EARLY HEALTH TECHNOLOGY ASSESSMENT OF POINT-OF-CARE AND LABORATORY DIAGNOSTICS

Methods and applications in acute and primary care

Michelle M.A. Kip

**Early health technology assessment of point-of-care and laboratory diagnostics:
Methods and applications in acute and primary care**

ISBN: 978-90-365-4485-6

DOI: 10.3990/1.9789036544856

Layout : Michelle M.A. Kip

Cover design: Tom J.H. Heerink

Printed by: Ipskamp Printing, Enschede

This thesis is part of the Health Science Series, HSS 18-20, department Health Technology and Services Research, University of Twente, Enschede, the Netherlands. ISSN: 1878-4968.

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

© Copyright 2018: Michelle M.A. Kip, Enschede, the Netherlands.

All rights reserved. No part of this publication may be reproduced without permission of the copyright holder.

EARLY HEALTH TECHNOLOGY ASSESSMENT OF POINT-OF-CARE AND LABORATORY DIAGNOSTICS

METHODS AND APPLICATIONS IN ACUTE AND PRIMARY CARE

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. T.T.M. Palstra,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 19 april 2018 om 14.45 uur

door

Michelle Maria Aleida Kip

geboren op 19 januari 1989
te Oldenzaal

Dit proefschrift is goedgekeurd door:

Prof. dr. G.C.M. Kusters (promotor)

Prof. dr. M.J. IJzerman (promotor)

Dr. ir. H. Koffijberg (copromotor)

SAMENSTELLING PROMOTIECOMMISSIE

Voorzitter

Prof. dr. Th.A.J. Toonen

Universiteit Twente

Promotors

Prof. dr. G.C.M. Kusters

Universiteit Twente

Prof. dr. M.J. IJzerman

Universiteit Twente

Copromotor

Dr. ir. H. Koffijberg

Universiteit Twente

Referent

Dr. R.M. Hopstaken

Saltro

Leden

Prof. dr. ing. A.J.H.M. Rijnders

Universiteit Twente

Prof. dr. ir. E.W. Hans

Universiteit Twente

Prof. dr. V. Scharnhorst

Universiteit Eindhoven

Prof. dr. G.J. Dinant

Universiteit Maastricht

Prof. dr. M.J. Postma

Rijksuniversiteit Groningen

Paranimfen

Marieke Weernink

Jorieke Postel

CONTENTS

Chapter 1	General introduction	9
Part I – Early health technology assessment of point-of-care tests to rule out acute coronary syndrome in primary care		23
Chapter 2	Early health technology assessment of future clinical decision rule aided triage of patients presenting with acute chest pain in primary care	25
Chapter 3	Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners	55
Chapter 4	The cost-utility of point-of-care troponin testing to diagnose acute coronary syndrome in primary care	89
Part II – Cost-effectiveness analysis of a procalcitonin-guided algorithm for antibiotic discontinuation in intensive care patients with sepsis		117
Chapter 5	A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis	119
Chapter 6	Procalcitonin to guide antibiotic stewardship in intensive care	141
Chapter 7	Cost-effectiveness of procalcitonin testing to guide antibiotic treatment duration in critically ill patients: results from a randomized controlled multicenter trial	145

Part III – Evaluation of an extensive laboratory analysis for the diagnosis and treatment of anaemia in general practice	167
Chapter 8 The effectiveness of a routine versus an extensive laboratory analysis in the diagnosis of anaemia in general practice	169
Chapter 9 Assessing the cost-effectiveness of a routine versus an extensive laboratory work-up in the diagnosis of anaemia in Dutch general practice	185
Part IV – Expert elicitation, stakeholder preferences and checklist development concerning the evaluation of diagnostic tests and biomarkers	221
Chapter 10 Using expert elicitation to estimate the potential impact of improved diagnostic performance of laboratory tests: a case study on rapid discharge of suspected non-ST elevation myocardial infarction patients	223
Chapter 11 Understanding the adoption and use of point-of-care tests in Dutch general practices using multi-criteria decision analysis	259
Chapter 12 Towards Alignment in the Reporting of Economic Evaluations of Diagnostic Tests and biomarkers: the AGREEDT checklist	293
Chapter 13 General discussion	325
Summary	335
Samenvatting	343
Dankwoord	351
Curriculum Vitae & Publications	359

1

General introduction

GENERAL INTRODUCTION

In the last few decades, the number of diagnostic tests available in medical practice has grown rapidly. These diagnostic tests encompass two main categories: diagnostic imaging devices and in vitro diagnostics (IVDs). According to the European Diagnostic Manufacturers Association, IVDs are defined as non-invasive tests that are used on biological samples (for example blood, urine or tissues) to determine the states of one's health [1]. IVDs provide valuable information on how the body is functioning, and they are used for prediction, screening, diagnosis, as well as for therapeutic monitoring of diseases [1]. They have a broad scope which ranges from simple self-tests (such as those for pregnancy testing and monitoring blood glucose levels) to sophisticated technologies performed in clinical laboratories [1].

Currently, there are over 40,000 different IVD products available for a wide range of medical conditions [2]. However, given the increasing demand for healthcare (attributable to the aging population in most developed countries and the increasing incidence of chronic diseases), as well as technological advances and the personalization of medicine, this number is expected to increase even further [3, 4]. In turn, this raises concerns about potential overuse of tests, associated negative health consequences, and rising healthcare costs [5-8]. This development further increases the need for thorough evaluations of the impact of new diagnostic tests, in terms of health benefits and costs, prior to their implementation in clinical practice. Besides, such evaluations also become increasingly important as the new regulation on medical devices, including IVD products, in the European Union places stricter requirements on clinical evaluation and post-market clinical follow-up of these medical devices, to address concerns over product safety and performance [9]. As this thesis focuses on IVDs, the term diagnostic test will be used to refer to IVDs instead of to diagnostic imaging devices (unless otherwise stated).

Point-of-care tests

Although diagnostic testing was historically performed in a central medical laboratory, during the past decades there has been a shift to testing sites in close proximity of patients. The use of these mostly biochemical testing modalities is referred to as point-of-care (POC) testing and the number of POC tests available has steadily increased since its widespread introduction [10]. Currently, POC testing encompasses a wide variety of procedures and technologies, for example pregnancy testing, cardiac biomarkers, blood glucose concentration, coagulation testing and malaria screening [11]. The most prominent advantage of POC testing involves the rapid turnaround time of test results, as compared with the 1 to 2 hour (or more) delay when using central laboratory testing [11]. In turn, this rapid availability of test results may improve both the efficiency and quality of care provided, for example by allowing earlier treatment initiation, by preventing unnecessary referrals from primary to secondary care, and by improving treatment adherence and patient satisfaction [11-13].

However, POC tests also create challenges that healthcare providers should address prior to implementing a new POC testing modality. A quality issue very specific to POC testing, is that those tests are often performed by nurses and clinical staff members with a limited technical background, indicating the need for adequate and continuous training [11, 14]. In addition, the implementation of POC testing may raise resistance because when compared to central laboratory testing, POC testing 1) is typically more expensive, 2) may have a lower diagnostic performance, and 3) may potentially lead to a shift in workload from the central laboratory to the nursing staff [11, 14].

Evaluation of diagnostic tests

In contrast to other medical interventions, diagnostic tests (both POC tests and tests performed in a central laboratory) primarily affect health outcomes indirectly (i.e., by guiding clinical decision making and by affecting downstream management decisions) [15]. Therefore, when evaluating diagnostic tests, it is crucial to take a broad perspective, and to investigate (all) downstream consequences of diagnostic tests prior to their implementation in clinical practice [16-23].

Clinical utility

The impact of diagnostic testing on health outcomes is referred to as 'clinical utility', which does not only incorporate potential benefits of subsequent or directed treatments, but also requires evaluating the full range of effects that a diagnostic test may have on patients [18]. This impact may be comprehensively measured in a randomized controlled trial (RCT), which is accepted to provide the highest quality of evidence to quantify the consequences of testing strategies [24]. However, in particular in early stages of product development, RCTs are often not feasible for ethical, financial, or other reasons [17, 25, 26]. For example, the restricted time frame of most RCTs does not allow assessment of the long-term consequences related to diagnostic testing strategies (e.g. radiation risk of imaging tests) [27]. In addition, owing to rapid technological developments of diagnostic tests, RCT results may be quickly outdated [27]. Furthermore, besides evaluating a test's clinical utility, a key principle of medical test evaluation is that the (potential) increase in costs should be in proportion with the improvement in health outcomes [18, 23]. Because of the abovementioned issues, decision analytic and cost-effectiveness models may provide a valid alternative to diagnostic RCTs. As such models can be used to combine evidence from different data sources (including studies on risks, diagnostic test accuracy, and evidence on outcomes), they can serve to assess and compare the impact of new diagnostic strategies on both health outcomes and costs [17, 25, 27, 28].

Health Technology Assessment

As healthcare budgets worldwide face an increasing pressure to reduce costs and improve efficiency without compromising health outcomes, managing the demand for laboratory tests and reducing inappropriate requesting is highly important [29]. This calls for a thorough

evaluation of the costs and benefits of (new) diagnostic tests, regarding both its implementation and actual application in clinical practice. Currently, many countries use health economic concepts as a basis for their health policy decision-making [30]. Simultaneously, the use of health economic tools by manufacturers, to inform investment decisions in the development process of medical technologies, is increasing [31-33]. The rationale behind these concepts is to compare the potential costs and benefits of new healthcare technologies with usual care strategies, which is referred to as Health Technology Assessment (HTA) [30, 34]. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) defines HTA as a form of policy research that examines the consequences of the application of a healthcare-technology both on the short- and long-term [35]. In such an assessment, both safety, efficacy, real-world effectiveness, patient-reported outcomes, costs and cost-effectiveness can be assessed, as well as the social, legal, ethical and political impact [35].

The form of economic evaluation generally used in healthcare is cost-effectiveness analysis (CEA), which is typically aimed at maximizing health outcomes at a constrained budget [36]. The most commonly used metric to quantify the impact on health outcomes in such an evaluation is a Quality-Adjusted Life Year, or QALY. This single metric combines the impact of healthcare programs and interventions on individuals' life years gained with their health-related quality of life [36]. Although the use of QALYs as the measure of effect in a CEA is often referred to as a cost-utility analysis (CUA) instead of as a CEA, those terms are sometimes used interchangeably. In order to allow a comparison of the cost-effectiveness of different technologies, and across different disease areas, the incremental cost-effectiveness ratio (ICER) can be calculated. The ICER expresses the additional unit cost per extra unit of effect (e.g. QALY) in the intervention group relative to its comparator [36]:

$$\text{Incremental cost-effectiveness ratio (ICER)} = \frac{\text{Costs}_{\text{intervention}} - \text{Costs}_{\text{comparator}}}{\text{QALYs}_{\text{intervention}} - \text{QALYs}_{\text{comparator}}}$$

Thus, if a cost-effectiveness threshold (also called ceiling ratio or willingness-to-pay threshold) is known and agreed upon, HTA can be used to investigate whether the costs of a diagnostic test are justified by its benefits. Thereby, HTA can support coverage and adoption decisions of new technologies in healthcare from a variety of perspectives, including economic, ethical and regulatory [31]. Besides guiding the abovementioned investment decisions, this technique is also used to increase the efficiency of product development [31, 37].

Health economic evaluation of new tests in patient pathways

Although evaluating diagnostic accuracy is an essential step in diagnostic test evaluation, most diagnostic accuracy studies do not compare the new test with existing tests or existing testing pathways [20]. As knowledge of other features of a new test, for example its availability and invasiveness, can help predict its most likely way of use, Bossuyt et al. have proposed an approach in which the accuracy of new tests is compared with that of existing tests or

existing pathways. They have defined three roles of a new test [20]: replacement, triage and add-on. To illustrate this, a test may replace an existing test when it is more accurate, less invasive, risky, or uncomfortable for patients, easier to perform or to interpret, or quicker to yield results. However, in order to predict whether a new test can replace an existing one, the diagnostic accuracy of both the existing and the new test needs to be investigated. The choice of the appropriate study design(s) should depend on the role of the test within the existing pathway. Also, although novel tests may be less accurate than existing tests, they may have other benefits, e.g. in terms of turnaround time, simplicity or costs. These tests may be used as triage, in which the new test is used before the existing test or testing pathway. The results from this triage test are then used to determine which patients continue the testing pathway. This may reduce the use of existing tests that are more invasive, cumbersome, or expensive. Alternatively, a test may be positioned after the existing pathway as an add-on in a subgroup of patients. Those tests may either increase the sensitivity of the existing pathway, or may be used to limit the number of false positives after the existing pathway [20]. The eventual role of a new test should be determined based on its impact on the accuracy of the existing testing pathway. Thereby, defining roles for new and existing tests is intended to contribute to evidence-based healthcare [20].

Early Health Technology Assessment

As mentioned previously, regular HTA focuses on an economic analysis of the costs of the medical technology and on the benefits to the patient [32]. Those benefits to the patient for example involve extending a patient's lifetime, and/or improving a patient's quality of life. Such a regular HTA is generally performed once device-specific data (e.g. technical accuracy) have become available [32]. However, owing to the high number of (new) diagnostic tests available, rising healthcare costs, as well as the many different roles a test may fulfill in a testing pathway, assessment of new medical tests should preferably be performed in early stages of product development, which is referred to as 'early HTA'. Early HTA has been defined as "all methods used to inform industry and other stakeholders about the potential value of new medical products in development, including methods to quantify and manage uncertainty" [38]. Thus, besides the assessment of a product's safety, effectiveness, and cost-effectiveness, early HTA considers other aspects relevant to inform upon decision making, including implementation barriers, possible uptake, product specifications, patient needs and market size [31, 39]. Therefore, early HTA does not only include assessing the expected costs and benefits of new medical products, but also involves methods to estimate the value of such new medical products and to identify preferences with regard to these products. Thereby, the insights obtained from early HTA can be used to guide decisions in early stages of technology development, to define minimum performance thresholds required for the new technology as compared to currently available technologies, and it may also serve to support pricing and reimbursement strategies in early product development stages [31, 33, 38-40]. Such assessment is expected to be of added value in improving the likelihood of successful product implementation [38]. Early HTA

considers existing regulatory requirements in several countries as well as mechanisms for obtaining reimbursement depending on the added value produced. Thereby, early HTA differs from regular HTA as it primarily informs research and development instead of government (agencies) in decisions regarding coverage [33, 39, 40].

Multi-criteria decision analysis

As mentioned previously, although cost-effectiveness may occasionally be the most dominant criterion in the appraisal of healthcare technologies, other factors are also incorporated in the decision-making process [34]. For example, aspects that may not be fully captured in a QALY include the severity of disease and the different perspectives of stakeholders on the diverse benefits of technological innovations in healthcare [30, 41]. Therefore, decision tools that can systematically integrate both the benefits and costs of medical innovations from multiple perspectives are of added value [30]. Multi-criteria decision analysis (MCDA) offers a set of methods and approaches for taking multiple criteria into account in decision making. MCDA can be used in early HTA to obtain clarity on which criteria are relevant, as well as on the importance of each criterion, thereby increasing the consistency, transparency, and legitimacy of these decisions [42].

Expert elicitation

As early HTA is aimed at evaluating new healthcare technologies in early development stages, the reference to 'early' in this setting also reflects the uncertainty in (and the lack of) clinical evidence that is available. Therefore, in particular in early HTA, ways to reduce or alleviate this uncertainty should be considered [38]. One way to handle evidence gaps in early-stage health economic models is by the use of expert elicitation, which can be used to estimate both unknown probabilities and unknown effect sizes [38]. This technique is also referred to as belief elicitation, and guidelines on the reporting of such elicitations have recently been published [43].

Disease areas considered within this thesis

Diagnostic tests can be performed in many different disease areas and for a variety of different purposes, ranging from monitoring a chronic condition in primary care, to setting treatment thresholds in the acute care setting. The aim of this thesis was to get insight in the full potential impact of diagnostic tests in terms of costs and health outcomes further downstream the management and treatment pathway. To achieve this, the current thesis focuses on three very different disease areas, each of which encompasses a major burden to society. These three disease areas (and accompanying patient populations) involve patients presenting with symptoms suggestive of acute coronary syndrome (ACS) in primary care, patients with sepsis in the intensive care, and anaemia patients presenting in primary care. In the following paragraphs, each of these three disease areas will be briefly introduced.

Acute coronary syndrome

Coronary heart disease is a major cause of mortality, accounting for about 7.4 million deaths in 2012 worldwide [44]. ACS is one of the most common manifestations of coronary heart disease, with chest pain as a leading symptom [45]. Owing to the potentially life-threatening consequences of ACS, accurate and rapid diagnosis remains crucial. However, almost 200,000 patients per year in the Netherlands contact their general practitioner (GP) with chest pain, and ~29% of these patients are immediately referred to the hospital [46]. The estimated incidence of a cardiovascular problem among these patients is however low (8-15%) [47-51]. Thus, adequate clinical management of chest pain patients represents a major challenge for GPs. In addition, these (unnecessary) referrals lead to unnecessary burden to the healthcare system as well as patient distress. Although POC cardiac markers (e.g. troponin) have become available, which may facilitate the rapid exclusion of ACS, their cost-effectiveness is unknown. This thesis will therefore investigate the expected value of these POC tests, as well as their potential impact on costs and health outcomes.

Sepsis

Despite ongoing advances in medical technology and clinical care, sepsis remains a common cause of morbidity and mortality among intensive care patients [52]. Approximately 2% of all hospitalized patients, and 11-20% of all intensive care patients are treated for severe sepsis [53]. As the presentation of sepsis is often non-specific, diagnosing patients with sepsis remains challenging [54]. Although rapid and adequate antibiotic therapy is crucial, prolonged duration of antibiotic therapy is undesirable because of increasing antibiotic resistance [55]. To support physicians in the decision on when to discontinue antibiotic therapy, a procalcitonin (PCT)-based antibiotic treatment algorithm can be used, although the cost-effectiveness of such an algorithm (as compared to current practice) is unknown.

Anaemia

Another common medical problem carrying substantial burden to both the health and quality of life of many individuals as well as to the healthcare system, concerns anaemia [56-65]. Adequate diagnosis and early initiation of appropriate treatment is therefore essential [66]. However, anaemia is often not considered a disease in itself, but instead as a sign of an underlying condition, or as a consequence of aging [67]. To illustrate this, a previous study revealed that 48% of nursing home residents were found to have anaemia [68].

The three most common underlying causes of anaemia are (1) iron deficiency anaemia, (2) anaemia of chronic disease, and (3) renal anaemia [69]. Besides anamnesis and physical examination, laboratory tests are required to diagnose the underlying cause. However, despite the use of laboratory tests, no underlying cause can be identified in 28-52% of anaemia patients [70]. This thesis will therefore examine the cost-effectiveness of two laboratory-based approaches for diagnosing (and treating) anaemia among patients presenting in primary care.

THIS THESIS

This thesis was set up to (1) investigate the health economic impact of (new) diagnostic tests or diagnostic testing strategies based on case studies in the abovementioned disease areas, (2) investigate which aspects affect the implementation and use of these diagnostic tests, and (3) provide recommendations and guidance on conducting (early) HTAs of diagnostic tests.

As mentioned previously, before a decision regarding the implementation of (new) diagnostic tests can be made, insight in their expected cost-effectiveness is required. In this respect, **chapter 2 to 4** will discuss aspects that affect the decision to implement and use a POC test to rule out ACS in patients presenting with chest pain in primary care. In addition, these chapters will provide insights in the minimum requirements such a test has to fulfill, as well as into the expected cost-effectiveness of the POC troponin test. More specifically, **chapter 2** describes an early HTA regarding the use of a POC test in this setting, and describes the minimum reduction in unnecessary hospital referrals of chest pain patients that should be achieved to make the use of such a test cost-effective. Subsequently, **chapter 3** shows the result of a questionnaire among GPs, about their preferences regarding the use of a POC troponin test to diagnose or exclude ACS, as well as its expected impact on referral decisions. Following this, **chapter 4** describes an extensive cost-effectiveness analysis concerning the expected impact of a POC troponin test as compared to current practice.

Besides decreasing referral rates of patients, diagnostic tests can have many other benefits. For example, the PCT test can be used to guide the duration of antibiotic therapy in intensive care patients with sepsis. **Chapter 5** illustrates the cost-effectiveness of such a PCT-guided antibiotic treatment algorithm, when using the results of (international) published literature. Elaborating on these results, **chapter 6** shows the response to a publication of a Dutch multicenter RCT into the safety and effectiveness of such a PCT-guided algorithm. This correspondence illustrates the impact of this algorithm on costs and antibiotic treatment duration when the results from this RCT are used to calculate the direct healthcare-related costs. **Chapter 7** reports the results of a trial-based analysis into the cost-effectiveness of this PCT-guided algorithm, based on the results from the previously published RCT.

Although the abovementioned analyses all consider the use and expected cost-effectiveness of (relatively) new tests or testing modalities, (early) HTA is also of added value in evaluating the expected impact of using a new *combination* of existing tests. Therefore, **chapter 8** and **chapter 9** illustrate the impact of using an extensive, fixed set of 14 laboratory tests on the correct diagnosis and treatment of anaemia in primary care, as compared to the situation in which GPs decide themselves which tests to request (the routine work-up). **Chapter 8** describes the impact of both laboratory work-ups on the percentage of patients diagnosed with the correct underlying cause of anaemia. Following this, **chapter 9** shows the results of a cost-effectiveness analysis, comparing the extensive laboratory work-up with the routine work-up,

in terms of the difference in treatment costs in primary care and/or referral to secondary care per additional patient that is diagnosed with the correct underlying cause of anaemia.

As scientific evidence is often incomplete when performing health economic evaluations, expert elicitations may be used to provide estimations for model input parameters. The use of those expert elicitations is illustrated in **chapter 10**. This chapter concerns a case study on the cost-effectiveness of the combined use of heart-type fatty acid-binding protein (H-FABP), high-sensitive troponin, and copeptin, as compared to conventional serial high-sensitive troponin testing, to rapidly rule out non-ST elevation myocardial infarction in patients presenting with chest pain in the emergency department.

Even though the results of a cost-effectiveness analysis indicate that the use of a diagnostic test is cost-effective, the use of these diagnostic tests in clinical practice is often limited. This indicates that further insights in factors that hamper or facilitate test implementation and use are required. Therefore, **chapter 11** analyzes features that affect the adoption of diagnostic tests in clinical practice, which is illustrated using a case study on two POC tests in primary care (i.e. the POC glycated haemoglobin [HbA1c] test and the POC C-reactive protein [CRP] test).

Obviously, diagnostic test evaluation is complex, especially because diagnostic tests and biomarkers (in general) do not affect health outcomes directly, but instead have an indirect impact by affecting patient management decisions. Because of this complexity, many economic evaluations of diagnostic tests do not consider all aspects relevant to the estimation of cost-effectiveness of diagnostic tests and biomarkers. Therefore, in this study a comprehensive reporting checklist is developed in **chapter 12**, consisting of 43 items that ideally should be included (or explicitly excluded) in a model-based health economic evaluation of a diagnostic test or biomarker.

