refrain from prescribing the standard statin treatment. Currently, all DM2 patients are classified in the highest cardiovascular risk category, and consequently, statins are indicated

for all DM2 patients.[4,5] This strategy does not account for the wide variety of cardiovascular risk in the DM2 population.[6] Our analysis indicated that withholding statins to DM2 patients with a low cardiovascular risk (10-year risk below 10%) will lead to one additional cardiovascular disease (CVD) event in every 75 patients. The maximum headroom of the biomarkers in this application was €119.09 in case the willingness to accept for one additional CVD case was €0, which is almost certainly not a realistic case. The headroom reduced to €0 when the willingness to accept for one additional CVD case exceeded €15,614. Thus, it is reasonable to assume that there is no commercial headroom for the PREDICCt biomarkers in this application. Investments in the further development of these biomarkers are therefore not advisable.

In addition to the analysis of biomarkers for the tertiary prevention of macrovascular complications, we also assessed the potential commercial headroom of biomarkers to be applied within the primary prevention of DM2 (**Chapter 4**). As the PREDICCt project did not provide such markers for evaluation, we based our analysis on biomarkers recently presented in literature.[7] We assessed the commercial headroom of a set of four biomarkers added to 11 classical risk factors to predict clinically incident diabetes, and estimated this to be €75. This result is in line with the findings presented in chapter 2, namely that there is a larger commercial headroom to be expected in the primary prevention of DM2 than in the tertiary prevention of macrovascular complications. Whether this commercial headroom is large enough to warrant investments in discovery research to find such biomarkers depends on many factors, mainly the cost required for the research and development of these biomarkers, as well as the production cost of the eventual biomarker-based technology.

We further explored the potential for novel biomarkers in the primary prevention of DM2. To that end, we optimized a current design of a stepwise screening program for prediabetes on costs and cases detected (**Chapter 5**). Our previous finding that there likely is a commercial headroom for novel

biomarkers in primary prevention programs was confirmed. Increasing the accuracy of the risk prediction tool used in stepwise screening lead to lower costs per case detected, as well as a higher detection rate. However, we also found that other aspects of the screening program, such as patient response rates and costs of screening instruments, had a larger impact on these outcomes than improved accuracy. We found that when the costs for the first steps of screening were reduced through communicating by email, the subsequent reduction in costs per case detected was of the same magnitude as when the accuracy of the risk prediction tool was increased from an area under the curve of 74.3% to an area under the curve of 90%. This represents an improvement in accuracy far greater than thus far demonstrated by DM2 risk prediction biomarkers, quite possibly an unrealistic improvement.[7,8] By far the largest improvements in cost per case detected and the detection rates were achieved when patient response rates improved. This has a positive impact on both outcomes and is the only way to increase the total detection rate of the screening program. We thus conclude that even though there is a commercial headroom for novel biomarkers in the primary prevention of DM2, it is unlikely that developing novel biomarkers is an efficient strategy to improve screening programs. Our work did not include a comparison of the return on investment of research on novel biomarkers versus strategies to increase patient response rates. However, taking into account the enormous research effort required to develop and market biomarker base technologies, it is unlikely that returns on investments in that area will exceed those that may be achieved in effective public health strategies.

The yield and efficiency of screening for prediabetes is certainly not the only obstacle that stands in the way of implementing this primary prevention strategy. When screening for patients with prediabetes, some patients with previously undiagnosed DM2 will inevitably be identified and subsequently treated. There is currently no robust evidence that the standard DM2 treatment is cost-effective in screen-detected DM2 patients. Standard treating protocols have been developed for the treatment of clinically detected

DM2 patients, who invariably have a higher risk for complications than screen-detected DM2 patients. No treatment guidelines for screen-detected DM2 patients have been developed yet. Even though the treatment of prediabetes patients is likely to be cost-saving, it is uncertain whether these savings are large enough to compensate for the additional cost of treating the previously undiagnosed DM2 patients. Previous studies on the cost-effectiveness of prediabetes screening were based on data from two to three decades ago. Since then, the lead-time of DM2 has decreased due to increased awareness and the standard of care for DM2 patients has seen drastic improvements. As a result, the previous studies likely present an overestimation of the health effects of prediabetes screening, and thereby a too favorable cost-effectiveness estimate. We used more recent estimates of the lead-time of DM2 and treatment effects and found that treating patients identified through a prediabetes screening program is not likely to be cost-effective (**chapter 6**). Even when more favorable assumptions are made regarding the health effects of treating screen-detected DM2, the incremental cost-effectiveness ratio for treating patients identified through prediabetes screening is only just below the willingness to pay threshold. However, because the costs of the screening program itself were not included in this analysis, it is unlikely that the screening program and subsequent treatment of identified patients combined will remain cost-effective.

Besides the search for novel biomarkers, the PREDICCt project also included the continued development of a number of prototype devices that were in different stages of development at the start of the project. We conducted an early economic evaluation of one such technology, the DiagnOptics Diab-spot. The results of this work are however not presented in this thesis as they cannot be published in scientific journals due to the confidentiality of the data and outcomes. DiagnOptics has eventually discontinued the development of the Diab-spot.

Conclusion of our assessment of the potential clinical and economic impact of the CTMM PREDICCt project

The CTMM consortium had set itself very ambitious objectives. Our analysis of the research strategy and output indicates that these objectives have not been achieved within the stated time horizon. We found that there is potential to achieve the defined objectives when research output would result in improvements in primary and tertiary prevention of DM2. However, the actual research output in the form of novel biomarkers for tertiary prevention is unlikely to provide significant clinical or economic value. Novel biomarkers to be used in primary prevention were not available to us for assessment. However, we have concluded that investments in public health innovations are more likely to contribute to the consortium's objectives than investments in biomarker research. Our work only assessed the output of the PREDICCt project, and then only part thereof. The objectives stipulated at the beginning of this chapter were formulated for CTMM as a whole. We can thus not state with full certainty that the objectives have not been met through the results of other consortia. However, we have little indication that the PREDICCt project is an outlier within CTMM with respect to its contribution to the overall CTMM goals.[9] As the output of the PREDICCt project has thus far not been shown to lead to a reduction in mortality or savings in healthcare costs, we are inclined to conclude that the objectives of CTMM to reduce mortality by 20% and reduce annual healthcare costs by 1 billion Euro have not been realized.

NOVEL METHODS FOR THE EARLY ECONOMIC EVALUATION OF TRANSLATIONAL RESEARCH AND BIOMARKERS

Our second research objective was to further the methodology available for the early economic evaluation of translational biomedical research. Our methods are building upon existing methods from early HTA and other disciplines. Below, we discuss the methodological advancements on three

areas of health economic evaluation presented in this thesis: priority setting for translational research, early economic evaluation of biomarkers, and the evaluation of stepwise screening programs.

Priority setting and resource allocation in biomedical translational research

We have demonstrated the applicability of multi-criteria decision analysis (MCDA) in priority setting in biomedical translational research projects (**chapter 2**). MCDA has previously been applied in the context of biomedical innovation[10,11], as well as government-sponsored technology development programs in other fields.[12,13] The work in this thesis demonstrates for the first time how it can be used to take societal and commercial aims into account when allocating funds in a translational research project.

Translational research aims to address societal objectives, rather than to develop fundamental knowledge. The societal objectives of translational research must be taken into account during the many decisions that have to be taken at the start and during a translational research project. Due to the very complex nature of healthcare provision, successful biomedical innovations have to satisfy numerous different and often conflicting requirements. This calls for the incorporation of expertise from a large variety of disciplines in the research and development process. In such a complex setting it is unlikely that decision makers are able to adequately assess and weigh all information relevant to the decision.[11] This can easily lead to an inefficient allocation of resources (i.e., investment in projects that have a lower probability than others in achieving the formulated societal objectives). In addition, the investment of large amounts of public funds calls for transparent and reproducible decision-making. For these reasons, decision-making approaches based on heuristics are not adequate in initiating and guiding large translational research projects and are expected to lead to suboptimal outcomes. Thus, formal decision-making frameworks such as MCDA are preferred.