In **chapter 13**, the main findings from the different studies are discussed. In addition, both the implications for clinical practice as well as recommendations for further research will be formulated for each of the three main aims of this thesis. Finally, the main conclusions will be drawn.

REFERENCES

- European Diagnostic Manufacturers Association, About EDMA. 2014.
- Rohr, U.P., et al., The Value of In Vitro Diagnostic Testing in Medical Practice: A Status Report. *PLoS One*, 2016. 11(3): p. e0149856.
- Wiener, J.M. and J. Tilly, Population ageing in the United States of America: implications for public programmes. *Int J Epidemiol*, 2002. 31(4): p. 776-81.
- Akhmetov, I. and R.V. Bubnov, Assessing value of innovative molecular diagnostic tests in the concept of predictive, preventive, and personalized medicine. *EPMA J*, 2015. 6: p. 19.
- Price, C.P., Evidence-based laboratory medicine: is it working in practice? *Clin Biochem Rev*, 2012. 33(1): p. 13-9.
- Thomas, R.E., et al., Interventions at the laboratory level to reduce laboratory test ordering by family physicians: Systematic review. *Clin Biochem*, 2015. 48(18): p. 1358-65.
- van Walraven, C. and C.D. Naylor, Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA*, 1998. 280(6): p. 550-8.
- Hickner, J., et al., Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. *J Am Board Fam Med*, 2014. 27(2): p. 268-74.
- European Commission, Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on medical devices, and amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009. 2012.
- St John, A. and C.P. Price, Existing and Emerging Technologies for Point-of-Care Testing. *Clin Biochem Rev*, 2014. 35(3): p. 155-67.
- Larsson, A., R. Greig-Pylypczuk, and A. Huisman, The state of point-of-care testing: a European perspective. *Ups J Med Sci*, 2015. 120(1): p. 1-10.
- Laurence, C.O., et al., Patient satisfaction with point-of-care testing in general practice. *Br J Gen Pract*, 2010. 60(572): p. e98-104.
- Gialamas, A., et al., Does point-of-care testing lead to the same or better adherence to medication? A randomised controlled trial: the PoCT in General Practice Trial. *Med J Aust*, 2009. 191(9): p. 487-91.
- Lewandowski, K., K. Gregory, and D. Macmillan, Assuring quality in point-of-care testing: evolution of technologies, informatics, and program management. *Arch Pathol Lab Med*, 2011. 135(11): p. 1405-14.
- Bossuyt, P.M.M. and K. McCaffery, Additional Patient Outcomes and Pathways in Evaluations of Testing, in Medical Tests-White Paper Series. 2009: Rockville (MD).
- Price, C.P. and R.H. Christenson, Evaluating new diagnostic technologies: perspectives in the UK and US. *Clin Chem*, 2008. 54(9): p. 1421-3.
- Moons, K.G., Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem*, 2010. 56(4): p. 537-41.
- Bossuyt, P.M., et al., Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem*, 2012. 58(12): p. 1636-43.
- Ferrante di Ruffano, L., et al., Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*, 2012. 344: p. e686.
- Bossuyt, P.M., et al., Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*, 2006. 332(7549): p. 1089-92.
- Gluud, C. and L.L. Gluud, Evidence based diagnostics. *BMJ*, 2005. 330(7493): p. 724-6.
- Van den Bruel, A., et al., The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *J Clin Epidemiol*, 2007. 60(11): p. 1116-22.
- Horvath, A.R., et al., From biomarkers to medical tests: the changing landscape of test evaluation. *Clin Chim Acta*, 2014. 427: p. 49-57.
- Schunemann, H.J., et al., Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*, 2008. 336(7653): p. 1106-10.
- Biesheuvel, C.J., D.E. Grobbee, and K.G. Moons, Distraction from randomization in diagnostic research. *Ann Epidemiol*, 2006. 16(7): p. 540-4.
- Bossuyt, P.M., J.G. Lijmer, and B.W. Mol, Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*, 2000. 356(9244): p. 1844-7.
- Koffijberg, H., B. van Zaane, and K.G. Moons, From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol*, 2013. 13: p. 12.
- Schaafsma, J.D., et al., Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol*, 2009. 62(12): p. 1248-52.
- Fryer, A.A. and W.S. Smellie, Managing demand for laboratory tests: a laboratory toolkit. *J Clin Pathol*, 2013. 66(1): p. 62-72.
- Wahlster, P., et al., Balancing costs and benefits at different stages of medical innovation: a systematic review of Multi-criteria decision analysis (MCDA). *BMC Health Serv Res*, 2015. 15: p. 262.
- IJzerman, M.J. and L.M. Steuten, Early assessment of medical technologies to inform prod-

- uct development and market access: a review of methods and applications. *Appl Health Econ Health Policy*, 2011. 9(5): p. 331-47.
32. Pietzsch, J.B. and M.E. Pate-Cornell, Early technology assessment of new medical devices. *Int J Technol Assess Health Care*, 2008. 24(1): p. 36-44.
 33. Vallejo-Torres, L., et al., Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach. *Int J Technol Assess Health Care*, 2008. 24(4): p. 459-64.
 34. Devlin, N.J. and J. Sussex, Incorporating multiple criteria in HTA: methods and processes. 2011, Office of Health Economics: London. p. 60.
 35. Bingefors, K., et al., Health care cost, quality, and outcomes: ISPOR book of terms. 2003, Lawrenceville, NJ: International Society for Pharmacoeconomics & Outcomes Research.
 36. Briggs, A. and K. Claxton, *Decision Modelling for Health Economic Evaluation*, in NY. 2006, Oxford University Press: New York.
 37. Chalmers, I., Biomedical research: are we getting value for money? *Significance*, 2006. 1: p. 172-5.
 38. IJzerman, M.J., et al., Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *Pharmacoeconomics*, 2017. 35(7): p. 727-740.
 39. Markiewicz, K., J.A. van Til, and I.J. MJ, Medical devices early assessment methods: systematic literature review. *Int J Technol Assess Health Care*, 2014. 30(2): p. 137-46.
 40. IJzerman, M.J., Webinar: Early Health Technology Assessment: potential to inform pricing and reimbursement during R and D. 2016, ISPOR New Zealand.
 41. Cohen, J. and W. Looney, What is the value of oncology medicines? *Nat Biotechnol*, 2010. 28(11): p. 1160-3.
 42. Thokala, P., et al., Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*, 2016. 19(1): p. 1-13.
 43. Iglesias, C.P., et al., Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations. *Pharmacoeconomics*, 2016. 34(11): p. 1161-1172.
 44. World Health Organization, *World Health Statistics: monitoring health for the SDGs, sustainable development goals*. 2016: Geneva.
 45. Roffi, M., et al., 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur Heart J*, 2016. 37(3): p. 267-315.
 46. Okkes, I.M., et al., Continue morbiditeitsregistratie Nijmegen. Episodes of care in family practice. Epimiological data based on the routine use of the International Classification of Primary Care (ICPC) in the Transition project of the Academic Medical Center/University of Amsterdam (1985-2003). 2005, Academic Medical Center/University of Amsterdam, Department of Family Medicine: Amsterdam.
 47. Bosner, S., et al., Chest pain in primary care: epidemiology and pre-work-up probabilities. *Eur J Gen Pract*, 2009. 15(3): p. 141-6.
 48. Nilsson, S., et al., Chest pain and ischaemic heart disease in primary care. *Br J Gen Pract*, 2003. 53(490): p. 378-82.
 49. Ruigomez, A., et al., Chest pain in general practice: incidence, comorbidity and mortality. *Fam Pract*, 2006. 23(2): p. 167-74.
 50. Rutten, F.H., et al., NHG-standaard acuut coronaair syndroom (eerste herziening). *Huisarts Wet*, 2012. 55(12): p. 564-70.
 51. Buntinx, F., et al., Chest pain in general practice or in the hospital emergency department: is it the same? *Fam Pract*, 2001. 18(6): p. 586-9.
 52. Dellinger, R.P., et al., Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med*, 2013. 41(2): p. 580-637.
 53. Damen, J., et al., De hemodynamische gevolgen van ernstige sepsis: pathofysiologie en een richtlijn voor de behandeling. *NJCC*, 2002: p. 19-29.
 54. Angus, D.C. and R.S. Wax, Epidemiology of sepsis: an update. *Crit Care Med*, 2001. 29(7 Suppl): p. S109-16.
 55. Van Boeckel, T.P., et al., Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *Lancet Infect Dis*, 2014. 14(8): p. 742-750.
 56. Carmel, R., Anemia and aging: an overview of clinical, diagnostic and biological issues. *Blood Rev*, 2001. 15(1): p. 9-18.
 57. Riva, E., et al., Association of mild anemia with hospitalization and mortality in the elderly: the Health and Anemia population-based study. *Haematologica*, 2009. 94(1): p. 22-8.
 58. Lucca, U., et al., Association of mild anemia with cognitive, functional, mood and quality of life outcomes in the elderly: the "Health and Anemia" study. *PLoS One*, 2008. 3(4): p. e1920.
 59. Robinson, B., Cost of anemia in the elderly. *J Am Geriatr Soc*, 2003. 51(3 Suppl): p. S14-7.
 60. Smith, R.E., Jr., The clinical and economic burden of anemia. *Am J Manag Care*, 2010. 16 Suppl Issues: p. S59-66.
 61. Shavelle, R.M., R. MacKenzie, and D.R. Paculdo, Anemia and mortality in older persons: does the type of anemia affect survival? *Int J Hema-*

- tol, 2012. 95(3): p. 248-56.
62. Zakai, N.A., et al., A prospective study of anemia status, hemoglobin concentration, and mortality in an elderly cohort: the Cardiovascular Health Study. *Arch Intern Med*, 2005. 165(19): p. 2214-20.
 63. Steensma, D.P. and A. Tefferi, Anemia in the elderly: how should we define it, when does it matter, and what can be done? *Mayo Clin Proc*, 2007. 82(8): p. 958-66.
 64. Culleton, B.F., et al., Impact of anemia on hospitalization and mortality in older adults. *Blood*, 2006. 107(10): p. 3841-6.
 65. Penninx, B.W., et al., Anemia in old age is associated with increased mortality and hospitalization. *J Gerontol A Biol Sci Med Sci*, 2006. 61(5): p. 474-9.
 66. Droogendijk, J., et al., Adherence to the Dutch general practitioner anemia guideline - a retrospective cohort study. *Perfusion*, 2013. 6(26): p. 192-198.
 67. Guralnik, J.M., et al., Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. *Blood*, 2004. 104(8): p. 2263-8.
 68. Artz, A.S., et al., Prevalence of anemia in skilled-nursing home residents. *Arch Gerontol Geriatr*, 2004. 39(3): p. 201-6.
 69. Stouten, K., et al., Relevance of mean corpuscular volume for the evaluation of anaemia in general practice. Manuscript submitted, 2017.
 70. Oosterhuis, W.P., et al., Prospective comparison of the flow chart for laboratory investigations for anaemia from the Dutch College of General Practitioners' guideline 'Anaemia' with a self-developed, substantive and logistical alternative flow chart'. *Ned Tijdschr Geneesk*, 2007. 151: p. 2326-32.

PART I

Early health technology assessment
of point-of-care tests to rule out acute
coronary syndrome in primary care

2

Early health technology assessment of future clinical decision rule aided triage of patients presenting with acute chest pain in primary care

Robert T.A. Willemsen
Michelle M.A. Kip
Hendrik Koffijberg
Ron Kusters
Frank Buntinx
Jan FC Glatz
Geert-Jan Dinant
and The 'RAPIDA' – Study Team ('RAPIDA': RAPid Test for
Investigating Complaints Possibly Due to Acute Coronary
Syndrome)

This chapter has been published as:

R.T.A. Willemsen, M.M.A. Kip, H. Koffijberg, R. Kusters, F. Buntinx, J.F.C. Glatz, G-J. Dinant.
*Early health technology assessment of future clinical decision rule aided triage of patients
presenting with chest pain in primary care.* Prim Health Care Res Dev, 2017: p. 1-13

Reprinted with permission.

ABSTRACT

The objective of the paper is to estimate the number of patients presenting with chest pain suspected of acute coronary syndrome (ACS) in primary care and to calculate possible cost effects of a future clinical decision rule (CDR) incorporating a point-of-care test (POCT) as compared with current practice. The annual incidence of chest pain, referrals and ACS in primary care was estimated based on a literature review and on a Dutch and Belgian registration study. A health economic model was developed to calculate the potential impact of a future CDR on costs and effects (ie, correct referral decisions), in several scenarios with varying correct referral decisions. One-way, two-way, and probabilistic sensitivity analyses were performed to test robustness of the model outcome to changes in input parameters. Annually, over one million patient contacts in primary care in the Netherlands concern chest pain. Currently, referral of eventual ACS negative patients (false positives, FPs) is estimated to cost €1,448 per FP patient, with total annual cost exceeding 165 million Euros in the Netherlands. Based on 'international data', at least a 29% reduction in FPs is required for the addition of a POCT as part of a CDR to become cost-saving, and an additional €16 per chest pain patient (ie, 16.4 million Euros annually in the Netherlands) is saved for every further 10% relative decrease in FPs. Sensitivity analyses revealed that the model outcome was robust to changes in model inputs, with costs outcomes mainly driven by costs of FPs and costs of POCT. If POCT-aided triage of patients with chest pain in primary care could improve exclusion of ACS, this CDR could lead to a considerable reduction in annual healthcare costs as compared with current practice.

BACKGROUND

A clinical decision rule (CDR) based on history and physical examination to safely rule out acute coronary syndrome (ACS) in primary care is not available [1-5]. Therefore, in primary care, up to 73% of patients with new or altered chest pain are immediately referred by the GP to the emergency department [6]. However, only a minority of those patients (up to 26% in literature) are subsequently diagnosed with an acute life threatening disease, for example ACS ('true referrals' or 'true positives' (TPs)) [6-10]. Patients that were referred and were found to be ACS negative ('false referrals' or 'false positives' (FPs)) were diagnosed with alternative diseases with advantageous courses [8, 9]. In this context, the term 'false' is used to indicate a referral in absence of ACS afterwards. The referral itself however is undisputed, as it is the result of a GP unable to exclude a potentially life threatening disease. On the other hand, incidentally, ACS is present in patients that were initially not referred ('false non-referrals' or 'false negatives' (FNs)) [8, 9].

ACS negative referrals (FPs) pose a significant burden on healthcare resources, and reduction of FPs can lead to increased patient comfort, while decreasing costs [11-13]. Therefore, specificity of a future CDR for ACS should be higher – and sensitivity should at the least be maintained – as compared with current practice that is based on a GPs' clinical judgement only. Recently, the efficiency of similar diagnostic processes in primary care was improved by introducing a cost-effective CDR, combining point-of-care tests (POCTs) with clinical findings, leading to less prescription of unnecessary antibiotics in lower respiratory tract infections and less unnecessary referral for suspected pulmonary embolism [14-16].

The availability of a validated CDR, incorporating a POCT measuring a biomarker of myocardial damage (eg, high-sensitive troponin (hsTn) or heart-type fatty acid binding protein (H-FABP)) is anticipated, however not yet available [17-20]. The majority of GPs expect future POCTs to be of added value in ruling out ACS [21].

Objectives

1. To estimate the number of patients with chest pain in primary care, their referral rates and the incidence of ACS among these patients;
2. To assess the minimum required reduction in ACS negative referrals (FPs) due to a future CDR, incorporating clinical findings and a POCT, to become cost-saving, assuming that the number of non-referrals among ACS patients (FNs) equals current practice;
3. To assess the impact of a relative decrease in ACS negative referrals (FPs), with decrements of 10%;
4. To assess the combined impact of simultaneously varying the referrals among non-ACS patients (FPs) and the costs of the POCT, to determine which combinations are expected to save costs.

METHODS

Estimation of annual patient numbers based on literature

International data and a Dutch registration study were used as separate sources to estimate numbers of true and false referrals (referred to as true and false positives respectively (TPs and FPs)) as well as true and false non-referrals (referred to as true and false negatives respectively (TNs and FNs)) in primary care in the Netherlands [22]. The international data were obtained through an extensive literature search. Those data will be referred to as 'international data', and will be used as the base case scenario in the remainder of this paper. However, as registration studies are rare, it was expected that part of the data were to be derived from studies describing relevant data that were primarily designed to meet other objectives. Therefore, we searched PubMed and Embase from January 1989 to May 2017 for chest complaints in primary care (see Appendix 1 for an overview of the literature search and strategy). Articles assumed relevant based on title/abstracts were read in order to select all studies supplying relevant data on the incidence of chest complaints in primary care, referral rates, and final diagnoses (when available). Additional relevant articles were identified from the references in these selected papers. Besides these sources, we used Dutch registry data (governmental data, data from the Dutch Heart Foundation and data from the Dutch Central Statistics Agency) [23-30]. Data from the literature review as well as from the Dutch registries were pooled and mean values with 95% confidence intervals (95% CI) were calculated. In most cases, the included studies did not describe all probabilities in the patient pathway (eg, the set of studies that described the percentage of patients that is referred to secondary care, differed from the set of studies that described the proportion of ACS positive cases among those referred patients). Therefore, the sample sizes of the pooled estimates differed across the different parameters. Eventually, the pooled international data were translated to estimated referral rates and incidence of ACS in the Netherlands. As a second scenario to model the patient pathway of Dutch patients with chest pain in primary care, a Dutch/Belgian registration study was used, and this scenario will be referred to as 'NL and B data' [22].

Model diagnostic process, resource use

A health economic model was developed to estimate the costs of the full diagnostic work-up. Costs were estimated based on several sources (see Table 1), and expressed in 2016 Euros [31]. The analysis was performed from a healthcare perspective, incorporating all direct medical costs that occur from the moment a patient presents with chest pain in primary care, until the patient was either referred to and diagnosed and treated in secondary care, or sent home following the GP consultation (without referral). As this time horizon is less than one year, discounting of costs and effects was not required. Calculations were performed using unit costs for assessment in primary care including POCT testing, ambulance transport to the hospital, and assessment in secondary care. Exclusion of ACS by a GP costs €18 without POCT, and increases to €63 when cost for usage of a POCT of €45 is included (based on the expected price of a H-FABP POCT which is currently in development). The cost for every patient that

is assessed in a hospital for cardiac analysis with and without the eventual presence of an underlying ACS is estimated at €5,735 and €1,426 respectively, including hospital transport by ambulance.

Outcome measures

The primary effectiveness measure was defined as the percentage of patients in whom ACS (including both unstable angina (UA) and acute myocardial infarction (AMI)) was correctly diagnosed or excluded when using the CDR as compared with current practice. The incremental cost-effectiveness ratio (ICER) was therefore expressed as incremental costs per patient when using a CDR (including POCT), as compared with current practice, and divided by the difference in the number of patients in whom the correct referral decision is made in both work-ups.

Table 1. Model input: cost data. Cost prices for relevant events in in- or excluding ACS in The Netherlands. Cost prices are based on the following sources: double consultation price GP in The Netherlands (*), estimations from manufacturer (**), cost price requested at large ambulance service in South of The Netherlands (***), average diagnosis-treatment combination tariffs of considerable number of Dutch hospitals of varying type (small, large, academic, urban, rural) (****). Abbreviations: 95%CI = 95% confidence interval; ACS = acute coronary syndrome; CCU = Coronary Care Unit; GP = general practitioner; H-FABP = Heart-type fatty acid binding protein; PCI = percutaneous coronary intervention; POCT = point of care test; VAT = value added tax

Parameter	Value [95% CI]	Distribution
Consultation at tariff GP (double)*	€18.00 [€10.18 - €27.93]	Gamma
POCT (including finger prick needle, and VAT)**	€45.00 [€26.08 - €69.52]	Gamma
Ambulance transport medium to high urgency (medical personnel A1/A2 drive, overhead costs call center, and VAT)***	€750.00 [€429.70 - €1,172.51]	Gamma
Analysis CCU, no ACS (diagnostic tests, medical personnel, hospital stay 1-2 days)****	€676.00 [€386.75 - €1,054.49]	Gamma
Analysis CCU, ACS present (diagnostic tests, medical personnel, hospital stay 3 days, PCI)****	€4,985.00 [€2,823.36 - €7,673.56]	Gamma

One- and two-way sensitivity analysis

A two-way deterministic sensitivity analysis was performed to obtain insight into the combined impact of simultaneously varying the cost of the POCT and the %FPs [32]. For subsequent analysis (ie, one-way and probabilistic sensitivity analyses), the minimum required relative decrease in FPs to obtain a strategy that is cost-saving compared with current practice was applied in the scenarios with POCT (while assuming that costs of the POCT remain unaffected). Following this, to identify which individual cost parameters drive the model outcome, given

fixed costs of the POCT and the minimum required relative decrease in FPs, we conducted a one-way deterministic sensitivity analysis [32]. In the one-way sensitivity analysis, the impact of a 25% decrease and increase in all cost input parameters on the costs per patient presenting with suspected ACS in primary care was analysed. As the impact on FP referrals in the CDR + POCT strategy was arbitrarily chosen (based on the minimum required reduction in FPs), it was decided to only incorporate the impact on costs in this one-way sensitivity analysis. In addition, this avoids double counting, as both the numerator and the denominator of the ICER are affected by a change in correct referral decisions.

Table 2. Model input: effectiveness data for three different base cases. Number of chest pain patients in the Netherlands and probabilities of ‘true positives’ (ACS positive referrals), ‘false positives’ (ACS negative referrals), ‘false negatives’ (ACS negative non-referrals) and ‘true negatives’ (ACS negative non-referrals), as well as the accompanying 95% CI and the distribution applied. Numbers and distributions are presented for three different scenarios: based on pooled analysis of international literature (international data), based on a Dutch/ Belgian cohort study (NL and B data) and based on a combination of sources (combined data). As the international data are based on the largest set of patients, those were used in the base case analysis of this article. *References: [6, 33-47]. **References: [22]. ***References: [6, 22, 34-47]. 95% CI = 95% confidence interval; ACS = acute coronary syndrome; B = Belgium; FN = false negatives; FP = false positives; NL = the Netherlands; TN = true negatives; TP = true positives.

Scenario (estimated annual number of chest pain patients for the Netherlands [95%CI])	Referred		Not referred		Distribution
	ACS (TP) [95% CI]	No ACS (FP) [95% CI]	ACS (FN) [95% CI]	No ACS (TN) [95% CI]	
International data* (n = 1,054,729 [1,047,881 - 1,061,578])	3.4% [2.8% - 4.0%]	10.8% [10.2% - 11.5%]	1.5% [0.8% - 2.2%]	84.3% [83.5% - 85.0%]	Beta
NL and B data** (n = 862,960)	6.1% [4.0% - 7.9%]	8.1% [6.0% - 11.1%]	0.3% [0.0% - 2.5%]	85.5% [80.6% - 89.1%]	Beta
Combined data*** (n = 1,054,729 [1,047,881 - 1,068,427])	6.8% [5.5% - 7.0%]	21.9% [21.6% - 23.3%]	0.3% [0.0% - 1.7%]	71.0% [69.6% - 71.7%]	Beta

Probabilistic sensitivity analysis

Distributions were assigned to all model parameters [32]. Subsequently, random samples were drawn for all model input parameters simultaneously. An overview of the type of distribution used for each input parameter, as well as the accompanying 95% confidence intervals (95% CI) is provided in Tables 1 and 2. A probabilistic sensitivity analysis (PSA) based on Monte Carlo simulation with 10,000 samples was performed to determine the effect of joint uncertainty in all model input parameters on model outcome.