Given its complexity, it is astounding that research priority setting and resource allocation for translational biomedical research is most often done without formal decision-making frameworks. For example, within the Horizon 2020 program of the European Commission, the translation of abstract societal objectives (e.g., improve longevity and quality of life of the European population) into calls for research proposals is not transparent and no analysis is presented to demonstrate that the formulated research topics have a reasonably high potential to achieve those objectives. In the process of selecting which research proposals to fund, reviewers score the proposals on a set of subjective criteria, (e.g., the extent that the proposed work is beyond the state of the art, and demonstrates innovation potential) and subsequently aim to reach consensus.[14,15] Again, no empirical data is used to estimate the potential impact of proposals. Considering the vast amounts of public funds that are distributed through programs like Horizon 2020 (€77 billion)[16] and CTMM (€321 million)[17], it is surprising that so little effort is being put in ensuring that the funds are allocated in a way that provides the largest chance of achieving the societal objectives. We believe that the method we applied in the context of CTMM can be adapted to be applied to other programs, such as Horizon 2020.

Compared to the current practice, more time and effort are likely required in order to use MCDA for priority setting and resource allocation. It may therefore not be a suitable approach for smaller funding programs where this investment cannot be justified. However, for programs the size of Horizon 2020 and CTMM, the additional investment required to implement formal decision-making is very small compared to the funding budget and can be easily justified from the perspective of due diligence towards the public. In the case of CTMM PREDICCt, the total budget was €18.4 million.[18] Conducting an MCDA assessment for priority setting would cost less than €75,000 and for around €200,000 a full package of priority setting and early HTA assessments could be done to optimize resource allocation at the start and during the first half of the project. This is just over 1% of the total budget.

One important added value of the MCDA methods as we have applied it in the case study in chapter 2 over most existing early HTA methods, is that it can prioritize between alternative conceptual approaches, rather than specific clinical applications. Most early HTA methods are based on the incremental assessment that is central to HTA. This requires the specification of a target population and comparator intervention, or in the case of diagnostics or prognostics, a clinical decision to be informed. During the priority setting phase at the start of translational research projects, it is often not viable to define a potential application in such detail and compare alternatives. However, it is a phase in which it is crucial for decision-making to be guided by the societal objectives of the project. The MCDA method allows the use of available information on the general clinical effectiveness and cost-effectiveness of the standard of care in different alternative areas of application on a more general level to be incorporated. The information on clinical and cost-effectiveness gathered as part of the MCDA process can serve as input for early HTA methods. In that way, the MCDA method can be the start of an efficient iterative appraisal process alongside a translational research project.

The early health economic evaluation of biomarkers

In **Chapter 3** and **Chapter 4**, we have demonstrated how early HTA methods can be adapted and applied to the early health economic assessment of novel biomarkers. Frameworks for the development of biomarker-based technologies place economic evaluation at the end of a number of assessment steps that must be passed.[19–21] This makes sense from a regulatory perspective, where assessments can be done sequentially, and only those candidates that pass an assessment go on to the next. However, such a sequential assessment framework has limited value for developers and investors. They have to make an estimate of the commercial potential of a biomarker candidate at each of the decision gates in their R&D process. As a technology must fulfill all criteria to be a commercial success, all criteria have to be considered in each

decision. Not determining the potential of a technology under development to fulfill some criteria until very far in the development process is a very risky and potentially costly strategy if a technology under development fails that assessment. An assessment of the strength of the association between the presence of a biomarker and a target disease currently marks the endpoint of biomarker discovery research. This may present a logical time point for a decision gate in translational research. However, methods to provide an early insight into the potential clinical impact and commercial potential of biomarker candidates at this early phase of development have been lacking.

To assess biomarker candidates on their association with a disease or predictive power, only a disease of interest has to be specified. To assess biomarker candidates on clinical impact and economic or commercial value requires a definition of how the eventual biomarker-based technology will be used in clinical practice. This requires a definition of the target population that goes beyond the disease (e.g., age, prior lines of treatment), as well as specification of the clinical decision the biomarker-based test intends to inform (i.e., what is done as a result of different possible outcomes of the test in terms of treatment or other interventions).[22] Such a clinical application definition is often lacking for diagnostic or prognostic tests in development.[23] That in itself poses a problem for well-supported decision-making in the early phases of technological development. It is possible that in an early phase of development a test still has the potential to be applied in many different clinical settings. Assessing the value in all clinical settings is not always a viable option. However, the aim in this phase of development is most often not to obtain an estimate of the total value of an innovation in all possible applications, but rather to demonstrate that clinical applications exist in which the technology is likely to have a large enough value to merit the continuation of development. To that end, specifying the most likely application or the application where the most value can be expected intuitively is sufficient. In case investors want more certainty that the continued development of a biomarker candidate is warranted, an estimate of the total post-market cash

flow can be obtained from the commercial headroom estimate and target population size of one or several defined applications. It can then be decided if this total post-market cash flow presents an acceptable risk-adjusted return on the investments required to bring the biomarker-based technology to market.

One key advantage of the method we demonstrated in **Chapter 3** is that it can be done relatively quickly and does not require much data beyond what is available from association studies. This means that such analyses can be done relatively inexpensively and are suitable to be performed for a number of different applications. This is an important quality for such an analysis, as predictive and prognostic tests are more often developed by smaller companies with limited R&D budgets instead of larger pharmaceutical companies. The method can be used to obtain a quantitative substantiation of a value story and enables the identification of inadequate product concepts, thereby reducing the waste of R&D resources. It fits in an efficient 'fail fast fail cheap' approach.[24] A limitation of this method, however, is that there is a considerable amount of uncertainty in the outcomes. The method will have limited value in cases where the point estimate of the outcome is ambiguous, i.e., when it does not indicate either a very large or a nonexistent commercial headroom.

Early modeling methods, such as demonstrated in **chapter 4** could also be applied at the same phase of development. They have the potential to provide more detailed insights and allow for more flexibility in assessing different scenarios compared to the methods demonstrated in chapter 3. However, early health economic modeling requires much more data and is methodologically more complex. As a result, it is more resource intensive. It is rarely feasible to assess multiple biomarker candidates - each with multiple possible clinical indications - using early health economic modeling unless all these candidates can be assessed using the same health economic model. Apart from the time and resource demands of early modeling, this method also often runs into

the problem of data availability when it is used to analyzed biomarkers or other diagnostic technologies. In order to accurately model the effects of changing the cut-off of the test, detailed data is required from which the relation between the risk level in a population and the effects of treatment in this population can be inferred. Such data is often not available.

The assessment of stepwise screening programs

Like all healthcare interventions, screening for prediabetes should be assessed on its cost-effectiveness. However, the complex mechanism by which different design aspects of the stepwise screening program have an impact on the downstream costs and health effects of the subsequent intervention make such an analysis difficult to perform. Conducting a randomized controlled trial of prediabetes screening would be able to answer these questions, but such a study is very unlikely ever to be conducted due to ethical objections.[25] Modeling studies are also hampered by this complexity and the lack of data to model such effects accurately. One way to simplify the economic evaluation of screening is to employ an investment perspective. In this perspective, screening is seen as an upfront investment that will later result in benefits in terms of improved health, lower healthcare expenditures, or both. It is then the objective to design a screening program in such a way that it most efficiently identifies those individuals that benefit from the treatment following identification. In this perspective, the total uptake and cost per identified patient are relevant as they indicate, respectively, the total effectiveness and efficiency of the screening program. Until now, few studies on the design of screening programs have employed this approach. In **chapter 5** we demonstrate how a stepwise screening program can be assessed in this manner. This method enables the assessment of the effects of changing many different design parameters of a stepwise screening program in isolation or in combination. Besides optimizing the design of the screening program, it also indicates where further improvements in aspects of the screening program yield the most benefit. This enables setting research priorities

over the entire scope of disciplines involved, from the development of more accurate (biomarker-based) screening instruments to public communication strategies aimed at improving participation.