RESULTS

Estimation of relevant patient numbers

In our literature search, 1,900 articles were assessed. The majority of articles were not considered relevant after screening the title and abstract. Of the remaining 17 articles, three additional articles were eliminated after full text evaluation, since no relevant data were obtainable from these papers [48-50]. Two articles reported as references in the remaining 14 articles were added, resulting in a final number of 16 articles that were included (Appendix 1) [6, 33-47]. In addition, data from regional and national Dutch and Belgian databases were used [24-30, 51]. In Appendix 6, all relevant data representing parts of the patient flow of patients with chest complaints, found in the selected articles and databases, are presented. All data were converted to absolute patient numbers in the Netherlands (Appendix 7). Numbers of TPs, FPs, TNs and FNs were calculated (see Table 2). Besides the estimated patient numbers based on the two previously defined data sets (ie, base case 'international data' and the 'NL and B data'), a third scenario (referred to as 'combined data') was defined. This data set was based on the international data, although one article was excluded because the health system and time setting in which this study was performed were considered not comparable with the current health system in the Netherlands [33]. In addition, in this scenario the incidence of FNs was based on the 'NL and B data', as the incidence of FNs in the base case 'international data' seemed higher than observed in Dutch daily practice [22]. All analyses were performed for each of these three scenarios, with the international data used as base case, as these data are based on the largest set of patients. The main results based on the 'NL and B data' and the 'combined data' are also presented in the text of this paper, the accompanying figures can be found in Appendix 2-5.

Patient flow, GP's sensitivity and specificity in current practice

The results of the literature review indicate that annually in primary care in the Netherlands, 1 054 729 [95% CI 1,047,881–1,061,578] patients consult a GP for chest pain. The referral rate among these patients was found to be 14.2% [95% CI 14.0–14.4]. Sensitivity and specificity of a GPs judgement in the current setting (not aided by a CDR), are 69 and 89% for the 'international', 95 and 91% for 'NL and B', and 96 and 76% for the 'combined data', respectively.

Required reduction of ACS negative referrals (FPs) due to a future CDR, effect of further reduction of FPs

When the cost price of a future POCT is set at €45, the minimum required relative reductions in FPs for the POCT strategy to become cost-saving are 29.0, 39.5 and 14.5% for the 'international data', the 'NL and B data', and the 'combined data', respectively (see Figure 1 and Appendix 2 and 3). In Table 3, the impact of a further relative reduction in FP rates on costs is shown for all three scenarios. For every additional absolute 10% reduction in %FPs, average additional cost savings per patient are €16 when using 'international data', €12 for 'NL and B data' and €31 for 'combined data' per chest pain patient.

Impact on health outcomes and costs

The minimum required relative decrease in %FPs, as obtained from the two-way SA, was used as input into the PSA, while assuming that costs of the PoCT would remain unchanged (ie, €45). The results of this analysis on the average total costs (both per patient as well as the total costs in the Netherlands), are shown in Table 4. To visualize the constitution of those total costs, results are split up into costs that are attributable to TPs, FPs, FNs and TNs, and as the corresponding fraction of total costs. The result of the 10,000 Monte Carlo simulations is shown in Figure 2 for the 'international data', whereas results for the 'NL and B data' and 'combined data' are shown in Appendix 4a and 4b, respectively.

Figure 1 Two-way SA for 'international data'. Deterministic two-way SA showing the combined effect of a relative reduction in ACS negative referrals (FPs, on x-axis), and of a variation in costs of a POCT (on y-axis), on the difference in total costs between POCT and current practice (ie, without POCT). The analysis was performed based on the 'international data'. When assuming that POCT would only impact the %FPs and incur costs of the POCT test (and leave all other model input parameters unaffected), a relative reduction of at least 29.0% in FPs is required to make the POCT strategy become cost-saving (as represented by the black square, assuming a cost price of a POCT test of €45.00). ACS = acute coronary syndrome; FPs = false positives; POCT = point-of-care test; two-way SA = two-way sensitivity analysis.

Table 3. Effect of a stepwise reduction of ACS negative referrals (FPs) on costs. This table shows the impact of steps of a 10% relative decrease in FP referrals (deterministic). At the top of the table, the costs for each of the base case scenarios is shown, depending on whether the POCT is used. Absolute and relative effects are given for all three scenarios ('international data', 'NL and B data' and 'combined data' respectively). If no reduction in FPs is achieved (0% change = base case for all three scenarios where the POCT is used) costs in all three scenarios will rise with the cost of a POCT (€ 45.13), as compared to current daily practice where a POCT is not used. Abbreviations: ACS = acute coronary syndrome; B = Belgium; FP = 'false positives'; NL = the Netherlands; POCT = point-of-care test.

Change in %FPs	International data		NL and B data		Combined	
	Base case without POC:		Base case without POC:		Base case without POC:	
	€366.57		€485.01		€722.46	
Effect on costs	Base case with POC:		Base case with POC:		Base case with POC:	
	€411.71		€530.14		€757.59	
	Effect on costs	% effect	Effect on costs	% effect	Effect on costs	% effect
-100%	€ -110.20	-30.1%	€ -71.93	-14.8%	€ -286.05	-39.6%
-90%	€ -94.66	-25.8%	€ -60.22	-12.4%	€ -236.73	-32.8%
-80%	€ -79.13	-21.6%	€ -48.52	-10.0%	€ -205.41	-28.4%
-70%	€ -63.60	-17.4%	€ -36.81	-7.6%	€ -174.09	-24.1%
-60%	€ -48.07	-13.1%	€ -25.10	-5.2%	€ -142.77	-19.8%
-50%	€ -32.53	-8.9%	€ -13.40	-2.8%	€ -111.46	-15.4%
-40%	€ -17.00	-4.6%	€ -1.69	-0.3%	€ -80.14	-11.1%
-30%	€ -1.47	-0.4%	€ 10.01	+2.1%	€ -48.82	-6.8%
-20%	€ 14.07	+3.8%	€ 21.72	+4.5%	€ -17.50	-2.4%
-10%	€ 29.60	+8.2%	€ 33.43	+6.9%	€ 13.81	+1.9%
0% (base case)	€ 45.13	+12.3%	€ 45.13	+9.3%	€ 45.13	+6.2%

Sensitivity of model outcome to changes in cost input parameters

The sensitivity of the model outcome to changes in individual cost input parameters, was measured using a one-way sensitivity analysis. The results are shown in tornado diagrams (see Figure 3 for 'international data', and Appendix 5 for 'NL and B data' and 'combined data'). Results indicate that the model outcome (expressed as cost per patient) is robust to changes in input parameters in all three scenarios. In addition, in all three scenarios, the model outcome is most sensitive to changes in costs of the POCT, while it is less sensitive to changes in costs of ambulance transportation to the hospital and costs of analysis at the coronary care unit among FPs. As each patient is assumed to undergo one consultation at the GP, and because the probability of correctly diagnosing ACS was assumed not to be affected by the POCT, the model outcome was not affected by changes in those input parameters.

Table 4 Costs per patient, converted to patient numbers in the Netherlands. This table presents where the average costs per patient are composed of, by splitting up those average costs into costs that are attributable to ACS positive referrals (TPs), ACS negative referrals (FPs), ACS positive non-referrals (FNs), and ACS negative non-referrals (TNs), as well as the accompanying percentage. Costs are given for all three base case scenarios ('international data', 'NL and B data' and 'combined data' respectively, without using POCT), and based on the results of the probabilistic analysis.

*References: [6, 33-47]

**Reference: [22]

***References: [6, 22, 34-47]

ACS = acute coronary syndrome; B = Belgium; FN = false negatives; FP = false positives; NL = the Netherlands; POCT = point-of-care test; TN = true negatives; TP = true positives.

Scenario	Referred				Not referred				Total costs in the	
	ACS (TP)		No ACS (FP)		ACS (FN)		No ACS (TN)		Netherlands	
	Costs	%	Costs	%	Costs	%	Costs	%	Costs per patient	
International data*	€194.00 (€116.35 – 293.90)	52.9	€157.22 (€108.36 - €215.09)	42.9	€0.28 (€0.12 - €0.50)	0.1	€15.25 (€8.71 - €23.66)	4.2	€366.71 (€268.07 - €483.72)	€385,736,500 (€282,739,424 - €510,189,897)
NL and B data**	€338.82 (€182.95 - €546.32)	71.3	€120.88 (€72.67 - €184.89)	25.4	€0.12 (€0.00 - €0.47)	0.0	€15.40 (€8.76 - €23.87)	3.2	€475.27 (€296.30 - €712.19)	€400,543,692 (€255,691,986 - €614,595,666)
Combined data***	€371.23 (€224.78 - €559.47)	52.5	€322.99 (€221.85 - €445.24)	45.7	€0.08 (€0.00 - €0.33)	0.0	€12.79 (€7.31 - €20.00)	1.8	€707.07 (€518.55 - €936.05)	€767,833,684 (€546,928,106 - €987,276,217)

Figure 2. Incremental cost-effectiveness plane based on 'international data'. This figure shows the result of 10 000 model simulations (PSA), and the mean value, based on the international data. Costs of a POCT are set at € 45, and reduction of ACS negative referrals (FPs) is assumed to be 29.0% (cost-neutral as compared with current practice, see Figure 1). ACS = acute coronary syndrome; FPs = false positives; POCT = point-of-care test; PSA = probabilistic sensitivity analysis.

DISCUSSION

Summary of main findings

We estimated that in the Netherlands (population 17 million) annually ~1 million patient contacts with GPs are about chest pain. In 14% of these contacts, direct referral to a cardiologist is made. Eventually, no more than a quarter of these referred patient is diagnosed with ACS (3.4% of all chest pain patients). As a result, 10.8% of all chest pain patients are referred while they eventually are diagnosed as not having ACS (FP). The estimated annual number of FPs in the Netherlands is 113,911, representing an economic burden of 162 million Euros. Improving triage of patients presenting with chest complaints to their GP could lead to a considerable reduction of FPs and thus to a reduction in both direct healthcare costs and patients' distress. We estimated the impact of a CDR incorporating a POCT in this diagnostic process, on the accompanying costs and effects (ie, number of patients in whom the correct referral decision is made). When using an estimated cost price of a POCT of €45, introduction of such test would be cost-saving if a relative reduction in FPs of at least 29% is achieved. This would imply a reduction in percentage of unnecessary referred patients from 10.8 to 7.7%. Such decrease would account for a cost saving of €47,106,755. Besides, for every 10% further reduction in %FPs, beyond the reduction of 29% where cost neutrality was reached, €16 per chest pain patient –

referred or not referred – is saved, accounting for an annual saved amount of ~16 million Euros in the Netherlands. In the two alternative scenarios, a required reduction in %FPs of 39.5% for the ‘NL and B data’ and 14.4% for the ‘combined data’ was found. Such a reduction of FPs seems achievable when compared with results of similar studies in the field of suspected pulmonary embolism [15]. Yet, a lower cost price of a POCT can attenuate the required reduction in FPs. Halving a POCT’s cost price to €22.50 leads to a minimum required reduction of FPs of only 14.5% for the POCT to become cost-saving, when based on the ‘international data’. Although the effect of preventing ACS negative referrals on societal costs has not been included in the current analysis, including those costs for both patients (and family or caregivers) would likely have increased the estimated cost savings that can be achieved by implementation of a POCT-aided CDR.

Figure 3. Tornado diagram of one-way SA's for 'international data'. Tornado diagram showing the impact of changes in input parameters on the difference in costs, based on 'international data'. Costs of a POCT are set at €45, and the reduction of ACS negative referrals (FPs) is assumed to be 29.0% (cost-neutral situation as compared with current practice, see Figure 1). All input parameters were varied with 25% below and above the mean value. ACS = acute coronary syndrome; CCU = coronary care unit; FPs = false positives; PCI = percutaneous coronary intervention; one-way SA = one-way sensitivity analysis; POCT = point-of-care test; VAT = value-added tax.

Sensitivity analysis

In the one-way model-sensitivity analysis, the model outcome proved to be robust for varying the model input parameters with -25% and +25% from the base case value. However, the starting point for variation in costs was based on a cost price for the POCT of €45, which was based on the cost prognosis of a POCT H-FABP test in development, and results may have been different when differently priced or different cardiac marker based POCTs (eg, POCT troponin) would have been used. Therefore, a wider range of costs was applied in the two-way sensitivity

analysis, as this allows to apply the model to a wider range of POCT cardiac markers for use in primary care.

Strength and weakness

A strength of this study is that, to our knowledge, this study is the first to describe the possible financial benefit of a CDR incorporating a POCT in chest pain patients presenting in primary care. We synthesized all available and relevant evidence, from different studies and countries, to make the best possible estimation of prevalence of chest complaints and referral rates in primary care. Moreover, we repeated the analysis using three different data sets for patient numbers of chest pain patients in primary care in the health economic model. As the base case model outcome is based on several international studies, it is likely that the results regarding the effectiveness can be generalized to other countries. However, as costs may differ strongly between countries, country-specific cost estimates are required to make reliable per-country calculations.

In our analyses it was assumed that the test is performed in all patients. However, when a CDR is already positive after only scoring a patient's clinical findings by the GP, the patient will most likely be immediately referred without performing a POCT. Therefore, as the use of POCT will likely be more sophisticated in daily clinical practice, the costs of POCT have (conservatively) been overestimated in the current analysis. On the other hand, wide availability of a validated POCT in the future could lower the threshold for using this test in daily practice, as expected based on previous research [21].

Still, some uncertainties remain. Despite the fact that patient numbers and referral rates are based on a thorough review of available literature, the number of studies that could eventually be included was limited for some model input parameters. In addition, some of those studies were relatively small. Although it might have been most straight-forward to use a Dirichlet distribution to estimate all four probabilities (TP, FP, TN and FN) simultaneously, this would have required to use a small number of patients across this distribution, thereby largely overestimating the uncertainty in model outcomes. Therefore, a Beta distribution was used instead, which allows a two-step approach. First, the probability of a patient being referred was estimated (which was based on a large number of patients), followed by estimating the probability of being either TP or FP (among referrals), or TN or FN (among non-referrals). As we assumed that the number of patients with and without ACS had to remain constant, patients could only switch from TP to FN, and from TN to FP (and vice versa). Consequently, we could not simultaneously incorporate uncertainty in the number of patients that either have or do not have ACS. However, as the result of Monte Carlo simulations shows that the uncertainty in costs is expected to be limited, and because the analysis has been performed for three different scenarios, we consider it unlikely that incorporating this uncertainty would have changed the conclusions.

Furthermore, a validated CDR does not yet exist and until such a diagnostic tool is available, the actual diagnostic performance of a CDR incorporating a POCT remains uncertain. In addition, an issue that could not be included in the current model (as it was considered too complex for the scope of this analysis), is the possible presence of other severe diseases causing chest pain in patients where ACS is excluded. However, in recent studies, nearly all FPs are diagnosed with not life threatening diseases [6, 22]. Another possible limitation may be caused by the fact that the number of FNs was assumed to remain constant in the current analysis. In current practice, patients that are initially not referred but are eventually diagnosed as having ACS (in the following days after assessment by a GP) might be under registered. Moreover, a POCT could reduce FN referrals, which likely leads to health gain. In addition, those FN rates were found to be relatively high in our 'international data' set, but appear to be much lower in recent studies [22], and based on the authors' experience in daily clinical practice. Therefore, the 'combined' data set is assumed to provide an estimate better reflecting current practice in the Netherlands, and in many other countries with similar healthcare systems. However, because of the assumptions that had to be made, the 'international' data were conservatively chosen as base case scenario in the current analysis.

CONCLUSION

If the use of a CDR including a POCT can reduce unnecessary referrals of chest pain patients to secondary care, this could considerably reduce healthcare costs. Our study provides insights in the minimum requirements (regarding the relative reduction in FPs, as well as the costs of the POCT) for the CDR strategy including a POCT to become cost-saving.

ACKNOWLEDGEMENTS

The authors thank VGZ health insurance for their assistance in determination of cost prices of cardiac care in secondary care facilities. They thank GGD Zuid-Limburg for their assistance in determination of cost prices of ambulance transport.

FINANCIAL SUPPORT

The study is funded by means of an unrestricted grant by FABPulous BV, the company that develops point-of-care H-FABP-tests. FABPulous BV had no role in data collection, data management, data analysis or interpretation of data. Publication of possible unfavorable outcome of our study was guaranteed.

CONFLICTS OF INTEREST

J.G. is chief scientific officer (CSO) at FABPulous BV. The remaining authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

- Grijseels, E.W., et al., Implementation of a pre-hospital decision rule in general practice. Triage of patients with suspected myocardial infarction. *Eur Heart J*, 1996. 17(1): p. 89-95.
- Winters, M.E. and S.M. Katzen, Identifying chest pain emergencies in the primary care setting. *Prim Care*, 2006. 33(3): p. 625-42.
- Body, R., et al., The value of symptoms and signs in the emergent diagnosis of acute coronary syndromes. *Resuscitation*, 2010. 81(3): p. 281-6.
- Gencer, B., et al., Ruling out coronary heart disease in primary care patients with chest pain: a clinical prediction score. *BMC Med*, 2010. 8: p. 9.
- Bosner, S., et al., Chest pain in primary care: is the localization of pain diagnostically helpful in the critical evaluation of patients?—A cross sectional study. *BMC Fam Pract*, 2013. 14: p. 154.
- Bruins Slot, M.H., et al., Diagnostic value of a heart-type fatty acid-binding protein (H-FABP) bedside test in suspected acute coronary syndrome in primary care. *Int J Cardiol*, 2013. 168(2): p. 1485-9.
- Chambers, J. and C. Bass, Atypical chest pain: looking beyond the heart. *QJM*, 1998. 91(3): p. 239-44.
- Buntinx, F., et al., Chest pain in general practice or in the hospital emergency department: is it the same? *Fam Pract*, 2001. 18(6): p. 586-9.
- Kohn, M.A., et al., Prevalence of acute myocardial infarction and other serious diagnoses in patients presenting to an urban emergency department with chest pain. *J Emerg Med*, 2005. 29(4): p. 383-90.
- Hocaoglu, C., M.Y. Gulec, and I. Durmus, Psychiatric comorbidity in patients with chest pain without a cardiac etiology. *Isr J Psychiatry Relat Sci*, 2008. 45(1): p. 49-54.
- Graff, L.G., et al., Impact on the care of the emergency department chest pain patient from the chest pain evaluation registry (CHEPER) study. *Am J Cardiol*, 1997. 80(5): p. 563-8.
- Dumville, J.C., et al., Non-cardiac chest pain: a retrospective cohort study of patients who attended a Rapid Access Chest Pain Clinic. *Fam Pract*, 2007. 24(2): p. 152-7.
- Mourad, G., et al., Societal costs of non-cardiac chest pain compared with ischemic heart disease—a longitudinal study. *BMC Health Serv Res*, 2013. 13: p. 403.
- Cals, J.W., et al., C-reactive protein point of care testing and physician communication skills training for lower respiratory tract infections in general practice: economic evaluation of a cluster randomized trial. *J Eval Clin Pract*, 2011. 17(6): p. 1059-69.
- Geersing, G.J., et al., Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. *BMJ*, 2012. 345: p. e6564.
- Little, P., et al., Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. *Lancet*, 2013. 382(9899): p. 1175-82.
- Nilsson, S., et al., Point-of-Care Troponin T Testing in the Management of Patients with Chest Pain in the Swedish Primary Care. *Int J Family Med*, 2013. 2013: p. 532093.
- Glatz, J.F. and R. Mohren, Plasma reference value of heart-type fatty acid binding protein, the earliest available plasma biomarker of acute myocardial infarction. *Health*, 2013. 5(8): p. 1206-9.
- Willemsen, R.T., et al., The value of signs, symptoms and plasma heart-type fatty acid-binding protein (H-FABP) in evaluating patients presenting with symptoms possibly matching acute coronary syndrome: background and methods of a diagnostic study in primary care. *BMC Fam Pract*, 2014. 15: p. 203.
- Glatz, J.F. and R. Renneberg, Added value of heart-type fatty acid-binding protein as plasma marker for the early evaluation of suspected acute coronary syndrome. *Clin Lipidol*, 2014. 9: p. 205-20.
- Kip, M.M.A., et al., Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners. *Prim Health Care Res Dev*, 2017: p. 1-12.
- Hoorweg, B.B., et al., Frequency of chest pain in primary care, diagnostic tests performed and final diagnoses. *Heart*, 2017. 103(21): p. 1727-1732.
- Okkes, I.M., et al., Episodes of care in family practice. Epidemiological data based on the routine use of the International Classification of Primary Care (ICPC) in the Transition Project of the Academic Medical Center/University of Amsterdam, in ICPC in the Amsterdam Transition Project. 2005, Academic Medical Center / University of Amsterdam: Amsterdam.
- Catholic University Leuven, Derived from analysis of registration data Catholic University Leuven. 2013.
- Academic Hospital Maastricht, Data derived from analysis of registration data Academic Hospital Maastricht. 2013.
- Dutch Heart Foundation. Percentage of patients with AMI in an emergency department. Data from Dutch Heart Foundation. 2013; Available from: <http://www.hartstichting.nl/hart-en-vaatziekten/feiten-en-cijfers-hart-en-vaatziekten>.
- Centraal Bureau voor de Statistiek (CBSa). Cardiovascular mortality. 2014; Available from: <http://statline.cbs.nl/Statweb/publication/?D->