Limitations

The methods we use in this thesis are built on the principles of HTA. This means that they stem from a societal decision context that expects policymakers to aim to provide the largest possible health benefits from the public resources allocated to healthcare provision. The relevance of this decision context for investors and developers depends both on their intrinsic motivation and extrinsic incentives. Investors and developers may have an intrinsic motivation to develop technologies that result in the largest possible health benefit. To some extent, this is to be expected from investors and developers in public-private translational research consortia. External incentives are formed by market regulators and technology purchasers who limit market access and demand for technologies that do not provide good value for money. The extent to which these internal and external incentives are applicable differ per country, type of technology, and project. For example, the market for medical technology is in most countries much less regulated than the market for pharmaceuticals, also in terms of the requirements on the cost-effectiveness of innovations. On the other hand, pharmaceutical companies demonstrate time and time again that their main objective is profit maximization (i.e. 'creating shareholder value'). Additionally, we observe that regulators and purchasers are not able to keep expensive new products off the market, which likely leads to the displacement of more efficient forms of treatment and prevention. Investors and developers that aim to maximize profit in poorly regulated markets with no regard for the possible harm their actions might cause due to the displacement of more efficient technologies will find little value in our methods. Those that aim to improve the health outcomes of the population will.

IMPLICATIONS FOR POLICY AND FURTHER RESEARCH

Initiators and funding programs of translational research projects

Agencies investing in translational research who are serious about addressing societal issues would be well advised to use formal decision-making methods when making funding decisions (as discussed extensively above). We have shown that formal decision-making methods can be applied in this setting and that they can identify suboptimal investment options. However, the commitment to optimally allocating resources should go beyond the methods used to inform decisions. In many research funding agencies (biomedical) fundamental scientists are over-represented in the committees that decide on the allocation of research funds (see CTTM for example[26]). This can be expected to impact funding decisions. First, it is likely that fundamental scientists are better able to determine the scientific merit, value, and feasibility of a research proposal, rather than its societal value. To some extent, this limit in expertise can be addressed using formal decision methods, when applied correctly. However, it is also likely that an over-representation of fundamental scientists leads to a conflict of interest. This can be very direct, e.g., when a researcher submits a proposal to a funding agency at which he or she is also a referee. It can also be indirect, such as when fundamental biomedical research as a profession is competing with other research professions for the same resources (e.g., public health researchers, health policy researchers, or health economists). Thus, translational funding agencies (like CTMM and H2020) should include input from experts from all areas relevant to medical innovation in the committees that decide on resource allocation. Apart from fundamental biomedical research, this also would include medical experts, (bio-)statisticians, epidemiologists, health economists, patient advocates, insurance companies/ national payers, manufacturers, hospital management, regulatory agencies, and many others. Obviously, a funding

committee comprised of all these experts can become impractical due to the number of participants and different viewpoints. This, again, is where formal decision methods such as MCDA can help by having a working committee of decision method experts who are impartial to the funding decision collect and synthesize expertise from all these experts to inform the allocation decision.

Personalized medicine and biomarkers

The vast majority of biomarker discovery research is fundamental research aimed at obtaining a better understanding of the molecular pathology underlying a disease. Unfortunately, such research has often been presented to investors, funding agencies, and the general public as translational research. This often involves the claim that a biomarker can be used in treatment stratification or risk prediction in a manner that has clinical value (see for example Van der Leeuw et al. [27]). For such a claim to be more than just a strategy to make fundamental research seem more clinically and societally relevant than it actually is, a plausible value hypothesis has to be formulated a priori. This fundamental premise of this value hypothesis differs between prognostic and predictive biomarkers.

For predictive biomarkers to be of value, there has to be an indication that there is heterogeneity in the underlying (molecular) pathology or pharmacokinetics in the target population. Sometimes multiple different pathological mechanisms are considered one and the same disease because they present the same symptoms, and the underlying pathologies have yet to be discovered.[28] In these cases, it is possible that differentiating treatment to each pathological mechanism provides better health outcomes. Developing a biomarker that is able to distinguish these different pathological mechanisms and thereby enabling this differentiated treatment can, in that case, have value. Pharmacogenetics has revealed that genetic differences can to some extent explain heterogeneity in treatment response.[29] These genetic differences impact on pharmacokinetics as they are related to liver enzymes

involved in the metabolism of certain pharmaceuticals. Biomarkers that can identify those patients that are not able to metabolize a drug can have value by avoiding resource waste on ineffective treatments.

In the case of prognostic biomarkers, the value hypothesis is that the distribution of disease risk in the population is very broad or possibly bimodal. In cases where there is a group with a very high risk and one with a very low risk, the classification of patients in these risk categories allows for more appropriate prevention or screening strategies. The value of risk stratification is reduced as the variation of risk in the population becomes smaller. As the health effects of preventive approaches are directly related to the level of risk, there is little value to be gained by stratifying on risk when there is little difference in risk within the population.

When it is not possible to formulate a value hypothesis grounded in current knowledge on the pathology and risk of the disease or pharmacokinetics, biomarker discovery research cannot be considered to be translational. This is the case for a large share of biomarkers published in literature. In case a value hypothesis is formulated, it can serve as a starting point for all methods we have described in this thesis from MCDA to early health economic modeling.

Type 2 diabetes prevention and screening

The incidence of DM2 is expected to continue to increase over the coming decades.[30] Given that prognosis it is understandable that policymakers and researchers are looking into ways of preventing or reducing the burden of disease from DM2, either through primary or tertiary prevention. In this thesis, we have attempted to provide an early assessment of the potential of primary and tertiary prevention strategies to address the health and economic burden of DM2 through their impact on the incidence of DM2 or DM2 related complications. However, both primary and tertiary DM2 prevention strategies have an impact which extends beyond their effect on DM2 and related

complications. These interventions should therefore be assessed within this broader context. Primary prevention of DM2 has to be seen within the context of the prevention of metabolic syndrome, that is, overweight or obesity and a sedentary lifestyle.[31] The main prevention tool, lifestyle intervention, is directly aimed at addressing these factors.[32] The target population of such interventions should be defined on risk for developing metabolic syndrome, not only the risk for DM2. The design, implementation, and evaluation of such interventions should be done within this appropriate target population, rather than patients with a high risk for DM2. The idea that the prevention of DM2 must be regarded within the larger scope of the metabolic syndrome had been addressed before the CTMM PREDICCt project had started.[33,34] Not following this broader approach limited the possible societal impact of the project.

Similarly, tertiary prevention of cardiovascular complications has to be regarded within a broader context of cardiovascular risk management. Over the past years, we have seen an increase in the integration of guidelines for DM2 treatment and cardiovascular risk management. For example, the American Diabetes Association and American Heart Association have published a joint guideline.[4] In this perspective, DM2 is regarded as one of many risk factors for numerous cardiovascular diseases. It should always be considered in conjunction with other risk factors of the metabolic syndrome.[31] This shift in perspective has not yet fully precipitated into the research and development of new prevention strategies in this field. New tertiary DM2 prevention strategies should be regarded as interventions addressing cardiovascular risk factors, and their health and economic impact should be assessed as such.

Assessing innovations for the prevention of DM2 and related complications from the narrow (at risk for) DM2 patient population perspective may lead to underestimating the health and economic effects of these interventions. This narrow perspective can also distort the relative potential of different innovations, when some alternatives only realize an impact in the DM2

patient population, whereas others additionally have a large effect in non-DM2 patients. This may lead to suboptimal decisions when prioritizing research lines or allocating research funding. Assessing multiple research alternatives must therefore always be done from the broader metabolic syndrome or cardiovascular risk management perspective.

The debate whether screening for undiagnosed DM2 is effective, whether or not in combination with prediabetes, still endures in the scientific community. The arguments for and against have not changed much over the past five years.[25,35–37] This is in no small part due to the fact that no randomized controlled trial of screening versus no screening has been conducted, and is not likely to be ever conducted.[25] It remains uncertain whether screening and early treatment of DM2 result in improved health outcomes. Circumstantial evidence indicates that population screening (i.e., screening every individual within a certain age group) is not likely to provide any benefits over current care. This is due to the increased awareness of DM2 in the general practice and the subsequent opportunistic screening (i.e., testing for DM2 when it is suspected based on symptoms).[38] As awareness with general practitioners and the general public on DM2 and its risk factors increases further, and as inexpensive point of care tests for HbA1c enable opportunistic general practice, it becomes increasingly less likely that population screening will be a cost-effective use of public health resources. Further research and technological innovation aimed at addressing the expected increasing burden of DM2 should therefore be focused elsewhere.