- M=SLNL&PA=7233&D1=655-659&D2=0&D3=0&D4=0,4,9,(1-1)-l&VW=T.
28. Centraal Bureau voor de Statistiek (CBSb). Inhabitants the Netherlands. 2014; Available from: <http://www.cbs.nl/nl-NL/menu/themas/bevolking/cijfers/extra/bevolkingsteller.htm>.
 29. Centraal Bureau voor de Statistiek (CBSc). GP visits in The Netherlands. 2014; Available from: <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=81178NED&D1=0-2,12-14,25-27,41,46,53&D2=a&D3=a&D4=0&D5=l&HD=130701-1102&HDR=G4,G3,T&STB=G1,G2> and <http://www.cbs.nl/nl-NL/menu/themas/gezondheid-welzijn/publicaties/artikelen/archief/2013/2013-3854-wm.htm>.
 30. Centraal Bureau voor de Statistiek (CBSd). Causes of death in The Netherlands. 2014; Available from: http://statline.cbs.nl/StatWeb/selection/default.aspx?VW=T&DM=SLNL&PA=7052_95&D1=0%2c79-91&D2=0&D3=0&D4=a&HDR=G1%2cG2%2cG3&STB=T.
 31. Centraal Bureau voor de Statistiek. Consumentenprijzen; prijsindex 2006 = 100. 2016 [cited 2016 March 31]; Available from: <http://statline.cbs.nl/Statweb/>
 32. Briggs, A. and K. Claxton, Decision Modelling for Health Economic Evaluation, in NY. 2006, Oxford University Press: New York.
 33. Rosser, W., et al., An exploratory report of chest pain in primary care. A report from ASPN. *J Am Board Fam Pract*, 1990. 3(3): p. 143-50.
 34. Klinkman, M.S., D. Stevens, and D.W. Gorenflo, Episodes of care for chest pain: a preliminary report from MIRNET. Michigan Research Network. *J Fam Pract*, 1994. 38(4): p. 345-52.
 35. Svavarsdottir, A.E., et al., Chest pain in family practice. Diagnosis and long-term outcome in a community setting. *Can Fam Physician*, 1996. 42: p. 1122-8.
 36. Carroll, K., et al., Prevalence and management of coronary heart disease in primary care: population-based cross-sectional study using a disease register. *J Public Health Med*, 2003. 25(1): p. 29-35.
 37. Nilsson, S., et al., Chest pain and ischaemic heart disease in primary care. *Br J Gen Pract*, 2003. 53(490): p. 378-82.
 38. Bakx, C., et al., First myocardial infarction in a Dutch general practice population: trends in incidence from 1975-2003. *Br J Gen Pract*, 2005. 55(520): p. 860-3.
 39. Ruigomez, A., et al., Chest pain in general practice: incidence, comorbidity and mortality. *Fam Pract*, 2006. 23(2): p. 167-74.
 40. Koek, H.L., et al., Incidence of first acute myocardial infarction in the Netherlands. *Neth J Med*, 2007. 65(11): p. 434-41.
 41. Verdon, F., et al., Chest pain in daily practice: occurrence, causes and management. *Swiss Med Wkly*, 2008. 138(23-24): p. 340-7.
 42. Bosner, S., et al., Chest pain in primary care: epidemiology and pre-work-up probabilities. *Eur J Gen Pract*, 2009. 15(3): p. 141-6.
 43. Yeh, R.W., et al., Population trends in the incidence and outcomes of acute myocardial infarction. *N Engl J Med*, 2010. 362(23): p. 2155-65.
 44. Haasenritter, J., et al., Ruling out coronary heart disease in primary care: external validation of a clinical prediction rule. *Br J Gen Pract*, 2012. 62(599): p. e415-21.
 45. Soler, J.K., et al (a). An international comparative family medicine study of the Transition Project data from the Netherlands, Malta and Serbia. Is family medicine an international discipline? Comparing incidence and prevalence rates of reasons for encounter and diagnostic titles of episodes of care across populations. *Fam Pract*, 2012. 29(3): p. 283-98.
 46. Andersson, P.O., et al., Consequences of high-sensitivity troponin T testing applied in a primary care population with chest pain compared with a commercially available point-of-care troponin T analysis: an observational prospective study. *BMC Res Notes*, 2015. 8: p. 210.
 47. Frese, T., et al., Chest pain in general practice: Frequency, management, and results of encounter. *J Family Med Prim Care*, 2016. 5(1): p. 61-6.
 48. Sox, H.C., Jr., et al., Using the patient's history to estimate the probability of coronary artery disease: a comparison of primary care and referral practices. *Am J Med*, 1990. 89(1): p. 7-14.
 49. Soler, J.K., et al (a), An international comparative family medicine study of the Transition Project data from the Netherlands, Malta and Serbia. Is family medicine an international discipline? Comparing diagnostic odds ratios across populations. *Fam Pract*, 2012. 29(3): p. 299-314.
 50. Haasenritter, J., et al., Causes of chest pain in primary care--a systematic review and meta-analysis. *Croat Med J*, 2015. 56(5): p. 422-30.
 51. Okkes, I.M., et al., Continue morbiditeitsregistratie Nijmegen. Episodes of care in family practice. Epidemiological data based on the routine use of the International Classification of Primary Care (ICPC) in the Transition project of the Academic Medical Center/University of Amsterdam (1985-2003). 2005, Academic Medical Center/University of Amsterdam, Department of Family Medicine: Amsterdam.

APPENDIX 1

Supplementary Figure S1. Literature search. Visualisation of search for quantitative data on care process of chest pain in primary care, contributing to supplementary Tables S1 and S2.

APPENDIX 2

2

Supplementary Figure S2. Two-way SA for 'NL and B data'. Deterministic two-way SA showing the combined effect of a relative reduction in ACS negative referrals (FPs, on X-axis), and of a variation in costs of a POCT (on y-axis), on the difference in total costs between POCT and current practice (i.e. without POCT). The analysis was performed based on the 'NL and B data'. When assuming that POCT would only impact the %FPs and incur costs of the POCT test (and leave all other model input parameters unaffected), a relative reduction of 39.5% in FPs is required to make the PoCT strategy cost-neutral (as represented by the black square, assuming a cost price of a POCT test of €45,00). Abbreviations: ACS = Acute Coronary Syndrome; FPs = 'false-positives'; POCT = point of care test; two-way SA = two-way sensitivity analysis

APPENDIX 3

Supplementary Figure S3. Two-way SA for 'combined data'. Deterministic two-way SA showing the combined effect of a relative reduction in ACS negative referrals (FPs, on X-axis), and of a variation in costs of a POCT (on y-axis), on the difference in total costs between POCT and current practice (i.e. without POCT). The analysis was performed based on the 'combined data'. When assuming that POCT would only impact the %FPs and incur costs of the POCT test (and leave all other model input parameters unaffected), a relative reduction of 14.5% in FPs is required to make the POCT strategy cost-neutral (as represented by the black square, assuming a cost price of a POCT test of €45,00). Abbreviations: ACS = Acute Coronary Syndrome; FP = 'false-positives'; POCT = point of care test; two-way SA = two-way sensitivity analysis

APPENDIX 4

2

a

b

Supplementary Figures S4 and S5. Incremental cost-effectiveness plane, based on 'NL and B data' (a), and 'combined data' (b). This figure shows the result of 10,000 model simulations (PSA), and the mean value, based on the 'NL and B' data (a) or the combined data (b). Costs of a POCT are set at € 45, and reduction of ACS negative referrals (FPs) is assumed to be 39.5% (a) and 14.5% (b) (cost-neutral as compared to current practice, see figures S2 and S3, respectively). Abbreviations: ACS = Acute coronary syndrome; FPs = 'false positives'; POCT = point of care test; PSA = probabilistic sensitivity analysis

APPENDIX 5

a

b

Supplementary Figures S6 and S7. Tornado diagram of one-way SA's for 'NL and B data' (a) and 'combined data' (b). Tornado diagram showing the impact of changes in input parameters on the difference in costs, based on the 'NL and B' data (a) and 'combined data' (b). Costs of a POC test are set at € 45, and the reduction of ACS negative referrals (FPs) is assumed to be 39.5% (a) and 14.5% (b) (cost-neutral situation as compared to current practice, see figures S2 and S3, respectively). All input parameters were varied with 25% below and above the mean value. Abbreviations: ACS = Acute coronary syndrome; CCU = Coronary Care Unit; FPs = 'false positives'; GP = general practitioner; PCI = percutaneous coronary intervention; one-way SA = one-way sensitivity analysis; POC = point of care; VAT = value added tax

APPENDIX 6

Supplementary Table S1

	Source	Country	Year	Population	Outcome	Weighted average of estimated patient numbers and percentages, with 95%CI between [] when appropriate and estimations of variation (multiplication of more than one 95% CI) between []	Estimated patient number for the Netherlands 2013, as used in figure 2, with 95%CI between [] when appropriate and estimations of variation (multiplication of more than one 95% CI) between []
a) Population in the Netherlands	CBS [27-30]	NL	2013	All inhabitants of the Netherlands	Total population: N = 16,779,575 Population aged 20-79: N = 12,205,982	-	-
b) Number of annual GP consultations	CBS [27-30]	NL	2013	All inhabitants of the Netherlands	Population visiting GP at least once: N = 12,015,601 Total number of GP consultations: N = 68,488,928	-	-
c) Incidence of chest pain as RFE GP	CMR Nijmegen [51]	NL	1995-2008	All 35,000 registered patients of selected GPs	Incidence of chest pain or pressure encountering a GP: 10.9/1000/yr	pooled data: 61,039 cases of chest pain in 3,964,274 consultations;	annual number of consultations for chest pain, NL 2013:
	Soler et al [45]	NL	1995-2005	All 15,318 registered patients of selected GPs	Incidence of chest pain or pressure encountering a GP: 30.6/1000/yr	weighted average of incidence of chest pain 1.54% [1.53-1.55]	1.54% of 68,488,928 consultations = 1,054,729 [1,047,881-1,061,578] (value C)
	Ruigómez et al [39]	UK	1996	All 3,000,000 registered patients of appr. 1500 GPs	Incidence of new-onset chest pain as reason for encountering a GP: 15.5/1000/yr	(after exclusion of Rosser [33])	1,054,729 [1,047,881-1,068,427] (value C)
	Nilsson et al [37]	SW	1998-2000	All 16,152 registered patients of selected GPs, aged 20-79	Incidence new onset chest pain 19.6/1000 patients aged 20-79/yr	[33] 60,052 cases among 3,892,749 consultations, weighted average of incidence of chest pain 1.54% [1.53-1.56]	
	Bösner et al [42]	D	2005-2006	All 190,000 consultations done by 74 participating GP's	Prevalence chest pain in patients aged ≥ 35 among all patients visiting GP: 0.7%		
	Verdon et al [41]	CH	2001	24,620 patients seen by 59 GPs	Prevalence new-onset chest pain among patients aged ≥ 16 among all patients visiting a GP: 1.8%		
	Svavarsdottir [35]	IC	1989-1990	5470 patients registered in one GP practice, making 28,050 consultations in 2 years	Percentage of GP consultations where chest pain is RFE: 0.68%		

	Frese [47]	D	1999-2000	8877 consultations in primary care	Percentage of GP consultations where chest pain is RfE: 3.0%		
	Frese [47]	NL	1985-2003	597,312 consultations in primary care	Percentage of GP consultations where chest pain is RfE: 1.7%		
	Rosser [33]	US/CAN	1985	registration of 71,525 consultations in primary care	Percentage of GP consultations where chest pain is RfE: 1.38%		
d) Number of patients with ACS in primary care	Verdon et al [41]	CH	2001	442 patients presenting new-onset chest pain	Incidence ACS among patients presenting with chest pain in primary care 1.5% (diagnosis set after 1 year follow up)	pooled data: 84 cases of ACS among 3,012 cases of chest pain	number of patients with ACS in primary care, NL 2013:
	Bösner et al [42]	D	2005-2006	1,212 patients ≥ 35, presenting with chest pain in primary care	Incidence ACS among patients presenting with chest pain in primary care: 3.6% (diagnosis set after 6 months follow up)	weighted average of incidence of ACS among cases of chest pain in primary care: 2.79% [2.24-3.46] (value D)	2.79% of 1,054,729 chest pain consultations (value C x value D) = 29,427 [23,473-36,731]
	Klinkman [34]	US	1994	399 episodes for chest pain in primary care	Incidence ACS among patients presenting with chest pain in primary care: 1.5%		(after exclusion of Rosser [33] (value C' x value D) 29,427 [23,473-36,968])
	Haasenritter [50]	D	2009-2010	844 patients with chest pain in primary care	Incidence ACS among patients presenting with chest pain in primary care: 2.5% (diagnosis set after 6 months follow up)		
	Andersson [46]	SW	2009-2011	115 patients with chest pain in primary care	Incidence ACS among patients presenting with chest pain in primary care: 5.2% (diagnosis set after 1 month follow up)		
	CMR Nijmegen [51]	NL & SW	1995-2008	All 35,000 patients presenting with chest complaints to a GP.	Relative number of patients that are immediately referred by the GP: 29%	pooled data: 15,266 referrals among 107,284 cases of chest pain	number of patients with chest pain in primary care directly referred to secondary care, NL 2013:
e) Absolute and relative number of referrals among all patients presenting chest pain to a GP.	Nilsson et al [37]	SW	1998-2000	Patients presenting with new-onset chest pain (exclusion of those suspected of stable coronary disease), N = 317	Relative number of patients that are immediately referred by the GP: 12%	weighted average of referral rate among cases of chest pain in primary care: 14.2% [14.0-14.4] (value E)	14.2% of 1,054,729 chest pain consultations (value C x value E) = 149,772 [146,703-152,867] (value E')

f) Referral rate among patients presenting with chest pain to a GP where who are suspected of ACS	Verdon et al [41]	CH	2001	442 patients presenting new-onset chest pain	Percentage of patients presenting with chest complaints to a GP, that are immediately referred by the GP: 16%. (Percentage suspected of severe acute disease: 20%)	(after exclusion of Rosser [33] 10,259 referrals among 35,759 cases, weighted average of referral rate among cases of chest pain in primary care: 28.7% [28.2-29.2]) (value E')	(after exclusion of Rosser [33] (value C' x value E') 302,707 [295,502-311,981]) (value E'')
	Rosser [33]	US/ CAN	1985	registration of 71,525 consultations in primary care	Percentage of patients presenting with chest complaints to a GP, that are immediately referred by the GP: 7%		
g) Absolute and relative number of patients presenting with chest pain in primary care where ACS / severe disease is considered	Bruins Slot et al [6]	NL	2006-2008	Selection of patients presenting with chest complaints to a GP <i>that are suspected for ACS by the GP</i> . N = 298.	Relative number of patients that are immediately referred by the GP: 73%	referral rate among cases of chest pain in primary care where ACS is suspected: 73.2% [67.7-78.0]	
	based on e) and f) in this table				In the total population presenting with chest pain to a GP, mean referral rate is 14.2% (see row e) in this table). Referral rate in the population of patients presenting chest pain to a GP who are suspected of ACS is 73% (see row f) in this table). Since the group of patients suspected of ACS in both populations is the same, the population of patients with chest pain suspected of ACS is assumed to be $14.2/73 = 19.5\%$ of the total population of patients with any chest pain, either suspected or unsuspected of ACS.	percentage of patients where ACS / severe disease is considered among patients presenting with chest pain in primary care: 20% (value G)	number of patients presenting with chest pain in primary care where ACS / severe disease is considered, NL 2013 (value C' x value G): 210,946 [209,576-212,316] (after exclusion of Rosser [33] (value C' x value G): 210,946 [209,576-213,685])
h) Absolute and relative number of ACS among referred patients	Verdon et al [41]	CH	2001	442 patients presenting new-onset chest pain	Percentage of patients with new presentation of chest pain suspected of severe acute disease: 20%		
	Nilsson et al [37]	SW	1998-2000	65 patients immediately referred by GP after presenting with new-onset chest pain	Percentage of ACS, resp AMI among patients with chest complaints that are referred by a GP, 29% resp 11% (outcome set 3 months after presentation)	pooled data: 67 cases of ACS among 283 referred patients weighted average of true referrals among referred patients with chest pain in primary care: 23.7% [18.9-29.1] (value H)	number of ACS among referred patients presenting with chest pain in primary care ('true referrals'), NL 2013 (value E'' x value H): 35,496 [27,727-44,484]
	Bruins Slot et al [6]	NL	2006-2008	Selection of study population that was referred by GP, N = 218	Percentage of ACS, resp AMI: 22% resp 17%,		(after exclusion of Rosser [33] (value E'' x value H) 71,742 [55,850 - 90,786])

i) Absolute and relative number of ACS among patients initially not referred	Nilsson et al [37]	SW	1998-2000	281 patient with new onset chest pain where ACS was not suspected by GP	Percentage of ACS among patients with chest complaints that were initially not suspected of ACS by a GP: 1.8% (outcome set 3 months after presentation)	ACS rate among non-referred patients with chest pain in primary care: 1.8% [0.7-4.3] (value I)	number of ACS among not referred patients presenting with chest pain in primary care ("false non referrals", NL 2013 (value I x (value C - value E"))): 16,289 [[6,279-39,257]] (after exclusion of Rosser [33] (value I x (value C' - value E"))): 13,536 [[5,151 - 33,236]]
j) Number of patients with chest pain at ER	Data from Academic Hospital Maastricht [25]	NL	2013	199,411 inhabitants of a region in the South-East of the Netherlands	Incidence of visiting cardiologic emergency department with chest pain (either referred by GP, brought by ambulance or self-referred): 12,54/1000 pts/year	Incidence of chest pain presentation at ER: 1,25 [1,21-1,30]	Number of patients with chest pain at ER, NL 2013: 209,745 [203,033- 218,134]

Supplementary Table S1. Table on literature search results. Data on numbers and referral rates of patients presenting with chest complaints in primary care are presented. Sources are presented and all numbers and rates are converted into absolute numbers in The Netherlands in the year 2013. These converted numbers are used in Figure 2, means are used in Figure 2 when more than 1 source is available. Abbreviations: 95%CI = 95% confidence interval; ACS = Acute coronary syndrome; AMI = acute myocardial infarction; B = Belgium; CAN = Canada; CH = Switzerland; CMR = continuous registration of morbidity; D = Germany; ER = emergency room; GP = general practitioner; IC = Iceland; MI = myocardial infarction; N = number; NL = the Netherlands; RFE = reason for encounter; SW = Sweden; UK = United Kingdom; US = United States of America

APPENDIX 7

Supplementary Table S2

	Source	Used in base case	Country ; year	Population, background	Outcome [estimation for variation, 95% CI when applicable]	Translation to (estimated) patient numbers, NL 2013 [estimation for variation, 95% CI when applicable]
a) Population in the Netherlands	CBS [27-30]	'international data' and 'NL and B'	NL; 2013	All inhabitants of the Netherlands	Total population: N = 16,779,575 Population aged 20-79: N = 12,205,982	N = 16,779,575 N = 12,205,982
b) Number of annual GP consultations	CBS [27-30]	'international data' and 'NL and B'	NL; 2013	All inhabitants of the Netherlands	Population visiting GP at least once: N = 12,015,601 Total number of GP consultations: N = 68,488,928	N = 12,015,601 N = 68,488,928
c) Incidence of chest pain as RFE GP	Dutch cohort study [22] Pooled data international studies [6, 33-47] Combination Dutch data and selected international data [6, 22, 34-47]	'NL and B' 'international data' 'combined data'	NL, B; 2015-2016 NL, UK, SW, D, Ch, Ic, US, Can; 1989-2008 NL, UK, SW, D, Ch, Ic, US, Can; 1989-2008	Registration of 22,294 consultations in primary care Pooled data: 61,039 cases of chest pain in 3,964,274 consultations Pooled data: 60,052 cases of chest pain in 3,892,749 consultations,	Percentage of GP consultations where chest pain is RFE: 1.26% Percentage of GP consultations where chest pain is RFE: 1.54% [1.53-1.55] Percentage of GP consultations where chest pain is RFE: 1.54% [1.53-1.56]	N = 862,960 N = 1,054,729 [1,047,881-1,061,578] N = 1,054,729 [1,047,881-1,068,427]
d) Number of patients with ACS in primary care	Dutch cohort study [22] Pooled data international studies [6, 33-47] Combination Dutch data and selected international data [6, 22, 34-47]	'NL and B' 'international data' 'combined data'	NL, B; 2015-2016 SW, D, Ch, US; 1994-2011 NL, UK, SW, D, Ch, Ic, US, Can; 1989-2008	281 patients with chest pain in primary care Pooled data: 84 cases of ACS among 3,012 cases of chest pain Pooled data: 84 cases of ACS among 3,012 cases of chest pain	Incidence ACS among patients presenting with chest pain in primary care: 6.45% Incidence of ACS among patients presenting with chest pain in primary care: 2.79% [2.24-3.46] Incidence of ACS among patients presenting with chest pain in primary care: 2.79% [2.24-3.46]	N = 55,654 N = 29,427 [23,473-36,731] N = 29,427 [23,473-36,731]

e) Absolute and relative number of referrals among all patients presenting chest pain to a GP.	Dutch cohort study [22]	'NL and B'	NL, B; 2015-2016	281 patients with chest pain in primary care	Relative number of patients that are immediately referred by the GP: 14.2%	N = 122,540
	Pooled data international studies [6, 33-47]	'international data'	NL, SW, CH, US, CAN; 1985-2008	Pooled data: 15,266 referrals among 107,284 cases of chest pain	Relative number of patients that are immediately referred by the GP: 14.2% [14.0-14.4]	N = 149,772 [146,703-152,867]
	Combination Dutch data and selected international data [6, 22, 34-47]	'combined data'	NL, UK, SW, D, CH, IC, US, Can; 1989-2008	Pooled data: 10,259 referrals among 35,759 cases of chest pain,	Relative number of patients that are immediately referred by the GP: 28.7% [28.2-29.2]	N = 302,707 [295,502-311,981]
g) Absolute and relative number of patients presenting with chest pain in primary care where ACS / severe disease is considered	Dutch cohort study [22]		NL, B; 2015-2016	281 patients with chest pain in primary care	Percentage of patients with new presentation of chest pain suspected of ACS: 38.1%	N = 328,788
	Pooled data international studies [6, 33-47]	'international data'	NL, SW, CH, US, CAN; 1985-2008		Percentage where ACS / severe disease is considered among patients presenting with chest pain, primary care: 20%	N = 210,946 [209,576-212,316]
	Combination Dutch data and selected international data [6, 22, 34-47]	'combined data'	NL, UK, SW, D, CH, IC, US, Can; 1989-2008		Percentage where ACS / severe disease is considered among patients presenting with chest pain, primary care: 20%	N = 210,946 [209,576-213,685]
h) Absolute and relative number of ACS among referred patients	Dutch cohort study [22]	'NL and B'	NL, B; 2015-2016	281 patients with chest pain in primary care	Percentage of ACS among patients with chest complaints that are referred by a GP ('true referrals'): 43% (outcome set at least 30 days after presentation)	True referrals' (ref+, ACS+): N = 52,692 False referrals' (ref+, ACS-): N = 69,848
	Pooled data international studies [6, 33-47]	'international data'	NL, SW; 1998-2008	Pooled data: 67 cases of ACS among 283 referred patients	Percentage of ACS among patients with chest complaints that are referred by a GP ('true referrals'): 23.7% [18.9-29.1]	True referrals' (ref+, ACS+): N = 35,496 [27,727-44,484] False referrals' (ref+, ACS-): N = 114,276 [105,288-122,045]
	Combination Dutch data and selected international data [6, 22, 34-47]	'combined data'	NL, UK, SW, D, CH, IC, US, Can; 1989-2008	Pooled data: 67 cases of ACS among 283 referred patients	Percentage of ACS among patients with chest complaints that are referred by a GP ('true referrals'): 23.7% [18.9-29.1]	True referrals' (ref+, ACS+): N = 71,742 [55,850 - 90,786] False referrals' (ref+, ACS-): N = 230,965 [211,921-246,857]

i) Absolute and relative number of ACS among patients initially not referred	Dutch cohort study [22]	'NL and B'	NL, B; 2015-2016	281 patients with chest pain in primary care	Percentage of ACS among patients with chest complaints that were initially not suspected of ACS by a GP ('false non-referrals'): 0.4% (outcome set at least 30 days after presentation)	'False non-referrals' (ref., ACS+): N = 2,962 'True non-referrals' (ref., ACS-): N = 737/458
	Pooled data international studies [6, 33-47]	'international data'	NL, SW; 1998-2008	Pooled data: 9 cases of ACS among 361 not referred patients	Percentage of ACS among patients with chest complaints that were initially not suspected of ACS by a GP ('false non-referrals'): 1.8% [0.7-4.3]	'False non-referrals' (ref., ACS+): N = 16,289 [6,279-39,257] 'True non-referrals' (ref., ACS-): N = 888,668 [865,700-898,678]
	Combination Dutch data and selected international data [6, 22, 34-47]	'combined data'	NL, B; 2015-2016	281 patients with chest pain in primary care	Percentage of ACS among patients with chest complaints that were initially not suspected of ACS by a GP ('false non-referrals'): 0.4% (outcome set at least 30 days after presentation)	'False non-referrals' (ref., ACS+): N = 3,008 [2,971 - 3,037] 'True non-referrals' (ref., ACS-): N = 749,014 [748,985-749,051]

Supplementary Table S2. Overview of three base cases based on various estimations of incidence of chest pain, referral rates and 'true and false (non-) referrals':

The first base case ('NL and B') is based on the results of a Dutch cohort study [23]. The second base case ('international data') is based on a pooled analysis of international studies on incidence of chest pain, incidence of ACS and referral rates in primary care. The third base case ('combined data') is based on a combination of base cases 'NL and B' and 'international data', using the international data set after exclusion of one reference because of a non comparable health care system in the study [50]. Moreover, in this third base case ('combined data'), the incidence of ACS positive non-referrals ('false non-referrals') was based on the Dutch and Belgian cohort study since it was considered more representative on this issue [23]. All three base cases were analysed separately in our model for cost efficiency (see remainder of article). All numbers and rates are converted into absolute numbers in the Netherlands in the year 2013. Abbreviations: 95%CI = 95% confidence interval; ACS = Acute coronary syndrome; B = Belgium; CAN = Canada; CH = Switzerland; CMR = continuous registration of morbidity; D = Germany; GP = general practitioner; Ic = Iceland; N = number; NL = the Netherlands; Ref = referral; RFE = reason for encounter; SW = Sweden; UK = United Kingdom; US = United States of America

3

Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners

Michelle M.A. Kip
Amber M. Noltes
Hendrik Koffijberg
Maarten J. IJzerman
Ron Kusters

This chapter has been published as:

M.M.A. Kip, A.M. Noltes, H. Koffijberg, M.J. IJzerman, R. Kusters. *Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners*. Prim Health Care Res Dev, 2017: p. 1-12

Reprinted with permission.

ABSTRACT

Aim

To investigate GPs' desire and perceived added value of point-of-care (POC) troponin, its effect on referral decisions, and test requirements.

Background

Excluding acute coronary syndrome (ACS) in primary care remains a diagnostic challenge for GPs. Consequently, referral rates of chest pain patients are high, while the incidence of a cardiovascular problem is only 8-15%. Previous studies have shown that GPs are interested in a POC troponin test. This test could enhance rapid exclusion of ACS, thereby preventing unnecessary patient distress, without compromising safety and while reducing costs. However, using this test is not recommended in current guidelines due to uncertainty in the test's potential added value, and the lower sensitivity early after symptom onset as compared with troponin tests in a regular laboratory.

Methods

An online survey containing 34 questions was distributed among 837 Dutch GPs in June 2015.

Findings

A total of 126 GPs (15.1%) completed at least 75% of the questions. 67.1% of GPs believe that POC troponin tests have moderate to very high added value. Although the availability of a POC test is expected to increase the frequency at which troponin tests are used, it likely decreases (immediate) referral rates. Of the responding GPs, 78.3% only accept 10 min as the maximum test duration, 78.1% think reimbursement of the POC device is required for implementation, and 68.9% consider it necessary that it can be performed with a finger prick blood sample. In conclusion, according to GPs, the POC troponin test can be of added value to exclude ACS early on. Actual test implementation will depend on test characteristics, including test duration, type of blood sample required, and reimbursement of the analyzer.

BACKGROUND

Coronary heart disease is one of the main causes of mortality. Worldwide, it accounted for about 7.4 million deaths in 2012 [1]. Acute coronary syndrome (ACS) is one of the main manifestations of coronary heart disease, with chest pain as a leading symptom. As ACS encompasses unstable angina pectoris and acute myocardial infarction (AMI), rapid diagnosis or exclusion of ACS is crucial to allow initiation of effective evidence-based treatment and management [2]. As a patient's signs and symptoms alone are often insufficient to rule out AMI in general practices [3], ~29% of all patients presenting with chest pain in the GP's office are immediately referred [4], while the estimated incidence of a cardiovascular problem among those patients is relatively low (8-15%) [5-9]. This suggests that between six out of seven and 11 out of 12 patients are referred to the hospital without having ACS. Those unnecessary hospital referrals will incur patient distress as well as high healthcare costs. In general, the aging population and the increasing number of patients with multiple (chronic) conditions, have led to an increase in healthcare usage, and consequently to rising healthcare costs [10]. Therefore, GPs are increasingly pressured to play a gatekeeping role in the healthcare system to reduce referrals from primary to secondary care. Because of the high referral rate among chest pain patients, there is large potential for improvement regarding the exclusion of ACS in primary care.

To guide their referral decision, GPs can order a troponin test at a regular laboratory to detect myocardial injury or necrosis [2, 11]. Currently available high-sensitive troponin tests have increased the diagnostic accuracy for AMI as compared with conventional assays, especially in patients presenting early after chest pain onset [2]. Although point-of-care (POC) tests for troponin have become available, the majority of those tests are less sensitive than the currently available laboratory troponin tests. Consequently, POC troponin tests may not be able to detect a small rise in a patient's troponin level very early after the onset of an AMI, indicating that those tests cannot be used for diagnosing AMI with certainty in the early hours after symptom onset [2, 12]. Nevertheless, previous research has shown that a POC troponin test might be of added value in excluding ACS in low-risk patients presenting at the GP with a longer duration of symptoms or whose symptoms have already resolved [13, 14].

Recently, an international survey study by Howick et al. (2014) found that 66% of GPs in five countries, including the Netherlands, the United States, the United Kingdom, Australia, and Belgium, would like to have access to a POC troponin test [15]. However, there may be a mismatch between anticipated benefits (improving GPs' ability to rule out ACS) and actual test performance. Furthermore, current Dutch guidelines discourage the use of a POC test to rule out ACS in general practices, while international guidelines do not give a recommendation on whether or not to use the POC troponin test in those patients [2, 16]. Despite the issues

with the limited diagnostic performance of this POC test, more sensitive POC troponin tests are currently being developed [17]. As those developments may allow safe use of this test in the nearby future, they might enhance GPs' ability to safely exclude and even diagnose ACS in primary care. However, it is likely that the decision whether or not to implement this test will not solely depend on its diagnostic performance. Therefore, early insights in additional barriers and facilitators regarding implementation of this test, the expected use of this test in clinical practice, as well as the impact on referral decisions, may help both test developers and clinicians in providing safe and efficient healthcare. In turn, this likely facilitates rapid and efficient implementation as soon as those more sensitive tests become available. Therefore, this study aims to investigate a broad range of GPs' preferences and requirements regarding POC troponin testing for patients presenting with chest pain in primary care, as well as its effect on GPs' referral decisions.

METHODS

A survey was constructed in LimeSurvey [18] to identify GPs' opinions and requirements concerning the use of POC troponin testing in chest pain patients, and the indicated impact on GPs' referral decisions. Qualitative interviews with ten GPs served as input for the design of this survey, by providing insight into the GP's diagnostic process and decision making in patients presenting during consultation hours with chest pain in Dutch primary care. Following this, the survey was tested in a pilot study with four GPs to evaluate feasibility and completeness. The final survey (in Dutch) consisted of 34 questions, divided in five elements.

The first element included seven questions concerning general information about the GPs (eg, age, gender, years of working experience, type of general practice). The second element (seven questions) was about current practice, including the distance to the nearest hospital and coronary intervention center, estimated referral rates of new hypothetical chest pain patients to the hospital, the role of GPs' gut feeling in this decision, the percentage of patients presenting with symptoms for ≥ 6 h, and about the different POC tests that are currently available in their general practice. The third element contained seven questions about: the expected added value of the POC troponin test, the frequency at which they estimate to currently order laboratory troponin tests in chest pain patients, the frequency at which they expect to perform POC troponin tests for this purpose, and the aspects favorably affected by implementation of this test. In this question, GPs were asked to choose all applicable answers from a predefined list, based on a study by Cals et al. (2014) which was slightly adapted to the use of a POC troponin test (the aspect 'diagnostic speed' was added, while aspects concerning medication use and therapy compliance were removed) [19]. The fourth element (10 questions) concerned the effect of a POC troponin test on GPs' referral rates using two hypothetical patient cases, which were designed to represent one low and one intermediate risk case of ACS (based on known risk factors but without actual prediction of their ACS risk).

For both cases, GPs were given the patient's signs, symptoms, medical history, medication use, and vital functions. For each case, GPs were asked if they would immediately refer this patient to the hospital, refer the patient for an appointment at an outpatient cardiology clinic, not refer the patient (indicating that no action is taken), or decide otherwise (open answer). GPs who decided to refer immediately were asked to additionally consider a negative POC troponin test result, which could have been available for this patient during consultation. They were asked whether this information would make them revise their decision. In contrast, GPs who decided not to refer immediately, were presented with a slightly elevated troponin level and posed the same question. GPs who still decided not to refer immediately were presented with a strongly elevated troponin level and again posed the same question. As cut-off levels for troponin may differ between laboratories, misinterpretation was prevented by mentioning whether the troponin level was either negative, slightly elevated, or strongly elevated. The fifth element contained three questions concerning the maximum acceptable test duration, necessity of reimbursement of the analyzer, and the requirement of performing this test using a finger prick blood sample. The full survey (translated to English) is provided in Appendix 1.

The survey was distributed among 837 Dutch GPs affiliated to one of seven participating GP associations, located in four provinces of the Netherlands (Overijssel, Gelderland, Utrecht and Noord-Brabant). All GPs were invited by email (by AN, or by the GP association to which they were affiliated) to fill out the survey from Tuesday June 9, 2015 to Wednesday June 24, 2015. After 1 week, a reminder was sent out to all non-responders. All responses were processed anonymously.

The representativeness of the responding GPs of the entire population of GPs in the Netherlands was evaluated by comparing the respondents' age and gender to data of GP registrations from the Netherlands Institute for Health Services Research (Nederlands instituut onderzoek van de gezondheidszorg, NIVEL), using the one sample t-test and the nonparametric binomial test ($\alpha = 0.05$). To investigate whether the added value of the POC troponin test is considered higher when the distance between the nearest hospital or coronary intervention center increased, this relationship was examined using the χ^2 test. A χ^2 test was also performed to investigate whether the indicated importance of GP's gut feeling in making a referral decision, affected the extent to which they changed their initial referral decision following a discordant POC troponin test result. Data were analyzed using R (version 3.2.3). The package mice (version 2.25) was used for multiple imputation of missing data (10 imputation sets) [20, 21]. Multiple imputation is known to yield more valid results than complete case analysis [22]. To assess the effect of missing values on our results, estimates from complete case analysis and multiple imputation were compared.

RESULTS

Characteristics of responding GPs

In total, 169 GPs (20.2%) participated in the survey. Results of 43 respondents were necessarily excluded because they were either not registered as a GP ($n = 6$) or because they only started the survey but completed less than 25% of the questions ($n = 37$). Of the remaining 126 GPs (15.1%), 10 GPs missed one question (7.9%), and one GP missed two questions (0.8%). They needed ~15 min to fill out the survey. Given the limited amount of missing values, results from multiple imputation were very similar to the results from complete case analysis (Table 1). Therefore, only results from multiple imputation are reported. Of the responding GPs, 80 (63.5%) were male (Table 1), the mean age was 48.4 years ($SD = 9.4$), the average working experience as GP was 16.2 years ($SD = 9.7$), and 107 GPs (84.9%) had their own general practice. There was no significant difference in mean age between the responding GPs and the NIVEL database ($n = 11,345$) (48.4 versus 48.9 years, $P = 0.57$) [23]. However, a larger proportion of the responding GPs were male, although not statistically significant (63.5 versus 55.2%, $P = 0.07$). The distance to the nearest hospital was <20 km for 95.1% of GPs, whereas the distance to the nearest coronary intervention center was <20 km for 64.4% (Table 1).

Current work-up

GPs indicated that 54.9% ($SE = 2.0\%$) of all patients presenting with chest pain report symptoms for ≥ 6 h. Of all patients presenting with chest pain, GPs estimate to refer 26.8% ($SE = 2.0\%$) immediately to the hospital (Figure 1). Of those immediate referrals, GPs had strong suspicion in 54.7% ($SE = 2.2\%$), were uncertain about the presence of ACS in 33.0% ($SE = 1.6\%$), while 12.2% ($SE = 1.8\%$) were referred to reassure the patient (no suspicion of ACS) (Figure 1). A total of 97.6% of GPs indicate that their gut feeling has moderate, high or very high impact on this decision (Table 1).

GPs' desire to use the POC troponin test and its perceived added value

In all, 67.1% of GPs believe that POC troponin tests have moderate, high or very high added value for diagnosing patients with chest pain in general practices (Table 2). Only five GPs (4.0%) already had a POC troponin analyzer in their general practice (for details about the availability of POC analyzers in general practices see Appendix 2). Three of those GPs consider this test to have high ($n = 1$) or very high added value ($n = 2$) for diagnosing ACS in general practices, while only one considers this test of barely added value. Furthermore, the perceived added value of this POC test (none, barely or moderate added value, as compared to high or very high added value) was significantly higher for GPs whose practices are situated ≥ 10 km from the nearest hospital ($P = 0.01$), as compared with GPs whose practices are situated closer to the hospital. However, this difference was not significant for the distance from the nearest coronary intervention center ($P = 0.16$).

Table 1. Characteristics of participating GPs, and the role of GPs' gut feeling in their referral decision.

	Complete cases	Multiple imputation
Participating GPs [n (%)]	115 (100.0)	126 (100.0)
Male [n (%)]	74 (64.3)	80 (63.5)
Age in years [mean (SD)]	49.0 (9.3)	48.4 (9.4)
Years of working experience as a GP [mean (SD)]	16.9 (9.7)	16.2 (9.7)
Independent GP (own practice) [n (%)]	98 (85.2)	107 (84.9)
Distance to nearest hospital (%) (km)		
<2	18.3	16.7
2-5	21.7	23.0
5-10	27.0	25.6
10-20	27.8	29.8
20-40	5.2	4.9
>40	0	0
Distance to nearest coronary intervention center (%) (km)		
<2	9.6	8.7
2-5	14.8	15.1
5-10	16.5	15.1
10-20	26.1	25.5
20-40	28.7	31.3
>40	4.3	4.3
Role of GPs' gut feeling in immediate referral decision (%)		
None	0.0	0.0
Barely	2.6	2.4
To a moderate extent	20.9	23.0
To a high extent	64.3	62.7
To a very high extent	12.2	11.9

This table shows the characteristics of the participating GPs, the distance from the general practice to the nearest hospital and to the nearest coronary intervention center, as well as the role of GPs' gut feeling in the decision to immediately refer a patient to the hospital.

Figure 1. Estimated referral rates. This figure shows the estimated referral rates of patients presenting with chest pain in primary care [26.8% (SE 2.0%)] of all patients presenting with chest pain in primary care who are immediately referred to the hospital.

Table 2. Expectations of GPs with regard to the added value of using point-of-care (POC) troponin testing. This table shows the expected added value, as indicated by GPs ($n = 126$), of the POC troponin test in diagnosing acute coronary syndrome in primary care, as well as the aspects that they consider to be favorably affected by this test.

Expected added value (%)	
None	6.3
Barely	26.5
To a moderate extent	28.3
To a high extent	31.7
To a very high extent	7.1
Aspects favorably affected (%)	
Diagnostic certainty	67.5
Diagnostic speed	55.6
Patient safety	42.1
Healthcare costs	41.3
Patient satisfaction	31.7
Substitution to primary care	30.2
Doctor-patient relationship	21.4
Work satisfaction	19.0
Doctor-patient communication	14.3
Other	0.8
None of the above	12.7

In general, GPs believe that benefits related to the use of POC troponin tests include: diagnostic certainty (67.5%), time until diagnosis (55.6%), patient safety (42.1%), and reduced healthcare costs (41.3%) (Table 2). Purposes for which GPs would use a POC troponin test (top four choices) are: to exclude ACS (76.2%), for reassuring patients (50.0%), to improve their consultation with the cardiologist (43.7%), and to confirm ACS (40.5%) (Appendix 3).

Figure 2 (a, b). (Expected) frequency of performing [point-of-care (POC)] troponin tests. This figure shows the (expected) frequency of performing troponin tests in a regular laboratory, and POC troponin tests in primary care, depending on the perceived suspicion of acute coronary syndrome (ACS).

To investigate whether the availability of a POC troponin analyzer will increase the frequency at which troponin tests are performed, Figure 2a shows the estimated frequency at which GPs request laboratory troponin tests at a regular laboratory, while Figure 2b shows the expected frequency of performing a POC troponin test assuming GPs would have access to a POC troponin analyzer in their general practice. A total of 46.3% of GPs estimate to never perform a laboratory troponin test in patients without suspicion of ACS. This probability is expected to decrease to 28.6% in case a POC troponin analyzer is available. Aggregating over the estimated frequencies of troponin use in Figure 2a and 2b, the probability that a troponin test is performed, in patients in whom the GP is uncertain about the presence of ACS, is 55.1% for POC tests, and 30.3% for laboratory tests. In patients strongly suspected of ACS, those probabilities are 26.3 and 4.9%, respectively.

Effect of POC troponin tests on GPs' referral decisions

For the patient case which was designed to represent a patient with a low risk of ACS, the probability of ACS as estimated by the GPs was 11.2% (CI = 9.0-13.5%). For the intermediate risk case this probability was estimated at 31.5% (CI = 27.7-35.3%). In all, 80.2% of GPs estimated a higher probability of ACS in the intermediate risk case as compared with the low risk case. Figure 3 shows to what extent a discordant POC troponin test result is expected to make GPs revise their initial referral decision in both patient cases. In other words, this figure shows to what extent GPs indicated to change their initial referral decision in two hypothetical situations: (1) the GP initially decided to immediately refer the patient to the hospital, but the POC troponin level is then found to be negative, and (2) the GP initially decided not to refer the patient, but the troponin level is then found to be slightly, or even strongly elevated.

For the low and intermediate risk case, respectively 2.4 and 26.2% of GPs stated they would immediately refer the patient. Confronting those GPs with a non-elevated troponin level would make 66.7 and 81.8% to revise their decision in the low and intermediate risk case, respectively. In contrast, in each case where the GP indicates not to refer immediately, presenting those GPs with a slightly elevated troponin level resulted in an immediately referral by 52.1 and 55.9%. Of the remaining GPs who still indicate not to immediately refer, presenting them a strongly elevated troponin level resulted in an immediate referral by 70.7 and 58.5% in the low and intermediate risk case, respectively. Furthermore, GPs who decided not to immediately refer indicated that the availability of an elevated POC troponin test result would likely change that decision: there was a strong decrease in the number of GPs who indicated not to take further action or to perform additional examinations (eg, an electrocardiogram or additional laboratory tests), while there was a strong increase in the frequency at which they indicate to consult a cardiologist regarding this patient case. Furthermore, a χ^2 test revealed that GPs' gut feeling (none, barely, or relatively much impact, as compared with high or very high impact) did not affect the probability of changing their immediate referral decision following a discordant POC troponin result (low-risk case: $P = 0.56$, intermediate risk case: $P = 0.86$).

Figure 3. Impact of point-of-care (POC) troponin testing on GPs' referral decisions. The estimated probability of acute coronary syndrome (ACS) in the low and intermediate risk case was 11.2 and 31.5%, respectively. To illustrate the flow of GPs through this figure, consider the example of 97.6% of all GPs who decide not to refer the low risk patient case based on a patient's signs, symptoms, medical history, medication use, and vital functions. The table that is presented describes the action taken (in %) of those GPs who decided not to refer this patient. Next, all GPs who decided not to immediately refer were asked to additionally consider a slightly elevated POC troponin level, which could have been available for this patient during consultation. Based on this additional information, they were again asked if they would refer this patient immediately. In the low risk case, 47.9% of those GPs still decide not to refer this patient. When those remaining GPs were asked to consider a strongly elevated troponin level in this patient, which could have been available during consultation, 70.7% of those GPs indicate to immediately refer this low risk patient, while 29.3% still indicates not to refer this patient. ACS = acute coronary syndrome, prob. = probability.

Requirements for an unobtrusive use of POC troponin tests

In all, 78.3% of GPs consider 10 min the maximum acceptable duration of the POC troponin test, and only 4.8% find 20 min or longer acceptable. Also, 78.1% state that reimbursement of the POC device is necessary, while 68.9% demands the test can be performed with a finger prick instead of a venous blood sample.

DISCUSSION AND CONCLUSION

GPs have a gatekeeping role in the health care system. Using diagnostic tests in primary care may improve GPs' ability to decide which patients to refer to secondary care, thereby improving quality of care provided, and preventing unnecessary costs and burden to patients. In particular, the availability of POC troponin tests may enhance GPs' ability to rule out ACS. The current study provides new insights into the potential impact of this test, as well as in factors that facilitate and impede implementation. Instead of focusing on test developers or policy makers in healthcare, our study explored the utility of troponin testing, and assessed how their referral behavior may change when POC troponin tests are implemented. Such information is required to understand and improve quality of care in primary care.

Impact on referral decisions

GPs have indicated that POC troponin results will change some of their referral decisions, indicating the tests' diagnostic utility. However, as the same two cases were presented to all GPs, the actual change may, of course, be different when other cases would be used. Also, GPs were provided with the POC troponin test result after making their initial referral decision, and this stepwise provision of information may have overemphasized the test result, and may thereby have artificially increased the impact of this test. However, as current POC troponin tests take ≥ 10 min, it is likely that GPs already made an initial referral decision before having the test result available. As literature indicates that clinicians tend to cling to their initial diagnosis even when contradictory evidence becomes available [24], it is unlikely that our approach led to a substantial overestimation of the potential effect of POC troponin testing in primary care.