Further early HTA research

The work presented in this thesis has strengthened the available methodology to assess and guide translational research from beginning to end. That is to say, from investment decision to market access and implementation. We mainly contributed to the first phase of this path, from investment decision to the phase of (pre-)clinical testing. The latter phase, from clinical testing

to market access (i.e., classical HTA) was already better developed and has seen continued development over the past years.[24,39–41] The methodology for early HTA will no doubt continue to be developed over the coming years, as questions about the efficiency and sustainability of health care systems will further come to the forefront.

The methods presented in this thesis have been developed within the context of a translational public-private consortium. We are however convinced that these methods are also relevant for private investors and developers of biomedical technologies. Private developers and public-private consortia differ on several important aspects, such as their investment horizon, willingness or ability to take risk, cost of capital, access to necessary expertise, and ultimately their objectives. Further research should be done to optimize the methods presented in this thesis for their application to support the decision-making of private developers of biomedical technologies, taking into account the conditions unique to private development in smaller enterprises. Subsequently, it should be confirmed empirically that these methods provide value for the investment and R&D decision-making in those companies. After all, a scientific field that is committed to maximizing the value created by the healthcare system should be equally committed to maximizing the value of their methods to achieve that goal.

REFERENCES

1. Center for Translational Molecular Medicine. *Business Plan*. (2006).
2. Noble, D, Mathur, R, Dent, T, Meads, C & Greenhalgh, T. Risk models and scores for type 2 diabetes: Systematic review. *BMJ* **343**, d7163 (2011).
3. Van Der Leeuw, J, Ridker, PM, Van Der Graaf, Y & Visseren, FLJ. Personalized cardiovascular disease prevention by applying individualized prediction of treatment effects. *European Heart Journal* **35**, 837–843 (2014).
4. Fox, CS, Golden, SH, Anderson, C, *et al*. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: A scientific statement from the American Heart Association and the American diabetes association. *Diabetes Care* **38**, 1777–1803 (2015).
5. American Diabetes Association (ADA). Standard of medical care in diabetes - 2017. *Diabetes Care* **40** (**sup 1**), s4–s128 (2017).
6. Kaasenbrood, L, Boekholdt, SM, Van Der Graaf, Y, *et al*. Distribution of Estimated 10-Year Risk of Recurrent Vascular Events and Residual Risk in a Secondary Prevention Population. *Circulation* **134**, 1419–1429 (2016).
7. Salomaa, V, Havulinna, A, Saarela, O, *et al*. Thirty-one novel biomarkers as predictors for clinically incident diabetes. *PLOS ONE* **5**, 1–8 (2010).
8. Echouffo-Tcheugui, JB, Dieffenbach, SD & Kengne, AP. Added value of novel circulating and genetic biomarkers in type 2 diabetes prediction: a systematic review. *Diabetes Research and Clinical Practice* **101**, 255–269 (2013).
9. Steuten, LM. Multi-Dimensional Impact of the Public–Private Center for Translational Molecular Medicine (CTMM) in the Netherlands: Understanding New 21 st Century Institutional Designs to Support Innovation-in-Society. *OMICS: A Journal of Integrative Biology* **20**, 265–273 (2016).
10. Frazão, TDC, Camilo, DGG, Cabral, ELS & Souza, RP. Multicriteria decision analysis (MCDA) in health care: A systematic review of the main characteristics and methodological steps. *BMC Medical Informatics and Decision Making* **18**, 1–16 (2018).
11. 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 in Health* **19**, 1–13 (2016).
12. Macharis, C & Bernardini, A. Reviewing the use of multi-criteria decision analysis for the evaluation of transport projects: Time for a multi-actor approach. *Transport Policy* **37**, 177–186 (2015).
13. Kiker, GA, Bridges, TS, Varghese, A, Seager, PTP & Linkov, I. Application of Multicriteria Decision Analysis in Environmental Decision Making. *Integrated Environmental Assessment and Management* **1**, 95–108 (2005).

14. European Commission. Evaluation process and results - H2020 Online Manual. *H2020 Online Manual* (2017). Available at: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/grants/from-evaluation-to-grant-signature/evaluation-of-proposals/eval_process_results_en.htm. (Accessed: 7th April 2019)

15. European Commission. Eligibility and evaluation criteria - H2020 Online Manual. *H2020 Online Manual* (2017). Available at: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/grants/from-evaluation-to-grant-signature/evaluation-of-proposals/elig_eval_criteria_en.htm. (Accessed: 7th April 2019)

16. European Commission. *Factsheet: Horizon 2020 budget*. (2013).

17. Lygature. The CTMM Portfolio | Lygature. (2019). Available at: https://www.lygature.org/ctmm-portfolio. (Accessed: 7th April 2019)

18. CTMM Predicct. *Biomarkers for the Prediction and Early Diagnosis of Diabetes and Diabetes-related Cardiovascular Complications: Output Report*. (2015).

19. Van den Bruel, A, Cleemput, I, Aertgeerts, B, Ramaekers, D & Buntinx, F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *Journal of Clinical Epidemiology* **60**, 1116–1122 (2007).

20. Sackett, D & Haynes, R. The architecture of diagnostic research. *BMJ* **324**, 539–541 (2002).

21. Pepe, MS, Etzioni, R, Feng, Z, *et al*. Phases of Biomarker Development for Early Detection of Cancer. *Cancer* **93**, 1054–1061 (2001).

22. Horvath, AR, Lord, SJ, StJohn, A, *et al*. From biomarkers to medical tests: The changing landscape of test evaluation. *Clinica Chimica Acta* **427**, 49–57 (2014).

23. Ioannidis, JPA & Bossuyt, PMM. Waste, Leaks, and Failures in the Biomarker Pipeline. *Clinical chemistry* **63**, 963–972 (2017).

24. IJzerman, MJ, Koffijberg, H, Fenwick, E & Krahn, M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *PharmacoEconomics* **35**, 727–740 (2017).

25. Simmons, D & Zgibor, JC. Should we screen for type 2 diabetes among asymptomatic individuals? Yes. *Diabetologia* **60**, 2148–2152 (2017).

26. Center for Translational Molecular Medicine. *Translating science into better healthcare*. (2016).

27. van der Leeuw, J, Beulens, JWJ, van Dieren, S, *et al*. Novel Biomarkers to Improve the Prediction of Cardiovascular Event Risk in Type 2 Diabetes Mellitus. *Journal of the American Heart Association* **5**, e003048 (2016).

28. Jameson, JL & Longo, DL. Precision Medicine — Personalized, Problematic, and Promising. *The New England Journal of Medicine* **372**, 2229–2234 (2015).

29. Scott, SA. Personalizing medicine with clinical pharmacogenetics. *Genetics in Medicine* **13**, 987–995 (2011).

30. International Diabetes Federation. *IDF Diabetes Atlas, 8th edition*. (2017).

31. Grundy, SM. Metabolic syndrome: Connecting and reconciling cardiovascular and diabetes worlds. *Journal of the American College of Cardiology* **47**, 1093–1100 (2006).

32. Magkos, F, Yannakoulia, M, Chan, JL & Mantzoros, CS. Management of the Metabolic Syndrome and Type 2 Diabetes Through Lifestyle Modification. *Annual Review of Nutrition* **29**, 223–256 (2009).

33. Griffin, SJ, Little, PS, Hales, CN, Kinmonth, a L & Wareham, NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. *Diabetes/Metabolism Research and Reviews* **16**, 164–171 (2000).

34. Spijkerman, A, Griffin, S, Dekker, J, Nijpels, G & Wareham, NJ. What is the risk of mortality for people who are screen positive in a diabetes screening programme but who do not have diabetes on biochemical testing? Diabetes screening programmes from a public health perspective. *Journal of Medical Screening* **9**, 187–190 (2002).