Although GPs prefer a very sensitive POC test allowing rapid and accurate detection of ACS, the high-sensitive troponin tests currently used in laboratories increase the frequency at which slightly elevated troponin levels are encountered in patients who do not have an AMI [11]. As our study indicates that a slightly elevated troponin may increase immediate referrals and consultations with cardiologists, this high-sensitivity may also increase healthcare costs.

As, according to the responding GPs, 54.9% of patients have ≥ 6 h complaints, the issue of the relatively low sensitivity of POC troponin very early after symptom onset will not apply to those patients. In addition, although this test is unlikely to be performed in patients with a high suspicion of ACS, GPs estimate that they only have a high suspicion in about half of the

26.8% of immediately referred patients. Thus, the POC troponin test might be of added value in a relatively large proportion of patients. In addition, as the referral decision after the POC troponin test was found to be independent from the estimated effect of gut feeling by GPs, POC troponin tests may improve GPs' ability to rule out ACS and thereby prevent referrals. Although this will decrease healthcare costs, the availability of a POC test might increase the frequency at which a patient's troponin level is measured, which is most likely explained by the shorter turn-around-time as compared with the laboratory test.

Strengths and limitations

A limitation of the current approach involves the risk of recall bias, as GPs are asked to make estimations concerning referral rates of chest pain patients and about the frequency with which they request laboratory troponin tests. Depending on the number of patients presenting with chest pain in the GP's office each month, which likely ranges from 2 to 4, a recall period of 3 months would mean that GP's estimates are based on 6-12 patients. Given the relatively large number of respondents in the current study, and the absence of reasons why GPs would systematically underestimate or overestimate these numbers, it is likely that our overall results are realistic and robust. Although it might have been possible to additionally use primary care databases to gather more detailed information on current management of chest pain patients, such databases cannot provide insights in the potential use of POC troponin tests. Finally, more accurate data might also be obtained by performing a pragmatic trial. However, such a trial will be costly and time-consuming, and the chosen approach already allows deriving valuable estimates of intended POC troponin at low cost and in a short time period.

Although the number of respondents was relatively large (126 responses that could be included in the final analysis), the low response rate of 15.1% represents a limitation to our study. However, this issue of low response rates among GPs has been reported previously [25]. To investigate the impact of this low response rate, results from our survey have been compared with the results of another recent Dutch survey study concerning POC tests in primary care [19]. Results indicate that the availability of POC tests in Dutch general practices was very similar (Appendix 2). Compared to this previous study, only the availability of POC C-reactive protein (CRP) tests and glycated hemoglobin (HbA_{1c}) tests increased strongly, which can be explained by the increase in scientific evidence that has become available and the inclusion of the CRP POC test in the guideline 'Acute Cough' of the Dutch College of General Practitioners [26-29]. In addition, we found that 67.1% of GPs perceive moderate to very high added value of POC troponin tests, whereas 65% of the GPs previously expressed their desire to use this test [19]. These similar results were found despite the possible overrepresentation of GPs located in more rural areas in the Netherlands. Also, although the distance between general practices and hospitals, even in rural areas, is relatively low in the Netherlands as compared with other countries, the relation found between the perceived added value of the POC troponin test and the distance of the general practice from a hospital does corroborate previous findings in

other countries [15]. Consequently, the overall perceived added value of the POC troponin test by GPs across the Netherlands might be somewhat lower than reported here.

Previous research has shown that GPs in different countries unanimously expressed the desire to have this test available in their office [15]. In addition, as issues of the aging population and the increased demand of healthcare usage are applicable to all (developed) countries, the results from this study performed in the Netherlands are expected to be generalizable to other countries.

Recommendations

As most currently available POC troponin tests do not yet meet GPs' requirements [30], designing a test that is fast and only requires a finger prick will be paramount to the implementation of POC troponin testing in Dutch primary care. Besides, reimbursement is considered crucial. Therefore, a few recommendations to enhance (efficient) implementation and use of POC troponin testing can be given. First, as the current study investigated the impact of only a few POC troponin levels on GPs' referral decisions, it is recommended to investigate this impact for multiple different troponin levels. Second, the two cases used in this study both concerned patients aged below 65 years. However, as previous research states that the use of different decision limits for high-sensitive troponin should be considered in elderly patients [31], GPs' referral decisions may also be dependent upon the patient's age. Third, the prognostic value of slightly elevated troponin levels from increasingly sensitive tests to guide therapy may be investigated [32]. Fourth, as research has shown that POC troponin tests may often be performed without prior ischemia assessment by the clinician [33], the introduction of (adapted) decision rules in primary care, like the MACS or HEART score developed for emergency care, might improve the correct triaging of chest pain patients and identify those that benefit most from additional POC troponin testing [34, 35].

Although the consideration to use POC troponin has to be made by each GP individually, this study gives insights in the barriers and facilitators regarding this implementation decision, and shows the expected impact when this test would be available in clinical practice. The findings presented in this study are thus not only relevant for GPs, but also for policy makers, as well as for developers of POC testing technology. In addition, previous research has indicated that a POC troponin test may be cost saving (by reducing emergency hospital referrals), although this may come at the expense of missed cases of ACS [36, 37]. However, despite those valuable insights, those studies were limited by a non-randomized design and a small number of AMI patients. Further research is therefore recommended to synthesize evidence on the frequency of POC troponin testing, the impact on (correct and incorrect) referral decisions, test costs, and the benefits from rapid and adequate treatment in a formal cost-effectiveness analysis.

In conclusion, although the findings of the current study indicate that the POC troponin test may be of added value for excluding ACS in patients considered at low or intermediate risk, as well as to reassure patients, improvements in its diagnostic performance are still needed to ensure their implementation. Despite the fact that most of the currently available POC troponin tests do not yet meet the GPs' requirements, ongoing developments are likely to increase the test's expected benefits. The insight obtained from this study can be used to guide further development, as well as to facilitate implementation of the POC troponin test in clinical practice.

ACKNOWLEDGEMENTS

The authors thank all GPs for their participation in this questionnaire.

FINANCIAL SUPPORT

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

CONFLICTS OF INTEREST

None.

REFERENCES

- World Health Organization, World Health Statistics: monitoring health for the SDGs, sustainable development goals. 2016: Geneva.
- Authors/Task Force Members, et al., 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur Heart J*, 2015.
- Bruyninckx, R., et al., Signs and symptoms in diagnosing acute myocardial infarction and acute coronary syndrome: a diagnostic meta-analysis. *Br J Gen Pract*, 2008. 58(547): p. 105-11.
- Okkes, I.M., et al., Continue morbiditeitsregistratie Nijmegen. Episodes of care in family practice. Epidemiological data based on the routine use of the International Classification of Primary Care (ICPC) in the Transition project of the Academic Medical Center/University of Amsterdam (1985-2003). 2005, Academic Medical Center/University of Amsterdam, Department of Family Medicine: Amsterdam.
- Bosner, S., et al., Chest pain in primary care: epidemiology and pre-work-up probabilities. *Eur J Gen Pract*, 2009. 15(3): p. 141-6.
- Nilsson, S., et al., Chest pain and ischaemic heart disease in primary care. *Br J Gen Pract*, 2003. 53(490): p. 378-82.
- Ruigomez, A., et al., Chest pain in general practice: incidence, comorbidity and mortality. *Fam Pract*, 2006. 23(2): p. 167-74.
- Rutten, F.H., et al., NHG-standaard acuut coronair syndroom (eerste herziening). *Huisarts Wet*, 2012. 55(12): p. 564-70.
- Buntinx, F., et al., Chest pain in general practice or in the hospital emergency department: is it the same? *Fam Pract*, 2001. 18(6): p. 586-9.
- Wiener, J.M. and J. Tilly, Population ageing in the United States of America: implications for public programmes. *Int J Epidemiol*, 2002. 31(4): p. 776-81.
- Thygesen, K., et al., Third universal definition of myocardial infarction. *Circulation*, 2012. 126(16): p. 2020-35.
- Bruins Slot, M.H., et al., Point-of-care tests in suspected acute myocardial infarction: a systematic review. *Int J Cardiol*, 2013. 168(6): p. 5355-62.
- Marshall, G.A., N.G. Wijeratne, and D. Thomas, Should general practitioners order troponin tests? *Med J Aust*, 2014. 201(3): p. 155-7.
- Tomonaga, Y., et al., Diagnostic accuracy of point-of-care testing for acute coronary syndromes, heart failure and thromboembolic events in primary care: a cluster-randomised controlled trial. *BMC Fam Pract*, 2011. 12: p. 12.
- Howick, J., et al., Current and future use of point-of-care tests in primary care: an international survey in Australia, Belgium, The Netherlands, the UK and the USA. *BMJ Open*, 2014. 4(8): p. e005611.
- Dutch College of General Practitioners. NHG-Standaard Acuut coronair syndroom (eerste herziening). 2012; Available from: <https://www.nhg.org/standaarden/volledig/nhg-standaard-acuut-coronair-syndroom-eerste-herziening#note-13>.
- Philips In-vitro Diagnostics. Minicare I-20, Enabling near patient blood testing in the acute care setting. 2016 [cited 2016 April 21]; Available from: <http://www.philips.nl/healthcare/product/HCNOCTN496/minicare-i20-enabling-near-patient-blood-testing-in-the-acute-care-setting>.
- LimeSurvey Project Team / Schmitz, C., LimeSurvey: An Open Source survey tool. 2015, LimeSurvey Project: Hamburg, Germany.
- Cals, J.W.L., et al., Sneltesten in de huisartsenpraktijk: huidig gebruik en behoefte aan testen in de toekomst. *Ned Tijdschr Geneeskde*, 2014. 158(A8210): p. 7.
- R Core Team, R: A language and environment for statistical computing. 2015, R Foundation for Statistical Computing: Vienna, Austria.
- van Buuren, S. and K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 2011. 45(3): p. 1-67.
- van der Heijden, G.J., et al., Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*, 2006. 59(10): p. 1102-9.
- van Hassel, D.T.P., A. Kasteleijn, and R.J. Kenens, Cijfers uit de registratie van huisartsen - peiling 2014. 2015, NIVEL: Utrecht. p. 1-32.
- Rajkomar, A. and G. Dhaliwal, Improving diagnostic reasoning to improve patient safety. *Perm J*, 2011. 15(3): p. 68-73.
- Parkinson, A., et al., Recruiting general practitioners for surveys: reflections on the difficulties and some lessons learned. *Aust J Prim Health*, 2015. 21(2): p. 254-8.
- Cals, J.W., et al., C-reactive protein point of care testing and physician communication skills training for lower respiratory tract infections in general practice: economic evaluation of a cluster randomized trial. *J Eval Clin Pract*, 2011. 17(6): p. 1059-69.
- Cals, J.W., et al., Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. *BMJ*, 2009. 338: p. b1374.

28. Lenters-Westra, E. and R.J. Slingerland, Three of 7 hemoglobin A1c point-of-care instruments do not meet generally accepted analytical performance criteria. *Clin Chem*, 2014. 60(8): p. 1062-72.
29. Verheij, T.J.M., et al., NHG-Standaard Acuut hoesten (Eerste herziening). *Huisarts Wet*, 2011. 54(2): p. 68-92.
30. Bingisser, R., et al., Cardiac troponin: a critical review of the case for point-of-care testing in the ED. *Am J Emerg Med*, 2012. 30(8): p. 1639-49.
31. Andersson, P.O., et al., Consequences of high-sensitivity troponin T testing applied in a primary care population with chest pain compared with a commercially available point-of-care troponin T analysis: an observational prospective study. *BMC Res Notes*, 2015. 8: p. 210.
32. Sherwood, M.W. and L. Kristin Newby, High-sensitivity troponin assays: evidence, indications, and reasonable use. *J Am Heart Assoc*, 2014. 3(1): p. e000403.
33. Ahmad, F.A., S. Dobbin, and A.D. Hargreaves, Implementation of point-of-care troponin T testing in clinical practice. *Br J Cardiol*, 2015. 22(79).
34. Body, R., et al., The Manchester Acute Coronary Syndromes (MACS) decision rule for suspected cardiac chest pain: derivation and external validation. *Heart*, 2014. 100(18): p. 1462-8.
35. Backus, B.E., et al., A prospective validation of the HEART score for chest pain patients at the emergency department. *Int J Cardiol*, 2013. 168(3): p. 2153-8.
36. Nilsson, S., et al., Point-of-Care Troponin T Testing in the Management of Patients with Chest Pain in the Swedish Primary Care. *Int J Family Med*, 2013. 2013: p. 532093.
37. Nilsson, S., et al., Cost consequences of point-of-care troponin T testing in a Swedish primary health care setting. *Scand J Prim Health Care*, 2014. 32(4): p. 241-7.

APPENDIX 1

3

Point-of-care troponin testing: improvement of the diagnostic process of patients presenting with chest pain?

Dear general practitioner,

Thank you in advance for participating in this questionnaire. Your answers can contribute to **improving the diagnostic process of chest pain patients in the general practitioner's office**. The questionnaire concerns the current diagnostic process and the referral decisions for patients with chest pain in the general practitioner's office and the potential role of **point-of-care troponin tests** in this process. This questionnaire also contains some questions about your preferences and requirements with regard to point-of-care troponin tests.

It will take approximately **10 minutes** to fill out the survey. Of course the answers you provide will be processed **anonymously**. This study is conducted by the University of Twente in collaboration with several general practitioner organizations in the Netherlands. The results are owned by the University of Twente and will only be used for **scientific purposes**. In case you have any questions and/or remarks, you can always contact us via telephone or email. Thank you again for your time and effort.

With kind regards,

A.M. (Maran) Noltes

Master student Health Sciences, University of Twente

Email:

Tel.:

1. Are you currently registered as a general practitioner?

☐ Yes

☐ No (*in case participants are not registered as a general practitioner, they are not allowed to fill out the survey, and therefore the survey ends*)

Diagnosing patients with (a)specific chest pain

This part of the survey contains questions about the referral decisions for patients with (a)specific chest pain and the diagnostic process of those patients in the general practitioner's office.

2. How many of the 100 patients who present with (a)specific chest pain in the general practitioner's office, do you on average immediately refer to the hospital?*

Please fill out your answer over here:

Per 100 patients, I refer out of 100 patients immediately to the hospital.

* by ambulance or by means of own transport

3. What percentage of patients that you immediately refer, do you refer because of a strong suspicion of ACS, uncertainty about the presence of ACS, or when you have no suspicion of ACS (for reassurance)?

Please fill out your answer here:

% *strong suspicion of ACS*

% *uncertainty about ACS*

% *no suspicion of ACS (for reassurance)*

ACS = acute coronary syndrome

4. What percentage of patients with (a)specific chest pain, who present in your general practice, have had complaints for ≥ 6 hours?

You may estimate the answer.

Please fill out your answer here:

About %

5. How often do you measure a patients troponin level in a medical laboratory in patients with (a)specific chest pain, in whom you after your anamnesis and physical examination, either have a strong suspicion of ACS, uncertainty about ACS, or no suspicion of ACS?

Choose the appropriate answer for each subquestion:

	<i>Never (0/10)</i>	<i>Occasionally (2/10)</i>	<i>Regularly (4/10)</i>	<i>Often (6/10)</i>	<i>Very often (8/10)</i>	<i>Always (10/10)</i>
<i>Strong suspicion of ACS</i>	☐	☐	☐	☐	☐	☐
<i>Uncertainty about ACS</i>	☐	☐	☐	☐	☐	☐
<i>No suspicion of ACS</i>	☐	☐	☐	☐	☐	☐

6. Do you or your medical assistant in the general practitioner's office and/or the general practitioners cooperation (occasionally) make an ECG in patients with (a)specific chest pain?

Please choose one of the following answer options:

- ☐ *No (if the GP chooses this answer, the questionnaire will continue with question 8)*
- ☐ *Only in the general practice*
- ☐ *Only in the general practitioners cooperation*
- ☐ *Both in the general practice and in the general practitioners cooperation*

ECG = electrocardiogram

7. For what reason(s) do you or your medical assistant perform an ECG in a patient with (a)specific chest pain?

Please select all possibilities:

- ☐ *To exclude ACS*
- ☐ *To confirm ACS*
- ☐ *Performing an ECG is a standard procedure in patients with chest pain*
- ☐ *To reassure the patient (and family)*
- ☐ *Because I sometimes have time to do this because I have to wait for an ambulance*
- ☐ *Because I want to maintain my expertise in this field*
- ☐ *To be able to have an informed consultation with the cardiologist on call*
- ☐ *Other, namely*

8. To what extent does your gut feeling play a role in deciding whether or not to immediately refer a patient with (a)specific chest pain?

	<i>None</i>	<i>Barely</i>	<i>To a moderate extent</i>	<i>To a high extent</i>	<i>To a very high extent</i>
<i>Gut feeling</i>	☐	☐	☐	☐	☐

Cases

In the following part, we will present you two cases of patients presenting with (a) specific chest pain in the general practice. After each case, we will ask you some questions about your treatment policy for this specific patient. Please treat those cases in the same way as you would treat real patients in your general practice. There are no right or wrong answers.

Case Mr. De Vries

Mr. de Vries (39 years) presents with chest pain on Monday morning during your consultation hour, at 8 am.

In the last half year, you have seen Mr. De Vries five times because of hay fever, allergic rhinitis, lower back pain, and migraine. Mr. de Vries does not have any other illnesses.

Medication list:

- Levocetirizine 1x/day 5 mg
- Fluticasone 50 mg/dose
- Paracetamol 2x/day 500 mg

The symptoms of Mr. De Vries began three days ago during cycle racing and have not fully disappeared since then. Because he remained somewhat worried about the cause of his symptoms, he decided to come to your consultation hour this morning. He experiences chest pain, and assigns it a score of 2/10. The pain is spread all over his chest. He is not able to clearly indicate the type of pain he feels. The pain does not radiate. When Mr. De Vries bends over, his pain somewhat relieves. During palpation of the chest wall, Mr. De Vries slightly panics and the pain worsens. There are no cardiovascular diseases in the patient's personal history or in his family history. It is known that Mr. De Vries does not smoke, has a healthy diet, and is physically active (cycle racing).

RR 130/80 mm Hg. Pulse: 82 beats/minute (regular). Saturation: 98%. Auscultation heart: no abnormalities (tones, souffles). Auscultation lungs: clean.

If the GP in question 6 answers to (occasionally) use an ECG, the additional information "ECG does not display any typical abnormalities" is provided. Otherwise "no ECG has been recorded" is shown.

9. Please estimate the probability that Mr. De Vries is having an ACS.

Please fill in your answer here:

%

10. What will be your treatment policy for Mr. De Vries?

Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

11. In case the GP decides to immediately refer the patient to the hospital, he/she is shown the following question, after which the questionnaire continues with question 14. In case the GP decides not to immediately refer the patient, the questionnaire continues with question 12.

There is a point-of-care troponin test device available in your general practice. The troponin concentration of Mr. De Vries is 17 ng/L (reference value < 40 ng/L). What would be your treatment policy for Mr. De Vries in case you would have received this additional information immediately? Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

12. In case the GP decides not to immediately refer the patient to the hospital, he/she is shown the following question:

There is a point-of-care troponin test device available in your general practice. The troponin concentration of Mr. De Vries is 63 ng/L (reference value < 40 ng/L). What would be your treatment policy for Mr. De Vries in case you would have received this additional information immediately? Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

In case the GP decides to immediately refer the patient, the questionnaire continues with question 14. In case the patient is not immediately referred, it continues with question 13.

13. In case the GP decides in question 12 not to immediately refer this patient to the hospital, he/she is shown the following question:

There is a point-of-care troponin test device available in your general practice. The troponin concentration of Mr. De Vries is 200 ng/L (reference value < 40 ng/L). What would be your treatment policy for Mr. De Vries in case you would have received this additional information immediately? Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

Case Mrs. Alsahi

Mrs. Alsahi (50 years) presents with chest pain on Thursday morning during your consultation hour, at 11 am, accompanied by her son.

In the past three years, you have seen Mrs. Alsahi three times because of diabetes type II, acute cough, and for a cholesterol measurement.

Mrs. Alsahi does currently not use any medication.

The symptoms of Mrs. Alsahi began two days ago when she was at work (cleaner). She experiences chest pain and pain in her upper abdomen, but she cannot indicate exactly where the pain is located. She assigns the pain a score of 3/10. The pain does not radiate to her limbs, neck or jaw, etc. She is somewhat short of breath (dyspnea) and is also very tired. Mrs. Alsahi does not look pale and does not sweat. She thinks she feels that the pain relieves when she changes her body position, but she is not able to clearly indicate this. Three months ago her cholesterol level was measured in your general practice. The ratio total cholesterol/HDL was 5.5 (total chol = 6mmol/L, HDL = 1.1 mmol/L). By that time it was decided to provide some lifestyle advices and not to start with medication. There are no cardiovascular diseases in her family history. Also, it is known that she does not smoke and does not drink alcohol. She is slightly overweight and has an average diet. Besides her work as a cleaner, she does not engage in any physical activities. Mrs. Alsahi is relatively calm, but her son insisted on consulting you because of the nature of her symptoms. He is worried about the cause.

RR 140/80 mm Hg. Pulse: 76 beats/minute (regular). Saturation: 98%. Auscultation heart: no abnormalities (tones, souffles). Auscultation lungs: clean.

If the GP in question 6 answers to (occasionally) use an ECG, the additional information "ECG does not display any typical abnormalities" is provided. Otherwise "no ECG has been recorded" is shown.

14. Please estimate the probability that Mrs. Alsahi is having an ACS.

Your answer should be between 0 and 100.

Please fill in your answer here:

 %

15. What will be your treatment policy for Mrs. Alsahi?

Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

16. *In case the GP decides to immediately refer the patient to the hospital, he/she is shown the following question. After this, the questionnaire continues with question 19. In case the GP decides not to immediately refer the patient, the questionnaire continues with question 15.*

There is a point-of-care troponin test device available in your general practice. The troponin concentration of Mrs. Alsahi is 5 ng/L (reference value < 40 ng/L).

What would be your treatment policy for Mrs. Alsahi in case you would have received this additional information immediately?

Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

17. *In case the GP decides not to immediately refer the patient to the hospital, he/she is shown the following question:*

There is a point-of-care troponin test device available in your general practice. The troponin concentration of Mrs. Alsahi is 58 ng/L (reference value < 40 ng/L).

What would be your treatment policy for Mrs. Alsahi in case you would have received this additional information immediately?

Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

In case the GP decides to immediately refer the patient, the questionnaire continues with question 19. In case the patient is not immediately referred, it continues with question 18.

18. *In case the GP decides in question 12 still not to immediately refer the patient to the hospital, he/she is shown the following question:*

There is a point-of-care troponin test device available in your general practice. The troponin concentration of Mrs. Alsahi is 196 ng/L (reference value < 40 ng/L).

What would be your treatment policy for Mrs. Alsahi in case you would have received this additional information immediately?

Please choose one of the following options:

- ☐ No referral.
- ☐ A short term appointment at the outpatient cardiology clinic.
- ☐ Immediate referral because of suspicion of ACS
- ☐ Other, namely

End of the cases. You have now completed 75% of the questionnaire. You may now continue with the penultimate part of the questionnaire, concerning your preferences and requirements with regard to point-of-care troponin tests.