35. Goyder, E, Irwig, L & Payne, N. Should we screen for type 2 diabetes? No. *BMJ* **345**, e4516 (2012).

36. Khunti, K & Davies, M. Should we screen for type 2 diabetes: Yes. *BMJ* **345**, e4514 (2012).

37. Shaw, JE. Does the evidence support population-wide screening for type 2 diabetes? No. *Diabetologia* **60**, 2153–2156 (2017).

38. Colagiuri, S, Hussain, Z, Zimmet, P, Cameron, A & Shaw, J. Screening for type 2 diabetes and impaired glucose metabolism: the Australian experience. *Diabetes Care* **27**, 367–371 (2004).`

39. Markiewicz, K, van Til, JA & IJzerman, MJ. Medical devices early assessment methods: systematic literature review. *International Journal of Technology Assessment in Health Care* **30**, 137–146 (2014).

40. Støme, LN, Moger, T, Kidholm, K & Kværner, KJ. Early assessment of innovation in a healthcare setting. *International Journal of Technology Assessment in Health Care* **35**, 1–10 (2019).

41. IJzerman, M & Steuten, LM. Early Assessment of Medical Technologies to Inform Product Development and Market Access: A Review of Methods and Applications. *Applied Health Economics and Health Policy* **9**, 331–347 (2011).

Chapter 8

Summary

Nederlandse samenvatting

Dankwoord

Other SHARE dissertations

About the author

SUMMARY

This thesis presents our work on the assessment of the output of the CTMM PREDICCt project and on the development of novel methods for the early evaluation of translational research. The CTMM PREDICCt project was a large public-private partnership that aimed to develop innovative biomarker-based technologies to allow identification of individuals at increased risk of type-2 diabetes mellitus (DM2) and related complications. This risk profiling was meant to lead to improvements in length and quality of life, and a reduction in health care costs. As a whole, CTMM aimed to reduce mortality by 20% and healthcare costs by 1 billion Euro by 2019. Right from the start of the PREDICCt project, we commenced the evaluation of the project output against these objectives. To that end, we have adapted existing and developed new methods for the early economic evaluation of biomarkers.

We first analyzed the chosen allocation of available research funds by looking at the potential value for biomarker-based technologies in different areas of DM2 prevention (**chapter 2**). We found that the development of novel techniques applied in secondary prevention would be a poor investment of research funds. Considerably more value was to be expected from investments in primary and tertiary prevention. The relative attractiveness of these options depended on the strategic preferences of the decision maker. In case a large clinical and commercial value was preferred, primary prevention was the more attractive investment target. However, this is a more risky and complex endeavor than developing markers for tertiary prevention.

The PREDICCt project identified a number of biomarkers that could be implemented in the tertiary prevention of cardiovascular complications of DM2. We defined a clinical application for these biomarkers and subsequently estimated their clinical and commercial value (**chapter 3**). We found the expected clinical and economic value of these biomarkers to be very small, and advise against further investments in their development.

As the project had not delivered any biomarkers for the primary prevention of DM2 by the time of our analysis, we used a large set of previously published biomarkers to explore the general potential for novel biomarkers to be of value in the primary prevention setting (**chapter 4**). In line with our conclusions from chapter 2, we found that novel biomarkers applied in primary prevention potentially provide a larger clinical and commercial value than those applied in tertiary prevention.

As primary prevention is a complex public health challenge in which many factors play a role, we further explored the potential for novel biomarkers in this field. Biomarkers in primary DM2 prevention are most likely applied in the first step of a stepwise screening program. We assessed the effects of improved risk stratification within a stepwise screening program, and compared that to other optimization strategies (**chapter 5**). Our previous finding that biomarkers that improve risk prediction have clinical and commercial value in primary prevention programs (**chapter 4**) was confirmed. However, we also found that other aspects of the screening program, such as patient response rates and costs of screening instruments, had a more profound impact on the efficiency and yield of screening than improved risk stratification. It is therefore highly uncertain that developing novel biomarkers is an efficient strategy to improve the efficiency and yield of screening programs. Taking into account the vast research efforts and investments required to develop and market biomarker-based technologies, it is unlikely that returns on investments in that area will exceed those in public health strategies aimed at improving other critical aspects in screening programs.

Another obstacle prohibiting the use of biomarkers for the primary prevention of DM2 is the uncertainty surrounding the efficiency of primary prevention of DM2 in general. There is little evidence on the optimal treatment strategy of screen detected DM2 patients, as it has long been considered unethical to run a trial on this. We demonstrated the importance that the lead-time of screening and the effects of early treatment due to screening have on the

cost-effectiveness of a primary prevention program (prediabetes screening) (**chapter 6**). Based on the most recent estimates of the lead-time of DM2 screening and clinical outcomes of current care for clinically detected DM2 patients, we concluded that the treatment of patients identified in a prediabetes screening program is unlikely to be cost-effective. Investing in the development of biomarkers for the primary prevention of DM2 that would be used in a stepwise screening program is not advisable as long as these major hurdles for screening are not addressed.

As part of our work, we have made methodological advancements in three areas of health economic evaluation: priority setting for translational research, early economic evaluation of biomarkers, and the evaluation of stepwise screening programs. Such methods are increasingly valuable in the current reality of increasing pressure on healthcare budgets and scrutiny of new medical technologies on their cost-effectiveness. They are not only relevant to the work of public funders of research and research consortia, but also to private developers of medical technologies.

The work presented in this theses does not cover all possible applications of biomarkers in the prevention of DM2 complications. Notably, the area of prevention of microvascular complications is lacking as the PREDICCt project had not developed any such markers at the time of evaluation. Nonetheless, we found few indications that the development of biomarkers will lead to a substantial improvement in health outcomes and a reduction in costs. We have therefore concluded that the objectives of CTMM as stated above have not been realized.

NEDERLANDSE SAMENVATTING

In dit proefschrift presenteren wij ons werk met betrekking tot de beoordeling van de uitkomsten van het CTMM PREDICCt project en het ontwikkelen van nieuwe methoden voor de vroege economische evaluatie van translationeel onderzoek. Het CTMM PREDICCt project was een groot publiek-privaat samenwerkingsverband dat ten doel had innovatieve biomarker-gebaseerde technologieën te ontwikkelen om daarmee personen met een hoog risico op het ontwikkelen van type-2 diabetes mellitus (DM2), of daaraan gerelateerde complicaties, te identificeren. Uiteindelijk zou deze risicoprofilering moeten leiden tot een verbetering in lengte en kwaliteit van leven van de Nederlandse bevolking en een besparing van zorgkosten. Het CTMM als geheel had zich ten doel gesteeld om in 2019 de mortaliteit met 20% te reduceren en 1 miljard Euro zorgkosten te besparen. Meteen aan het begin van het PREDICCt project zijn we begonnen met de evaluatie van de uitkomsten aan de hand van deze doelstellingen. We hebben daarvoor bestaande methoden voor de vroege economische evaluatie van biomarkers aangepast en nieuwe methoden ontwikkeld.

Allereerst hebben we de gekozen toewijzing van de beschikbare onderzoeksgelden geëvalueerd door te kijken naar de potentiële waarde van biomarker-gebaseerde technologieën in verschillende gebieden van DM2-preventie (**hoofdstuk 2**). We stelden vast dat de ontwikkeling van nieuwe technieken voor secundaire preventie een slechte investering van onderzoeksgeld zou zijn. Van investeringen in primaire en tertiaire preventie was aanzienlijk meer waarde te verwachten. De relatieve aantrekkelijkheid van deze opties hing af van de strategische voorkeuren van de besluitvormer. Wanneer een grote klinische en commerciële waarde prioriteit heeft, was investeren in primaire preventie het meest aantrekkelijk. Deze aanpak is echter riskanter en moeilijker te realiseren dan het ontwikkelen en implementeren van markers voor tertiaire preventie.

Het PREDICCt project heeft een aantal biomarkers geïdentificeerd die gebruikt zouden kunnen worden in de tertiaire preventie van cardiovasculaire complicaties. We hebben een klinische toepassing voor deze biomarkers gedefinieerd en vervolgens een schatting gemaakt van hun klinische en commerciële waarde (**hoofdstuk 3**). We stelden vast dat de verwachte klinische en economische waarde van deze biomarkers zeer gering is en adviseren om geen geld te investeren in hun verdere ontwikkeling.

Ten tijde van ons onderzoek waren er binnen het PREDICCt project geen biomarkers geïdentificeerd die toegepast konden worden in de primaire preventie van DM2. Om die reden hebben we de potentiele waarde van het toepassen van nieuwe biomarkers in de primaire preventie van DM2 in zijn algemeenheid ingeschat aan de hand van biomarkers die in de wetenschappelijke literatuur beschreven waren (**hoofdstuk 4**). In lijn met onze conclusies uit hoofdstuk 2 stelden we vast dat biomarkers toegepast in primaire preventie potentieel een grotere klinische en commerciële waarde vertegenwoordigen dan biomarkers toegepast in tertiaire preventie.

Aangezien primaire preventie een complexe uitdaging is voor de volksgezondheid waarbij veel verschillende factoren een rol spelen, hebben we de mogelijke waarde van nieuwe biomarkers op dit gebied verder onderzocht. De meest waarschijnlijke toepassing van biomarkers in de primaire preventie van DM2 is in de eerste stap van een stapsgewijs screeningsprogramma. We onderzochten de effecten van verbeterde risicostratificatie met behulp van biomarkers op de efficiëntie en opbrengst van stapsgewijze screening en vergeleken deze met de effecten van andere strategieën om screening te optimaliseren (**hoofdstuk 5**). Onze eerdere bevinding dat biomarkers die de risicovoorspelling verbeteren klinische en commerciële waarde hebben (**hoofdstuk 4**) werd bevestigd. We stelden echter ook vast dat andere aspecten van het screeningprogramma, zoals de respons van patiënten en de kosten van de screeningsinstrumenten, een meer substantiële impact hadden op de efficiëntie en opbrengst van stapsgewijze screening dan

een verbeterde risicostratificatie. Het is daarom hoogst onzeker dat het ontwikkelen van nieuwe biomarkers een doelmatige strategie is om de efficiëntie en opbrengst van de screeningprogramma's te verbeteren. Rekening houdend met de enorme onderzoeksinspanningen die nodig zijn om biomarker-gebaseerde technologieën te ontwikkelen en op de markt te brengen, is het onwaarschijnlijk dat het rendement op investeringen in dat gebied groter zullen zijn dan investeringen in andere aspecten van het screeningsprogramma.

De onzekerheid over de doelmatigheid van primaire preventie van DM2 is een ander obstakel voor de ontwikkeling van biomarkers voor die toepassing. Het is onduidelijk wat de optimale behandelstrategie is bij door screening opgespoorde diabetes patiënten, aangezien studies hiernaar wegens ethische redenen niet zijn uitgevoerd. Bij het screenen voor een bepaalde ziekte ontstaat de zogeheten lead time: de tijd tussen het moment dat een patiënt geïdentificeerd wordt middels screening en de tijd dat deze patiënt zonder screening klinisch gediagnosticeerd zou zijn. De lengte van deze lead-time en de effecten van het eerder behandelen van DM2 als gevolg van de screening hebben invloed op de kosteneffectiviteit van een primair preventieprogramma (prediabetes screening) (**hoofdstuk 6**). Op basis van de meest recente schattingen van de lead-time van DM2-screening en de gezondheidsuitkomsten van de huidige zorg voor klinisch gediagnosticeerde DM2-patiënten, concluderen we dat het onwaarschijnlijk is dat de behandeling van patiënten die geïdentificeerd zijn in een prediabetes-screeningprogramma kosteneffectief is. Investeren in de ontwikkeling van biomarkers voor de primaire preventie van DM2 die gebruikt zouden worden in stapsgewijze screening is niet aan te raden zolang deze belangrijke obstakels voor screening niet worden aangepakt.

Als onderdeel van ons werk hebben we methoden doorontwikkeld in drie toepassingsgebieden van gezondheidseconomische evaluatie: het stellen van prioriteiten voor translationeel onderzoek, vroegtijdige

economische evaluatie van biomarkers en de evaluatie van stapsgewijze screeningsprogramma's. Dergelijke methoden worden steeds waardevoller in de huidige tijd waarin budgetten in de zorg steeds verder onder druk komen te staan en er kritischer wordt gekeken naar de doelmatigheid van nieuwe medische technologieën. Deze methoden zijn niet alleen relevant binnen publiek gefinancierd onderzoek, maar ook voor particuliere ontwikkelaars van medische technologieën.

In het in dit proefschrift gepresenteerde onderzoek hebben we niet alle mogelijke toepassingen van biomarkers voor de preventie van DM2-complicaties geëvalueerd. Met name de preventie van microvasculaire complicaties ontbreekt, aangezien het PREDICCt-project ten tijde van onze evaluatie nog geen biomarkers voor die toepassing had ontwikkeld. Toch vonden we in ons onderzoek weinig aanwijzingen dat de ontwikkeling van dergelijke biomarkers zal leiden tot een aanzienlijke verbetering van de gezondheidsresultaten en een besparing van zorgkosten. Om die reden concluderen wij dat de doelstellingen van CTMM zoals hierboven beschreven niet zijn gerealiseerd.

DANKWOORD

Gezondheidseconomen vertonen geen uitstelgedrag, zij verdisconteren hyperbolisch. Deze grap maakte ik vaak als iemand mij, terecht, wees op de keuzes die ik maakte ten aanzien van mijn tijdsbesteding. Nu er een tastbaar resultaat ligt zou ik dat willen aanvullen met: *'Good things come to those who wait'*. Ik hoop dat jullie het het lange wachten waard vonden.

Ik wil op deze plaats graag mijn dank en waardering uitspreken naar iedereen die op welke wijze dan ook heeft bijgedragen aan dit proefschrift en iedereen wiens geduld ik op de proef heb gesteld.

Beste Erik. Eindelijk ligt hij er dan. De laatste jaren ontving ik met enige regelmaat een kattenbelletje dat zowel informeerde naar de voortgang als motiveerde om het momentum erin te houden. Bedankt dat je in al die tijd het vertrouwen in de goede afloop niet bent verloren. Je was altijd snel met het leveren van waardevol commentaar op manuscripten, zodat ik snel verder kon als ik een periode de tijd had om aan het proefschrift te werken. Tijdens onze overleggen in Groningen bood je een waardevolle blik buiten de ivoren toren van de wetenschap. Je hebt met input vanuit je bestuurlijke ervaring ons werk relevanter weten te maken. Onze discussies over de bredere wetenschappelijke en maatschappelijke context van ons onderzoek hebben mij altijd zeer gemotiveerd. En het was leuk om tijdens de meetings regelmatig even verhalen over mooie motorreizen uit te wisselen.

Beste Douwe. Je hebt meer dan wie dan ook bijgedragen aan het tot stand komen van dit proefschrift. Het was voor mij een bijzondere reis, leerzaam op vele vlakken. Het kostte mij aan het begin van onze samenwerking enige tijd om te wennen aan je stijl en werkwijze. Tijdens een overleg kon je na een opmerking van mijn kant gerust vijf minuten in volledige stilte naar het papier staren en nadenken. In het begin wist ik me nooit raad met deze lange denkpauzes. Soms kwam je daarna tot een conclusie die mij niet aanstond.

Dan dacht ik: het zal wel, ik trek mijn eigen plan. Om vervolgens na twee dagen tot de conclusie te komen dat er inderdaad geen betere aanpak of oplossing was dan hetgeen jij had bedacht. Ik heb er nog dagelijks plezier van dat je me in de eerste maand meteen een R-script onder de neus schoof en me motiveerde om te leren coderen. De beste gesprekken hadden we wat mij betreft tijdens het avondeten in het ziekenhuis. We doken dan wat minder de details in en reflecteerden op ons vakgebied of de grote lijnen van het project. De beste stukken van het proefschrift zijn dan ook na het eten geschreven. Ik heb ook goede herinneringen aan onze activiteiten buiten de muren van het ziekenhuis. Van het kijken van vele voetbalwedstrijden in de kroeg, tot het jaarlijkse zaalvoetbaltoernooi waar we steevast onderaan eindigden. Wij waren immer de winnaars van de derde helft, tot aan het crashen van neon-splash techno feestjes aan toe. Maar het meest memorabel was toch wel aan onze trip in de VS, na afloop het SMDM congres in Phoenix. Een roadrip met hiken in de Grand Canyon en Zion park, eindigend in Las Vegas. Een onvergetelijke tijd. Douwe, ik ben je zeer dankbaar voor alles wat je me hebt geleerd en de leuke tijd in Groningen!