Point-of-care troponin tests

In this part you will be asked some questions about your personal preferences and requirements with regard to point-of-care troponin tests. No right or wrong answers can be given.

19. Which of the following point-of-care tests* are available in your general practice(s) where you work, for instance in the form of test strips?

Select all answers that apply:

- ☐ CRP (*C-reactive protein*)
- ☐ Glucose
- ☐ Urine (*for example: nitrate, leukocytes*)
- ☐ Pregnancy test
- ☐ Hemoglobin
- ☐ HbA1c
- ☐ Cholesterol
- ☐ Troponin
- ☐ D-dimer
- ☐ BNP (*brain natriuretic peptide*)
- ☐ Chlamydia
- ☐ PT-INR
- ☐ Other, namely:

* Point-of-care tests (bedside tests) involve devices or test strips that can be performed at the bedside of patients or in the general practice to rapidly set a diagnosis.

20. To what extent do you think that the point-of-care troponin test has added value for diagnosing patients with (a) specific chest pain in the general practice (office hours) and at the general practitioner cooperation (outside office hours)?

Please choose the most appropriate answer:

	<i>Not at all</i>	<i>Barely</i>	<i>To a moderate extent</i>	<i>To a high extent</i>	<i>To a very high extent</i>
<i>GP office</i>	☐	☐	☐	☐	☐
<i>GPC</i>	☐	☐	☐	☐	☐

21. How often would you perform a point-of-care troponin test in patients with (a) specific chest pain in which, following anamnesis and physical examination, you either have a strong suspicion of ACS, uncertainty about the presence of ACS, and have no suspicion of ACS?

Please choose the most appropriate answer:

	<i>Never (0/10)</i>	<i>Occasionally (2/10)</i>	<i>Regularly (4/10)</i>	<i>Often (6/10)</i>	<i>Very often (8/10)</i>	<i>Always (10/10)</i>
<i>Strong suspicion of ACS</i>	☐	☐	☐	☐	☐	☐
<i>Uncertainty about ACS</i>	☐	☐	☐	☐	☐	☐
<i>No suspicion of ACS</i>	☐	☐	☐	☐	☐	☐

22. For which purposes would you like to use a point-of-care troponin test?

Select all answers that apply:

- ☐ *None*
- ☐ *To exclude ACS*
- ☐ *To confirm ACS*
- ☐ *To reassure a patient*
- ☐ *To be able to provide better service to patient*
- ☐ *To be able to have an informed consultation with the cardiologist on call*
- ☐ *Other, namely*

23. Which of the aspects mentioned below would, in your opinion, be positively affected by the implementation of point-of-care troponin tests in your general practice?

Select all answers that apply:

- ☐ *Diagnostic certainty*
- ☐ *Diagnostic speed*
- ☐ *Patient safety*
- ☐ *Patient satisfaction*
- ☐ *Doctor-patient relationship*
- ☐ *Communication with patients*
- ☐ *Job satisfaction*
- ☐ *Healthcare costs*
- ☐ *Substitution to primary care*
- ☐ *None of the above*
- ☐ *Other, namely*

24. Please indicate which factors play the most important role for your decision to implement a point-of-care troponin test in your general practice, and rank them from 1 to 9.

Please swipe all factors from the left to the right column. Place the factor that you consider the most important on top, and place the factor you consider the least important at the bottom.

- *High diagnostic reliability*
- *High user-friendliness*
- *Perform test with finger prick instead of venous blood withdrawal*
- *Quick availability of test result*
- *Uptake of this test in the Dutch general practitioner guidelines*
- *Small and easy transportable device*
- *Requires little training and maintenance*
- *Integrating it with the general practice information system and other point-of-care devices*
- *Low costs*

25. What do you consider to be the maximum acceptable time between conducting a point-of-care troponin test in general practice and the availability of the test result?

Please choose one of the following answers:

- ☐ *1 minute*
- ☐ *5 minutes*
- ☐ *10 minutes*
- ☐ *15 minutes*
- ☐ *20 minutes*
- ☐ *30 minutes*
- ☐ *60 minutes*
- ☐ *90 minutes*
- ☐ *120 minutes*

26. Is performing the point-of-care troponin test using a finger prick blood sample (instead of a venous blood withdrawal) a requirement for implementing this test in your general practice?

- ☐ *Yes*
- ☐ *No*

27. Is reimbursement of this point-of-care troponin device by the health insurer or by a medical laboratory (instead of having to pay it as a general practice) for you a requirement for implementing this test in your general practice?

- ☐ *Yes*
☐ *No*

Personal information

In this last part of the questionnaire you are asked to provide some personal information, for instance age and gender. Of course, those results will be processed anonymously.

You have now completed 90% of the questionnaire. We will ask for some personal information after which you have completed the survey.

What is your gender?

- ☐ *Female*
☐ *Male*

28. What is your age?

Your answer must be between 18 and 100.

Please fill in your answer here:

years

29. For how many years have you already worked as a general practitioner (excluding your time in training)? Please fill in the answer in years.

Your answer must be between 0 and 100.

Please fill in your answer here:

years

30. In what way are you currently working as a general practitioner?

Select all answers that apply:

- ☐ *Own general practice*
☐ *Employed by another general practitioner*
☐ *Locum general practitioner*

31. In what type of general practice are you currently employed?

Select all answers that apply:

- ☐ None
- ☐ In a solo practice
- ☐ In a duo practice
- ☐ In a group practice
- ☐ In a "HOED" (multiple general practitioners located in one building, mostly with some form of collaboration)

32. To what organization(s) are you connected and in which town is your general practice located?

Select all answers that apply:

- ☐ Twentse Huisartsen Onderneming Oost Nederland
- ☐ Federate Eerstelijnszorg Almelo (FEA)
- ☐ Huisartsenzorg Oost-Achterhoek
- ☐ Huisartsendienst Arnhem
- ☐ Huisartsendienst Gelderse Vallei
- ☐ Chronos Zorggroep
- ☐ Doktersdienst Groningen (DDG)
- ☐ Other, namely...

* In case you are employed at multiple general practices, please fill out this question for the general practice where you are employed most of your time.

This is the last question of this questionnaire.

33. How far from the nearest hospital and from the nearest coronary intervention center is your general practice located*? Note: this could be the same hospital.

Choose the most appropriate answer for both subquestions:

	< 2 km	2 to 5 km	5 to 10 km	10 to 20 km	20 to 40 km	> 40 km
Distance to nearest hospital	☐	☐	☐	☐	☐	☐
Distance to nearest coronary intervention center	☐	☐	☐	☐	☐	☐

* In case you are employed by multiple general practices, please fill out this question for the general practice where you are employed most of your time.

End of the questionnaire

You have reached the end of this questionnaire. Thank you very much for your participation!

In case you have any questions and/or remarks about the questionnaire or about this study, please write them down in the field below.

In case you would like to receive the results of this study, please leave your email address in the field below.

Thank you very much for your participation. It is highly appreciated if you forward this questionnaire to other GPs. You can do this by forwarding the email you received, or by copying and pasting the link below in your own email.

(Link to the questionnaire)

With kind regards,

A.M. (Maran) Noltes

Master student Health Sciences, University of Twente

E-mail: ...

Tel: ...

APPENDIX 2

Point-of-care devices in general practices in the Netherlands

Table 1. Respondents' answer to the question "Which of the following point-of-care devices are available in the general practice(s) you work at?" (n = 126)

Test	n (%)
Glucose	122 (96.8)
Urine test strip	119 (94.4)
Pregnancy test	108 (85.7)
CRP (C-reactive protein)	101 (80.2)
Hemoglobin	67 (53.2)
HbA _{1c}	24 (19.0)
D-dimer	14 (11.1)
Troponin	5 (4.0)
Other	2 (1.6)
BNP (brain natriuretic peptide)	2 (1.6)
Cholesterol	0 (0.0)
PT-INR	0 (0.0)
Chlamydia	0 (0.0)

3

APPENDIX 3

Purposes for which GPs expect to use a POC troponin test

Figure 1. Purposes for which GPs expect to use a POC troponin test. GPs were asked to choose all applicable answers.

4

The cost-utility of point-of-care troponin testing to diagnose acute coronary syndrome in primary care

Michelle M.A. Kip
Hendrik Koffijberg
Marco J. Moesker
Maarten J. IJzerman
Ron Kusters

This chapter has been published as:

M.M.A. Kip, H.Koffijberg, M.J. Moesker, M.J. IJzerman, R. Kusters. *The cost-utility of point-of-care troponin testing to diagnose acute coronary syndrome in primary care.* BMC Cardio Dis, 2017. 17(1):p. 213.

Reprinted with permission.

ABSTRACT

Background

The added value of using a point-of-care (POC) troponin test in primary care to rule out acute coronary syndrome (ACS) is debated because test sensitivity is inadequate early after symptom onset. This study investigates the potential cost-utility of diagnosing ACS by a general practitioner (GP) when a POC troponin test is available versus GP assessment only.

Methods

A patient-level simulation model was developed, representing a hypothetical cohort of the Dutch population (>35 years) consulting the GP with chest complaints. All health related consequences as well as cost consequences were included. Both symptom duration, selection of patients in whom the POC troponin test is performed, and test performance at different time points were incorporated. Health outcomes were expressed as Quality-Adjusted Life Years (QALYs). The main outcome parameters involve the effect of POC troponin testing on (in) correct hospital referrals, QALYs, and costs.

Results

The POC troponin strategy decreases the referral rate in non-ACS patients from 38.46% to 31.85%. Despite a small increase in non-referral among ACS patients from 0.22% to 0.27%, the overall health effect is negligible. Costs will decrease with €77.25/patient (95%CI €-126.81 to €-33.37).

Conclusions

The POC troponin strategy is likely cost-saving, by reducing hospital referrals. The small increase in missed ACS patients can be partly explained by conservative assumptions used in the analysis. Besides, current developments in POC troponin tests will likely further improve their diagnostic performance. Therefore, future prospective studies are warranted to investigate whether those developments make the POC troponin test to a safe and cost-effective diagnostic tool for diagnosing ACS in general practices.

BACKGROUND

Each year almost 200,000 patients in the Netherlands contact their general practitioner (GP) with chest pain [1]. Chest pain is a leading symptom of acute coronary syndrome (ACS), which encompasses both unstable angina pectoris (UAP) and acute myocardial infarction (AMI) [2]. Among AMI patients, two subgroups can be distinguished based on a deviation on the electrocardiogram (ECG), which are ST elevation myocardial infarction (STEMI) and non-ST elevation myocardial infarction (NSTEMI) [2]. However, only approximately 18% of all patients who present with chest pain to the hospital are eventually diagnosed with ACS [3]. Clinical management is a major challenge for GPs, because unnecessary referrals lead to unnecessary costs and patient distress. So, given the potentially life-threatening consequences, accurate and rapid diagnosis of ACS remains crucial.

Besides a patient's signs, symptoms, medical history, and the information obtained from the ECG, imaging modalities (for example coronary angiography), and laboratory biomarkers of myocardial necrosis (mainly involving high-sensitive troponin), provide additional diagnostic value [2]. As those diagnostic tools are often not readily available in primary care, current clinical guidelines recommend immediate hospital referral when ACS is suspected [4]. However, the introduction of point-of-care (POC) troponin tests, accompanied by a shorter turn-around-time, offers a low-cost diagnostic test for GPs. Although costs per tests are mostly higher for POC tests as compared to regular laboratory tests [5], the use of POC troponin as a diagnostic aid to rule out ACS might improve referral decisions and hence lead to overall cost savings [6]. This advantage needs however to be weighed against the lower sensitivity of this test especially in the early hours after symptom onset [2, 6-8]. Consequently, international guidelines do not give a recommendation on the use of POC troponin [2], while its use is discouraged in Dutch guidelines [9]. In addition to collecting clinical evidence, proper evaluation of the cost-utility of POC troponin testing in primary care is required. Although randomized controlled trials (RCTs) would provide the best source of evidence, these are commonly very expensive, time consuming, and often do not allow assessment of long-term outcomes or risks [10, 11]. Therefore, an early stage health economic model may provide a valid alternative to quantify the effect of POC troponin testing on costs, and on short and long term outcomes. In this paper, we present such model-based health economic analysis of the use of POC troponin testing to diagnose ACS in primary care, as compared to current practice.

METHODS

Structure of the model

We developed a patient-level simulation model to reflect the diagnostic pathway of chest pain patients presenting to the GP (Figure 1). This model was used to analyse the cost-utility of a POC troponin test to exclude ACS in addition to GP examination, as compared to GP examination without the test. As test and treatment decisions in this model are based (partially) on patient's

characteristics and patient's history (including duration of symptoms, previous test outcomes, and previous management decisions) adequately reflecting the diagnostic pathway of patients requires simulating individual patients. Whereas such decisions are simulated for entire groups in a cohort (state-transition) model, they can be simulated on individual patient level when using a patient-level simulation model.

Figure 1. Structure of the decision tree. Black triangles represent branches in which patients are referred to the ED (because of suspected ACS), while grey triangles represent branches in which patients are sent home. ACS = acute coronary syndrome, ECG = electrocardiogram, GP = general practitioner, hosp. = hospital, POC = point-of-care, prob. = probability, STEMI = ST elevation myocardial infarction.

The primary outcome measure was the incremental costs per quality-adjusted life year (QALY) gained for the POC strategy as compared to current practice. Secondary outcome measures were the percentage of chest pain patients that were either correctly or incorrectly referred to the hospital or sent home (expressed as true positives [TPs], true negatives [TNs], false positives [FPs], and false negatives [FNs]). Other secondary outcome measures included mortality and heart failure rates in both strategies. All costs and effects were evaluated from a societal point of view, over a lifelong time horizon.

Current evidence

The model was populated using data from literature. The following paragraphs present a summary of evidence used, for further details see Appendix 1. An overview of the assumptions used is provided in Appendix 2.

Patient pathways

In this patient-level model, each hypothetical patient was assigned an age, gender, and disease state (i.e. ACS or no ACS) based on the population distribution derived from a cross-sectional diagnostic study [12, 13]. For ACS patients in primary care, the probabilities of having either UAP, NSTEMI or STEMI were based on a prospective study [14].

Of all chest pain patients presenting to the GP, 14.7% was immediately referred to the hospital without diagnostic testing, because of a high perceived probability of ACS by the GP (Table 1) [15]. All patients that were referred to the hospital were assumed to be transported by ambulance. Among the patients that were not immediately referred, all GPs who had an ECG in their practice (45%) [16], were assumed to use it in those patients. In case the patient had a STEMI, there was a 50% probability that this was diagnosed in case an ECG was performed [17]. Those patients were also immediately referred to the hospital. In current practice, the GP decides on referral of the remaining patients without POC troponin. The sensitivity and specificity of this decision (i.e. 88.3% and 72.2%, respectively) were derived from Nilsson et al. [18].

For the POC strategy, the time from symptom onset to presentation was estimated for each patient separately [19], and then categorized according to time since symptom onset as a) <4.0 h, b) 4.0 – 5.5 h, c) 5.5 – 7.0 h, d) 7.0 – 10.0 h, and e) >10.0 h. As it is known that the performance of the POC troponin test is less in patients presenting shortly after symptom onset [7], it was assumed that this test was not performed in patients presenting with <4.0 h symptoms [7, 19]. In those patients, the referral decision was based on the judgment of the GP as in current practice [18]. In the remaining patients, it was assumed that the POC troponin test was performed, and the sensitivity and specificity for time categories b), c), d) and e) were derived from Diercks et al. [7]. In case the result of the POC troponin test did not correspond with what would have been the (initial) referral decision of the GP in current practice (i.e. without POC

troponin), a probability of the GP revising the initial referral decision was incorporated in the model [15]. In both strategies, all patients who were not immediately referred to the hospital were assumed to be sent home without receiving medical care.

Mortality and heart failure rates were dependent on whether patients had ACS, and if present, whether ACS was correctly diagnosed. In addition, as the POC troponin test was assumed to delay time to treatment with 10 min, using this test would increase the mortality rate in AMI patients with 0.2% [21]. This 10 min duration of the POC troponin test was chosen as previous research has shown that GPs consider this the maximum acceptable duration of the POC troponin test [15]. Productivity losses incorporated lost working time due to GP visits, emergency department visits, and due to UAP, STEMI, NSTEMI, heart failure, and mortality [22-24].

Quality of life

The quality-adjusted life expectancy (QALE, based on combining expected quality of life with life expectancy) were obtained directly from a systematic review and economic model for STEMI, NSTEMI and UAP patients as well as for healthy patients [25-27]. Separate QALE estimates were used for categories of age and gender [25-27]. A missed diagnosis of ACS indirectly lowered quality of life by increasing the risk of heart failure and mortality. For heart failure patients, the QALEs were estimated by multiplying the reported utility of heart failure patients with their reported average life expectancy [28, 29].

Costs

Costs of both a GP consultation and ambulance transportation to the emergency department were derived from the Dutch cost manual [24], whereas costs of performing an ECG were derived from the Dutch Healthcare Authority [30]. As no tariff currently exists for the POC troponin test, and because a troponin test performed in a regular laboratory costs approximately €8 [30], the costs of a POC troponin test were conservatively estimated at €15. Average diagnosis and treatment costs per (suspected) ACS patient were derived from the Dutch Healthcare Authority [3]. Healthcare costs for the remaining lifetime of STEMI, NSTEMI, and UAP patients were derived from Goodacre et al. [25-27]. Lifetime healthcare costs per patient with heart failure were based on Dutch data concerning the point prevalence of heart failure [31], the annual healthcare costs of those patients [32], and combined with the average remaining life expectancy [28]. All costs were converted to 2015 using the consumer price indices as reported by the Dutch Central Bureau of Statistics, and converted to Euros if necessary (exchange rate 1.271 GBP = €1.000, 31 March 2016) [33, 34]. All costs and effects were discounted at an annual rate of 4.0% and 1.5%, respectively, in accordance with Dutch costing guidelines [24]. An overview of all cost parameters, including costs of productivity losses, is shown in Appendix 1.

Table 1: input parameters concerning patient pathways used in the model. This table shows the parameter, the value used in the model, the 95% confidence interval, the distribution used, and the data source. ACS = acute coronary syndrome, CI = confidence interval, ECG = electrocardiogram, GP = general practitioner, POC = point-of-care, STEMI = ST elevation myocardial infarction.

Parameter	Category	Probability	95% CI	Distribution	Reference
Perceived probability of ACS by GP	High probability	14.7%	8.2% to 22.5%	Beta	[15]
ECG	Performed at GP's office	57.4%	56.0% to 58.7%	Beta	[16]
	Sensitivity for STEMI	50.0%	37.6% to 62.6%	Beta	[17]
Time from symptom onset to presentation (male)	<4 h	47.5%	38.9% to 55.4%	Dirichlet	[19, 20]
	4–5.5 h	9.0%	5.2% to 14.5%	Dirichlet	[19, 20]
	5.5–7 h	7.3%	3.4% to 11.8%	Dirichlet	[19, 20]
	7–10 h	9.4%	5.2% to 14.9%	Dirichlet	[19, 20]
	>10 h	26.8%	20.0% to 34.6%	Dirichlet	[19, 20]
Time from symptom onset to presentation (female)	<4 h	54.8%	47.1% to 62.7%	Dirichlet	[19, 20]
	4–5.5 h	9.4%	5.1% to 13.9%	Dirichlet	[19, 20]
	5.5–7 h	6.5%	3.1% to 10.7%	Dirichlet	[19, 20]
	7–10 h	8.0%	4.6% to 13.1%	Dirichlet	[19, 20]
	>10 h	21.2%	15.2% to 27.8%	Dirichlet	[19, 20]
Diagnostic performance of GP, without POC troponin	Sensitivity	88.3%	79.2% to 95.2%	Beta	[18]
	Specificity	72.2%	67.9% to 76.3%	Beta	[18]
Sensitivity of POC troponin at different duration of symptoms	4–5.5 h	66.7%	55.5% to 77.0%	Beta	[7]
	5.5–7 h	79.2%	69.3% to 87.5%	Beta	[7]
	7–10 h	84.7%	75.3% to 91.9%	Beta	[7]
	>10 h	87.5%	78.8% to 94.2%	Beta	[7]
Specificity of POC troponin at different duration of symptoms	4–5.5 h	95.9%	94.2% to 97.3%	Beta	[7]
	5.5–7 h	94.4%	92.5% to 96.1%	Beta	[7]
	7–10 h	93.4%	91.3% to 95.3%	Beta	[7]
	>10 h	92.6%	90.4% to 94.5%	Beta	[7]
Revise referral decision after discordant POC troponin	Negative troponin	80.6%	73.3% to 87.1%	Beta	[15]
	Positive troponin	83.3%	76.3% to 89.3%	Beta	[15]

Probabilistic sensitivity analysis

A probabilistic sensitivity analysis was performed by means of Monte Carlo simulations, using 10,000 iterations of 20,000 hypothetical, unique patients. Distributions were assigned to all parameters under investigation [35]. An overview of the distributions and the 95%CI for each parameter is provided in Table 1, and Appendix 1.

Expected value of perfect information (EVPI) analysis

To determine the value of collecting additional information on this topic, and potentially enhance decision-making on the usefulness of POC troponin in primary care a value of information analysis was performed. The EVPI was calculated to estimate the expected costs of uncertainty in the current model, using a value of information tool developed by the University of Sheffield [36].

Cost-effectiveness acceptability curve (CEAC)

To evaluate the probability that the POC troponin strategy is cost-effective, a CEAC was constructed for a willingness-to-pay (WTP) threshold ranging from €0/QALY to €100,000/QALY. A separate CEAC was constructed under the assumption that use of the POC troponin test is only acceptable when it does not lead to health loss (i.e. does not increase the rate of missed ACS patients) and has acceptable cost-effectiveness.

RESULTS

Probabilistic sensitivity analysis

The results indicate that the POC troponin strategy costs on average €1,144 (95% CI €892 to €1,451) per patient, as compared to €1,221 (95% CI €955 to €1,541) for current practice. The cost savings for the testing strategy are €77 (95% CI €-127 to €-33) per patient. The probability that patients were correctly not referred to the hospital (TNs) would increase from 57.90% to 64.51% (+6.61%). Consequently, unnecessary referrals (FPs) decreased with 6.61% (i.e. from 38.46% to 31.85%). However, the probability of patients being correctly referred to the hospital (TPs) showed a very small decrease from 3.42% to 3.37% (i.e. -0.05%), which in turn resulted in an increase of patients incorrectly not referred (FNs) from 0.22% to 0.27% (i.e. 54 vs. 45 per 20,000 simulated patients). In turn, this results in a very small increase in the probability of patients dying of ACS (i.e. 61 vs. 60 per 20,000 patients), as well as a small increase in the incidence of heart failure (i.e. 121 vs. 119 per 20,000 patients). The overall change in QALYs from POC troponin use is negligible (-0.0004 when rounded to 4 decimals, i.e. 3.5 h in full health).