Graag wil ik op deze plaats ook mijn dank uitspreken aan de artsen die ons geholpen hebben om ons werk zo goed mogelijk te laten aansluiten bij de klinische praktijk, in het bijzonder mijn coauteurs Stefan Bakker en Jan Westerink. Stefan, ik heb je benaderbaarheid enorm gewaardeerd. Ik mocht je zelfs oppiepen wanneer je in de kliniek stond (wat ik nooit durfde, want artsen oppiepen doe je toch alleen als er echte problemen zijn). Jan, ik stuurde je een manuscript de dag voordat ik op vakantie ging, omdat ik ervan uitging dat het commentaar wel twee weken op zich zou laten wachten. Mijn verbazing was dan ook groot dat ik het al terug had voordat ik mijn huis had verlaten. Je was snel maar ook zeer scherp met je commentaar en was een waardevolle bondgenoot in het pareren van de kritiek van klinische reviewers.

Vanuit het CTMM en het PREDICCt project hebben ook vele mensen bijgedragen aan ons onderzoek. In het bijzonder gaat mijn dank uit naar

Marten Hofker die enorm veel heeft betekend voor het PREDICCt project. Het is spijtig dat hij er niet meer is om het eindresultaat in handen te krijgen. Vanuit CTMM volgde Erna Erdtsieck-Ernste ons onderzoek altijd met zeer veel interesse en probeerde onze bevindingen te benutten in de besluitvorming.

Geachte leescommissie, beste professor Frijlink, professor Moons en professor Severens, hartelijk dank voor jullie bijdrage in het beoordelen van mijn proefschrift en het zitting nemen in de corona.

Naast de vele mensen die dit werk professioneel hebben ondersteund is er ook een grote groep mensen die het mede mogelijk heeft gemaakt middels sociale ondersteuning. Ik ben erg blij dat twee mensen die heel erg belangrijk voor mij zijn geweest in de periode dat ik in Groningen woonde mij als paranimf zullen bijstaan tijdens de verdediging, de definitieve afsluiting van dat hoofdstuk.

Lieve Leanne, waarde (enig) mede lid van *'dit is niet mijn alma mater'*, de club voor Groningse promovendi die niet in Groningen gestudeerd hebben. Al snel na je komst op de vierde verdieping vonden we elkaar in onze behoefte om doordeweeks wat meer sociale activiteiten te ontplooien, omdat we de weekenden meestal buiten 'stad' spendeerden. Toch had het wegwerken van een bourgondische hoeveelheid Pauwel Kwak op een dinsdagavond ook wel zo zijn nadelen (voor de productiviteit de volgende dag). Maar de sfeer die jij bracht op de vierde verdieping en de vele borrels en feetjes hebben er zeker aan bijgedragen dat ik in Groningen ben blijven doorwerken aan dit proefschrift. Ik ben blij dat je voor mij, na al zovelen promovendi te hebben bijgestaan, nog een keer in de rol van paranimf wil kruipen. En nu je stilletjes om de hoek bent komen wonen: gaan we verder met de dinsdag borrels?

Jonas, we kennen elkaar inmiddels langer wel dan niet. Jouw aanwezigheid in Groningen was voor mij een belangrijke reden om de geboden promotieplek ook daadwerkelijk aan te nemen. Ik had niet kunnen vermoeden dat dat het begin zou zijn van een reis naar de uithoeken van deze wereld. Van de

binnenlanden van IJsland in een luxe Lexus, tot de westkust van Australië in *good-old* Gus the Snail. En natuurlijk vele avonturen dichter bij huis: op de motor voor de bliksem uit racen in het Sauerland en met de trailer op de autobahn achter je eigen banden aan. Van muziek maken, optreden met bedenkelijke bands tot racen in Porsches en karts. *Never a dull moment.* Ik heb ontelbare dierbare herinneringen aan de vele dingen die we hebben meegemaakt. De afstand tussen onze woonplaatsen is inmiddels weer wat groter, maar de band blijft en de avonturen zullen blijven komen.

Qi, you came in after me, but graduated way before me. Good work! Thanks for all the good times in and outside of the office. I enjoyed our many table tennis matches. I treasure the memories of the many conferences we went to. And I often miss your extensive knowledge on all the Chinese restaurants in the Netherlands when picking a place to eat!

Anne-Rixt, je werd van een sportmaatje een feestmaatje. De gezamenlijke marathon in Zwolle en de vele andere hardloopwedstrijden zijn voor mij een sportief hoogtepunt in mijn leven. Soms niet op topsnelheid omdat de avond ervoor de cocktails rijkelijk hadden gevloeid. Mogelijk tijdens een van die memorabele strooptochten door de stad met de drie Epema zusters. Maar Rixie, wanneer is nou eigenlijk dat feest in 't Koetshuys?

De dagelijkse beslommeringen van het promovendileven werden een stuk leuker door alle gezellige collega's op de vierde verdieping. Leanne, Anne, Qi, Anna, Kim Gert, Brechtsje, Daan, Marloes, Jisca, Leyla, Lilian, Pepijn. De lunches, pubquizzen, Epi-uitjes, sportdagen, tafeltennispartijen en ander vertier maakte het UMCG een fijne werkplek.

Anna, al snel na mijn verhuizing naar Groningen werd je van 'vriendin-van' een goede vriendin. Ik heb warme herinneringen aan de vele avonden met proleten-tv, zingen en muziek maken. De avondjes aan de Schoolholm zorgden ervoor dat ik snel kon aarden in een nieuwe stad waar ik niet veel mensen

kende. De afstand tussen Utrecht en Enumachill speelt ons enigszins parten, maar ik beloof mijn best te doen om wat vaker naar het ommeland af te reizen.

Ook buiten Groningen deden velen vrienden hun best om mij te motiveren, al dan niet middels sarcastische grappen. Dank aan iedereen voor alle feestjes, festivals, vakanties en ander vertier dat het leven mooi maakt. Ik kan niet iedereen bij naam noemen, jullie weten wel wie jullie zijn! Sjoerd, dank voor de wetenschappelijke discussies en mij voorbereiden op de verdediging (ik kan me niet voorstellen dat debatteren met de corona moeilijker is dan met jou). Aart, de weddenschap wie er eerder klaar zou zijn heb je na al die jaren dan toch verloren. Snel maar weer naar een feestje? Walter, dank dat je me bijschoolt in sociale wetenschappen, humaniora en kwalitatief onderzoek, dank voor alle scherpe gesprekken en bovenal dank voor een fijne vriendschap. Gertjan, van collega tot goede vriend en trouw feestmaatje achtervolg jij me al een aantal jaar door het hele land en ik jou op de racefiets. Wie had kunnen raden dat we jaren nadat we met ons tweeën het Panaxea kantoor bevolkten we gebroederlijk met het vingertje wapperend bij een set van Vunkie zouden staan? Dat er nog vele festivals en kilometers in je wiel mogen volgen!

Jan-Maarten en Lia, helaas weten mijn huidige collega's van het uitblijven van mijn promotie, dus zo'n legendarische grap als destijds bij mijn master zit er helaas niet in. Het is mooi dat we elkaar zo veel jaar na de VU nog regelmatig zien. Jullie tips, aanmoedigingen en sarcastische grappen hebben me gemotiveerd om het proefschrift af te maken.

Matthijs, het is voor een groot deel aan jou te danken dat dit proefschrift er nu ligt en niet over een jaar of twee. Sommige directeuren werken met de wortel en de stok. Jij gooide die wortel weg en pakte nog een stok. En toen lag er drie weken later een proefschrift. Ik ben blij dat je me naar iMTA hebt gehaald waar ik veel van je kan leren en met je kan lachen. Ik kijk er naar uit nog lang met je te kunnen samenwerken.

Lieve Addy en Wim. Jullie bijdrage aan dit werk begon lang voordat ik er tien jaar geleden aan begon. Het begon namelijk bij de intellectuele stimulering in onze opvoeding en het voorbeeld van een publicerende wetenschapper in huis. Ik mag zo meteen dan wel gepromoveerd zijn, maar ik denk niet dat ik ooit een hogere *h-index* zal behalen dan jij, Addy, dus het is wel duidelijk wie van ons de echte wetenschapper is. Jullie steunden mijn keuzes in opleiding en loopbaan ongeacht welke weg ik was ingeslagen. Ook toen ik gaandeweg steeds meer vanuit biomedische wetenschappen naar de 'bla-bla' studie, voor anderen beter bekend als economie, ging. Jullie stonden ook altijd klaar voor eerste hulp bij verhuizingen, klussen en ander ongemak. En het was fijn om terug te kunnen vallen op de rust en afzondering van jullie huis om meters te maken met het schrijven. Ik kon wel eens kortaf zijn op vragen hoe het nou met het proefschrift ging. Dat was echter meer een frustratie ten aanzien van mijn eigen gedrag dan dat ik de vraag vervelend vond. Ik waardeer jullie interesse in mijn werk zeer en neem nu er geen druk meer op de ketel staat graag een keer de tijd om in detail uit te leggen wat ik in dit boekje allemaal heb opgeschreven.

Seroj, ook jouw invloed begon al vroeg. Als grote broer volgde ik jou in veel activiteiten, van Lego bouwen tot aan brommers sleutelen. Vaardigheden die onmisbaar zijn in wetenschappelijk onderzoek: creativiteit, technisch inzicht en durven aanpakken zijn in die tijd en door jouw voorbeeld gevormd. Ik ben zeer blij dat jij de omslag van mijn proefschrift hebt ontworpen. Ik vind het resultaat fantastisch en heb veel bewondering voor je talent om in korte tijd de abstracte verhalen van de verschillende hoofdstukken te vangen in een ontwerp dat zowel iets van het onderzoek verbeeld als ook een gezamenlijke interesse van ons.

Tot slot, het belangrijkst voor het laatst. Lieve Mirjam. We zitten hier samen op de bank te lachten om het feit dat ik echt niet weet wat ik hier over jou moet schrijven, terwijl we nog uitbrakken van een dag festivallen met een hele bups vrienden. Dat beeld zegt veel over ons. We kunnen samen goed lachen, we genieten van dezelfde dingen en we kunnen ons rijk rekenen met een hele

hoop fijne vrienden om ons heen. Wat mij betreft de belangrijkste dingen in het leven. Wat je voor mij betekent en hoe veel je me hebt geholpen, het is te veel om op te noemen. Die dingen zeg ik liever tegen je dan dat ik ze hier opschrijf. Een ding weet ik zeker: ons avontuur samen gaat nog veel langer duren dan mijn promotie!

RESEARCH INSTITUTE SHARE

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen/ University of Groningen.
Further information regarding the institute and its research can be obtained from our internet site: *http://www.share.umcg.nl/*

More recent theses can be found in the list below.
(supervisors are between brackets)

2019

Oosterhaven J
Hand eczema; impact, treatment and ourtcome measures
(*dr MLA Schuttelaar, prof PJ Coenraads*)

Postma DBW
Affordance-based control in running to catch fly balls
(*prof KAPM Lemmink, dr FTJM Zaal*)

Nuenen FM van
Screening of distress and referral need in Dutch oncology practice
(*prof HBM van de Wiel, dr JEHM Hoekstra-Weebers, dr SM Donofrio*)

Olthof SBH
Small-sided games in youth soccer; performance and behavior compared to the official match
(*prof KAPM Lemmink, dr WGP Frencken*)

Yin H
Epidemiology and treatment of mental disorders in a rapidly developing urban region in China; a study of prevalence, risk factors and e-applications
(*prof RA Schoevers, dr KJ Wardenaar*)

Kuipers DA
Design for transfer; figural transfer through metaphorical recontextualization in games for health
(*prof JPEN Pierie, dr JT Prins*)

Belak A
The health of segregated Roma: first-line views and practices
(*prof SA Reijneveld, prof A Geckova, dr JP van Dijk*)

Kieffer TEC
Adaptation and modulation of memory and regulatory T cells in pregnancy
(*prof SA Scherjon, dr MM Faas, dr JR Prins*)

Kuiper D
Health of offspring of subfertile couples
(*prof M Hadders-Algra, prof A Hoek, prof MJ Heineman*)

Huang Y
Use of real-world evidence in pharmacoeconomic analysis
(*prof FM Haaijer-Ruskamp, prof MJ Postma, dr P Vemer*)

Cuijpers LS
Coordination dynamics in crew rowing
(*prof KAPM Lemmink, dr FTJM Zaal, dr HJ de Poel*)

Dittmar D
Multiple aspects of contact allergy; immunology, patch test methodology and epidemiology
(*dr MLA Schuttelaar, prof PJ Coenraads*)

Beemster T
Quality improvement of vocational rehabilitation in patients with chronic musculoskeletal pain and reduced work participation
(*prof MF Reneman, prof MHW Frings-Dresen, prof CAM van Bennekom, dr JM van Velzen*)
28.08.2019 EXPAND

Geerse O
The impact of lung cancer; towards high-quality and patient-centered supportive care
(*prof HAM Kerstjens, prof MY Berger, dr TJN Hiltermann, dr AJ Berendsen*)

Dierselhuis EF
Advances of treatment in atypical cartilaginous tumours
(*prof SK Bulstra, prof AJH Suurmeijer, dr PC Jutte, dr M Stevens*)

Gils A van
Developing e-health applications to promote a patient-centered approach to medically unexplained symptoms
(*prof JGM Rosmalen, prof RA Schoevers*)

Notenbomer A
Frequent sickness absence; a signal to take action
(*prof U Bultmann, prof W van Rhenen, dr CAM Roelen*)

Bishanga DR
Improving access to quality maternal and newborn care in low-resource settings: the case of Tanzania
(*prof J Stekelenburg, dr YM Kim*)

Tura AK
Safe motherhood: severe maternal morbidity and mortality in Eastern Ethiopia
(*prof SA Scherjon, prof J Stekelenburg, dr TH van den Akker*)

Vermeiden CJ
Safe motherhood : maternity waiting homes in Ethiopia to improve women's access to maternity care
(*prof J Stekelenburg, dr TH van den Akker*)

Schrier E
Psychological aspects in rehabilitation
(*prof PU Dijkstra, prof JHB Geertzen*)

Malinakova K
Spirituality and health: their associations and measurement problems
(*prof SA Reijneveld, prof P Tavel, dr JP van Dijk*)

For earlier theses visit our website

ABOUT THE AUTHOR

Gimon de Graaf was born on the 15th of November 1984 in Nijmegen. In 2005 he received a bachelor of science from University College Utrecht, where he majored in life sciences and obtained a minor in economics. After serving a year as president of student association AEGEE-Utrecht, he enrolled in the master programme in biomedical sciences at the Vrije Universiteit Amsterdam. As part of this study, Gimon completed research internships at the European Centre for Disease Prevention and Control in Stockholm (2007) and the Ministry of Health of Benin in Cotonou (2008). In 2009 he started his PhD research at the department of epidemiology of the University Medical Center Groningen, the results of which are presented in this thesis. Afterwards, in 2014, Gimon started working as a consultant at Panaxea in Enschede. While at Panaxea he developed an online rapid health technology assessment tool for the Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing (MAFEIP), and contributed to a proposal for the BeNeLuxA collaboration on horizon scanning for pharmaceuticals. Since 2017 Gimon is working at the institute for Medical Technology Assessment (iMTA) in Rotterdam. There, he is able to apply his expertise on the early economic evaluation of tests within numerous consultancy projects. Recently, Gimon has completed a study on the value assessment and implementation of predictive tests in the Netherlands for ZonMw. Gimon is a proponent of a healthy work-life balance. He likes to go cycling or mountain biking, travels a lot, likes diving, enjoys making music and going to concerts and festivals.