Given the negligible impact on health outcomes, it is not useful to calculate an incremental cost utility ratio (ICUR), as this outcome is very unstable for tiny differences in health outcomes as estimated in this study. When combining differences in health outcomes with differences in costs, an incremental net monetary benefit of €69 was found. In the POC strategy, the POC

Table 2. Outcomes of Monte Carlo simulations. This table shows the results of 10,000 iterations of 20,000 patients, showing the costs and remaining quality-adjusted life expectancy per patient. Both the 95% CI and the absolute and relative effect are provided. CI = confidence interval, FN = false negative, FP = false positive, HF = heart failure, POC = point-of-care, TN = true negative, TP = true positive.

Parameter	Without POC troponin (95% CI)	With POC troponin (95% CI)	Absolute effect (POC vs. non-POC) (95% CI)	Relative effect (POC vs. non-POC) (95% CI)
Primary care costs per patient	€291 (€189 to €421)	€262 (€174 to €375)	€-28 (€-50 to €-12)	-9.78% (-13.36% to -5.75%)
Hospital costs per patient	€446 (€308 to €620)	€398 (€277 to €552)	€-48 (€-77 to €-26)	-10.68% (-14.27% to -7.12%)
Lifetime costs per patient	€378 (€243 to €554)	€384 (€246 to €566)	€7 (€-17 to €33)	1.74% (-4.49% to 8.86%)
Costs of productivity loss per patient	€107 (€65 to €169)	€99 (€60 to €160)	€-8 (€-19 to €3)	-7.30% (-16.94% to 2.70%)
Total costs per patient	€1,221 (€955 to €1,541)	€1,144 (€892 to €1,451)	€-77 (€-127 to €-33)	-6.33% (-9.80% to -2.85%)
Quality adjusted life expectancy	11.29 (9.41 to 13.33)	11.29 (9.41 to 13.33)	-0.00 (-0.00 to 0.00)	-0.00% (-0.02% to 0.01%)
Probability TP	3.42% (2.47% to 4.53%)	3.37% (2.44% to 4.47%)	-0.05% (-0.15% to 0.06%)	-1.37% (-4.08% to 1.64%)
Probability FP	38.46% (33.97% to 45.00%)	31.85% (28.29% to 38.70%)	-6.61% (-8.30% to -4.99%)	-17.19% (-20.86% to -12.66%)
Probability TN	57.90% (51.52% to 62.38%)	64.51% (57.82% to 68.14%)	6.61% (4.99% to 8.30%)	11.42% (8.34% to 15.06%)
Probability FN	0.22% (0.08% to 0.46%)	0.27% (0.14% to 0.46%)	0.05% (-0.06% to 0.15%)	20.91% (-15.72% to 135.02%)
Probability death	0.30% (0.19% to 0.45%)	0.30% (0.19% to 0.46%)	0.00% (-0.01% to 0.03%)	1.53% (-4.11% to 9.30%)
Probability HF	0.59% (0.40% to 0.83%)	0.60% (0.41% to 0.84%)	0.01% (-0.03% to 0.05%)	1.85% (-4.24% to 8.91%)

troponin test was used in approximately 40% of all chest pain patients presenting to their GP. This strategy could result in annual cost savings in the Netherlands of approximately €14 million [1]. A detailed overview of all model outcomes, including the 95% CI, is shown in Table 2, and the incremental cost-effectiveness plane is shown in Figure. 2. In Table 2, the costs are divided in four categories: primary care costs involves the costs from presentation at the GP including potential hospital referral by ambulance. Hospital costs involves the costs from hospital admission until discharge. Lifetime costs involves the costs of follow-up after STEMI. Also, the effect on (correct) referral decisions, as well as the effect on the probability of heart failure and mortality is shown.

More specifically, the average cost savings of €77.25 per patient are mainly attributable to a decreased referral rate, resulting in a reduction in costs in primary care with €28.41 per patient (as caused by a decrease in hospital referrals by ambulance), accompanied by a decrease in costs of hospital treatment of €47.62 per patient. Although the lifetime costs per patient are slightly higher (i.e. +€6.58), which is mainly attributable to a small increase in heart failure rates, those higher costs are compensated by a decrease in productivity loss per patient (-€7.80), caused by a decrease in hospital referral rates.

Figure 2. Incremental cost-effectiveness plane, showing the result of 10,000 model iterations of 20,000 patients, the mean value, and the ICUR threshold of €20,000/QALY. ICUR = incremental cost-utility ratio, POC = point-of-care, QALY = quality-adjusted life year.

EVPI results

Results indicate that the overall EVPI per person presenting with chest pain in primary care is estimated to be €0.22. Based on a decision relevance of ten years, and an estimated 182,897

individuals presenting with chest pain in this period, this results in an overall EVPI per year of €403,364 in the Netherlands [1]. Thus, giving this low EVPI, it is unlikely that performing more research, to decrease uncertainty, would be considered an efficient use of resources. An overview of the expected value of partial perfect information for groups of model input parameters is shown in Appendix 3.

CEAC results

The CEAC indicates that the probability of the POC troponin strategy to be cost-effective ranges from 100.0% to 52.7%, for a WTP threshold of €0 to €200,000/QALY. For a WTP of €20,000/QALY, this probability was 98.4% (Figure 3). Thus, when society is unwilling to pay any money for an increase in health outcomes (i.e. a WTP of €0/QALY), the POC troponin strategy has to be cheaper compared to current practice to be considered cost-effective, regardless of the effect on health outcomes. On the other hand, a very high WTP indicates that society is unwilling to accept any loss in QALYs. Therefore, an increase in WTP decreases the probability that the POC troponin strategy is cost-effective. However, when the POC test was also required to provide equal or better health outcomes compared with usual care (non-inferiority), the probability of the POC troponin strategy to be cost-effective remained stable at 28.2% for a WTP ranging from €0 to €200,000/QALY (Figure 3). Although this suggests that the POC troponin strategy has a large chance of resulting in health loss, it should be noted that the actual absolute health loss is likely to be vanishingly small. For example, previous research has indicated that a change in health utility of 0.03 is the minimum clinically important difference, and that smaller differences can be considered to be not clinically important [37]. In our analysis the difference in utility is much smaller than 0.03, as even over the lifetime horizon the accumulated corresponding health loss equals only -0.0004. Indeed, when a very small decrease of 0.002 QALYs is chosen as cut-off for the maximum acceptable health loss, the probability that the POC troponin strategy is cost-effective already remains very high, decreasing only from 100.0% to 97.3% for a WTP ranging from €0 to €200,000/QALY (Figure 3).

DISCUSSION

Using a POC troponin test in primary care is expected to increase the probability at which the GP is able to correctly exclude ACS (TNs), accompanied by a decrease in unnecessary hospital referrals (FPs). However, the rate of correct referrals (TPs) decreases slightly, while the rate at which ACS patients are incorrectly not referred (FNs) shows a small increase. Despite this, the average per patient effect on QALYs is very small (i.e. -0.0004). Costs per chest pain patient are expected to decrease with approximately €77, indicating that the POC troponin strategy is likely cost-saving as compared to current practice.

Strengths

The main strength of the current study is that it was designed to accurately represent the current work-up of GPs, in chest pain patients. To achieve this, three real decisions were

built into the model, in which the POC troponin test is unlikely to be used: (1) when the GPs considers a patient at high risk of ACS, (2) when the GP diagnoses STEMI using an ECG, or (3) when a patient presents <4 h after symptom onset. Consequently, the POC troponin test is only used in approximately 40% of all chest pain patients presenting to the GP. Furthermore, as it is unlikely that GPs base their referral decision solely on a POC troponin test result, the probability of GPs adhering to this result was also incorporated. Although those assumptions inevitably limit the potential impact of the POC troponin test, they are crucial to obtain a realistic estimate of the test's potential cost-utility.

Figure 3. Cost-effectiveness acceptability curve of the POC troponin strategy versus current practice, for a WTP ranging from €0/QALY to €200,000/QALY. The black line represents the base case cost-effectiveness acceptability curve. The 'non-inferiority' and '0.002 QALYs' line represent scenarios in which POC troponin has acceptable cost-utility, and is either at least equally effective, or does not involve a decrease of ≥ 0.002 QALYs. QALYs = quality-adjusted life years.

Secondly, the use of a patient-level model has several advantages. First of all, probabilities of ACS are both age and gender specific, allowing to estimate costs, QALYs and productivity losses at the individual patient level. In addition, a duration of symptoms was assigned to each patient, allowing to incorporate time-dependent sensitivities and specificities, which is an important characteristic of the POC troponin test [7].

Furthermore, all assumptions used in the analysis were conservative regarding the potential impact of POC troponin, which may have led to underestimating both cost savings and health

benefits. First, the delay by POC troponin testing was set at 10 min, thereby assuming that the test is performed after the consultation by the GP, whereas it is likely that the test can already be performed during the patient's anamnesis and physical examination. Besides cost savings, this work-up would also prevent the potential negative health impact caused by a delay in treatment. Second, it was assumed that GPs would perform the test in all remaining chest pain patients, and therefore also in patients that are not actually suspected of having ACS. However, GPs can probably exclude ACS with certainty in some patients without using the POC troponin test, thereby saving costs. Third, it was assumed that the 14.7% of patients that are perceived to be at high risk of ACS by their GP (and who are immediately referred), corresponded to the patients with the highest actual probability of ACS. This assumption most likely overestimates the ability of the GP to identify high-risk patients.

Limitations

There are two limitations to be mentioned. First, evidence on the correlation between several input parameters were lacking. For example, no evidence was available on the relationship between (a) the pre-test probability of ACS, (b) diagnostic accuracy of the GP in current practice, and (c) the diagnostic accuracy of the POC troponin test. Similarly, no evidence was available on the potential influence of patient's characteristics, such as age and gender, on the probability that the GP revises his or her initial referral decision following a discordant POC troponin result. Therefore, changes in referral decisions made in practice may concern different patients than those with a change in referral decision in our model. However, results of a systematic review and meta-analysis indicate that it was not possible to define an important role for signs and symptoms in diagnosing AMI or ACS [38]. Accordingly, it is unlikely that this limitation strongly affects model outcomes.

Second, as this study is based on a hypothetical cohort of patients, the impact of POC troponin testing on health outcomes is dependent on the literature used in the model. However, evidence regarding the diagnostic performance of GPs (without the use of POC) is limited. The drawback of the study used in the current analysis, is that the sensitivity of the GP was based on their 'action in daily practice' in chest pain patients. This definition however, included besides referral to the hospital, also exercise testing, prescribing medication, and requesting a second opinion [18]. In addition, the diagnostic performance may also be dependent on the type of POC analyzer used, but the number of studies that report the diagnostic performance of POC troponin for (a) specific time(s) since symptom onset is limited. Furthermore, more sensitive POC troponin tests have already become available [39]. All in all, the current model may thus have overestimated the rate of missed ACS patients due to POC troponin use. In absence of real-life patient data, this model-based analysis gives a first insight in to the potential cost-effectiveness of the POC troponin test. Our results indicate that future prospective studies into the safety and efficacy of using POC troponin in primary care are warranted.

Implications for practice

Results from a previous study indicate that GPs' main requirements regarding POC troponin tests are a) test result within 10 min, b) a test performed with a finger prick, and c) reimbursement of the POC analyzer [15]. Recently developed POC troponin tests may already comply with the desired turn-around-time and the finger prick blood sample [40]. In addition, literature indicates that the diagnostic performance of recently developed POC troponin tests is comparable to the results obtained in a routine laboratory. Consequently, the accompanying increase in negative predictive value likely prevents that ACS patients are missed, thereby making use of the POC troponin test a cost-effective strategy. In turn, further research is necessary to investigate whether this lower limit of detection may allow use of this test to rule out ACS in patients presenting within four hours after symptom onset. In addition, including the expected improvement on patient's well-being (and consequently on QALYs) as achieved by decreasing unnecessary hospital referrals would further increase the potential benefit of this test.

CONCLUSION

Owing to the expected reduction in hospital referrals, the POC troponin strategy is likely cost-saving. The potential (accompanying) small increase in missed ACS patients can partly be explained by conservative assumptions used in the analysis. In addition, ongoing developments in POC troponin tests will likely further improve their diagnostic performance. Therefore, the use of a POC troponin testing strategy is likely to become cost-effective in the nearby future. When data from new prospective (randomized) studies into the safety and diagnostic accuracy of the use of POC troponin tests become available, the health economic model presented here allows rapid reassessment into the benefits of such a diagnostic tool for diagnosing ACS in general practices.

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This study was not funded.

Authors' contributions

MK, HK, MM and RK were involved the conception and design of the study. MK, HK, and MM designed the health economic model. All authors were involved in analyzing the health economic model and interpretation of the model outcomes. MK drafted the manuscript, and

HK, MM, MIJ and RK were major contributors in critically reviewing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

REFERENCES

- Okkes, I.M., et al., Continue morbiditeitsregistratie Nijmegen. Episodes of care in family practice. Epidemiological data based on the routine use of the International Classification of Primary Care (ICPC) in the Transition project of the Academic Medical Center/University of Amsterdam (1985-2003). 2005, Academic Medical Center/University of Amsterdam, Department of Family Medicine: Amsterdam.
- Authors/Task Force Members, et al., 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur Heart J*, 2015.
- Nederlandse Zorgautoriteit. Open data van de Nederlandse Zorgautoriteit. 2016 March 1, 2016 [cited 2016 March 29]; Available from: <http://www.opendisdata.nl/>.
- Bruins Slot, M.H., et al., Diagnosing acute coronary syndrome in primary care: comparison of the physicians' risk estimation and a clinical decision rule. *Fam Pract*, 2011. 28(3): p. 323-8.
- St John, A. and C.P. Price, Economic Evidence and Point-of-Care Testing. *Clin Biochem Rev*, 2013. 34(2): p. 61-74.
- Nilsson, S., et al., Cost consequences of point-of-care troponin T testing in a Swedish primary health care setting. *Scand J Prim Health Care*, 2014. 32(4): p. 241-7.
- Diercks, D.B., et al., Diagnostic accuracy of a point-of-care troponin I assay for acute myocardial infarction within 3 hours after presentation in early presenters to the emergency department with chest pain. *Am Heart J*, 2012. 163(1): p. 74-80 e4.
- Bruins Slot, M.H., et al., Point-of-care tests in suspected acute myocardial infarction: a systematic review. *Int J Cardiol*, 2013. 168(6): p. 5355-62.
- Dutch College of General Practitioners. NHG-Standaard Acuut coronair syndroom (eerste herziening). 2012; Available from: <https://www.nhg.org/standaarden/volledig/nhg-standaard-acuut-coronair-syndroom-eerste-herziening#note-13>.
- Bossuyt, P.M., J.G. Lijmer, and B.W. Mol, Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*, 2000. 356(9244): p. 1844-7.
- Moons, K.G., Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem*, 2010. 56(4): p. 537-41.
- Bosner, S., et al., Gender differences in presentation and diagnosis of chest pain in primary care. *BMC Fam Pract*, 2009. 10: p. 79.
- Bosner, S., et al., Chest pain in primary care: epidemiology and pre-work-up probabilities. *Eur J Gen Pract*, 2009. 15(3): p. 141-6.
- Bruins Slot, M.H., et al., Diagnostic value of a heart-type fatty acid-binding protein (H-FABP) bedside test in suspected acute coronary syndrome in primary care. *Int J Cardiol*, 2013. 168(2): p. 1485-9.
- Kip, M.M.A., et al., Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners. *Prim Health Care Res Dev*, 2017: p. 1-12.
- van Hassel, D., et al., De Toekomstvisie Huisartsenzorg 2011, waar staat de huisartsenzorg anno 2014? 2015, NIVEL: Utrecht. p. 1-85.
- Martin, T.N., et al., ST-segment deviation analysis of the admission 12-lead electrocardiogram as an aid to early diagnosis of acute myocardial infarction with a cardiac magnetic resonance imaging gold standard. *J Am Coll Cardiol*, 2007. 50(11): p. 1021-8.
- Nilsson, S., K. Ortoft, and S. Molstad, The accuracy of general practitioners' clinical assessment of chest pain patients. *Eur J Gen Pract*, 2008. 14(2): p. 50-5.
- Bruins Slot, M.H., et al., Gender differences in pre-hospital time delay and symptom presentation in patients suspected of acute coronary syndrome in primary care. *Fam Pract*, 2012. 29(3): p. 332-7.
- R Core Team, R: A language and environment for statistical computing. 2015, R Foundation for Statistical Computing: Vienna, Austria.
- Nallamothu, B., et al., Relationship of treatment delays and mortality in patients undergoing fibrinolysis and primary percutaneous coronary intervention. The Global Registry of Acute Coronary Events. *Heart*, 2007. 93(12): p. 1552-5.
- Mourad, G., et al., Societal costs of non-cardiac chest pain compared with ischemic heart disease--a longitudinal study. *BMC Health Serv Res*, 2013. 13: p. 403.
- Ghushchyan, V., K.V. Nair, and R.L. Page, 2nd, Indirect and direct costs of acute coronary syndromes with comorbid atrial fibrillation, heart failure, or both. *Vasc Health Risk Manag*, 2015. 11: p. 25-34.
- Hakkaart-van Roijen, L., et al., Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. 2015, Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam: Rotterdam. p. 1-120.
- Goodacre, S., et al., Systematic review, meta-analysis and economic modelling of diagnostic strategies for suspected acute coronary

- syndrome. *Health Technol Assess*, 2013. 17(1): p. v-vi, 1-188.
26. Polanczyk, C.A., et al., Emergency department triage strategies for acute chest pain using creatine kinase-MB and troponin I assays: a cost-effectiveness analysis. *Ann Intern Med*, 1999. 131(12): p. 909-18.
 27. Ward, S., et al., A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technol Assess*, 2007. 11(14): p. 1-160, iii-iv.
 28. Alter, D.A., et al., The average lifespan of patients discharged from hospital with heart failure. *J Gen Intern Med*, 2012. 27(9): p. 1171-9.
 29. Sullivan, P.W. and V. Ghushchyan, Preference-Based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making*, 2006. 26(4): p. 410-20.
 30. Nederlandse Zorgautoriteit, Tarieventabel DBC zorgproducten en overige producten per 1 januari 2016 (bijlage bij TB-CU-2066-01). 2016.
 31. Rutten, F.H., M.J.J.C. Poos, and P.M. Engelfriet. Hoe vaak komen hartfalen voor en hoeveel mensen sterven eraan? In: *Volksgezondheid en Toekomst Verkenning*. 2014 [cited 2016 March 31]; Available from: <http://www.naalkompas.nl/gezondheid-en-ziekte/ziekten-en-aandoeningen/hartvaatstelsel/hartfalen/omvang/>.
 32. Volksgezondheidenzorg.info. Cijfertool Kosten van Ziekten. 2015 [cited 2015 November 20]; Available from: https://kosten-vanziektentool.volksgezondheidenzorg.info/tool/nederlands/?ref=k vz_v2l1b1p4r3c0i0t-1j0o3y6a-1g0d207s54f0z0w2.
 33. Centraal Bureau voor de Statistiek. Consumentenprijzen; prijsindex 2006 = 100. 2016 [cited 2016 March 31]; Available from: <http://statline.cbs.nl/Statweb/publication/?DM=SL-NL&PA=71311NED&D1=0&D2=0&D3=64,77,90,103,116,129,142,155,168,181,194,219,232,245,258,l&HDR=G1,T&STB=G2&VW=T>
 34. Belastingdienst. Wisselkoersen. 2016 [cited 2016 March 31]; Available from: <http://www.belastingdienst.nl/rekenhulpen/wisselkoersen/>.
 35. Briggs, A. and K. Claxton, *Decision Modelling for Health Economic Evaluation*, in NY. 2006, Oxford University Press: New York.
 36. Strong, M., J.E. Oakley, and A. Brennan, Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decis Making*, 2014. 34(3): p. 311-26.
 37. Drummond, M., Introducing economic and quality of life measurements into clinical studies. *Ann Med*, 2001. 33(5): p. 344-9.
 38. Bruyninckx, R., et al., Signs and symptoms in diagnosing acute myocardial infarction and acute coronary syndrome: a diagnostic meta-analysis. *Br J Gen Pract*, 2008. 58(547): p. 105-11.
 39. Aldous, S., et al., Comparison of new point-of-care troponin assay with high sensitivity troponin in diagnosing myocardial infarction. *Int J Cardiol*, 2014. 177(1): p. 182-6.
 40. Philips In-vitro Diagnostics. Minicare I-20, Enabling near patient blood testing in the acute care setting. Available from: <http://www.philips.nl/healthcare/product/HCNOCTN496/minicare-i20-enabling-near-patient-blood-testing-in-the-acute-care-setting>.

APPENDIX 1

Model input parameters

Table 1a. model input parameters concerning patient pathways used in the model, showing the parameter, the value used in the model, the 95% confidence interval, the distribution used, and the data source. ACS = acute coronary syndrome, AMI = acute myocardial infarction, CI = confidence interval, ECG = electrocardiogram, ED = emergency department, GP = general practitioner, HF = heart failure, NSTEMI = non-ST elevation myocardial infarction, POCT = point-of-care test, STEMI = ST elevation myocardial infarction, UAP = unstable angina pectoris. *Calculated probabilities based on the total occurrence of ACS, and the age and gender specific probabilities of patients presenting to the GP with chest pain.

Parameter	Category	Probability	95% CI	Distribution	Reference
Population age distribution of chest pain patients at the GP (male)	35 – 44 years	21.2%	17.9% to 24.9%	Dirichlet	[1]
	45 – 54 years	21.3%	17.7% to 24.5%	Dirichlet	[1]
	55 – 64 years	20.8%	17.6% to 24.5%	Dirichlet	[1]
	65 – 74 years	23.4%	19.9% to 27.0%	Dirichlet	[1]
	≥ 75 years	13.3%	10.6% to 16.3%	Dirichlet	[1]
Population age distribution of chest pain patients at the GP (female)	35 – 44 years	17.0%	14.3% to 19.9%	Dirichlet	[1]
	45 – 54 years	20.5%	17.4% to 23.5%	Dirichlet	[1]
	55 – 64 years	18.3%	15.4% to 21.3%	Dirichlet	[1]
	65 – 74 years	25.4%	22.4% to 29.0%	Dirichlet	[1]
	≥ 75 years	18.9%	16.0% to 21.8%	Dirichlet	[1]
Probability of ACS for age (male)	35 – 44 years	1.7%	0.2% to 4.6%	Beta	[2]*
	45 – 54 years	1.5%	0.1% to 4.5%	Beta	[2]*
	55 – 64 years	3.8%	1.1% to 8.1%	Beta	[2]*
	65 – 74 years	7.1%	3.3% to 12.1%	Beta	[2]*
	≥ 75 years	10.3%	4.4% to 18.4%	Beta	[2]*
Probability of ACS for age (female)	35 – 44 years	1.0%	0.0% to 3.4%	Beta	[2]*
	45 – 54 years	0.9%	0.0% to 2.9%	Beta	[2]*
	55 – 64 years	2.2%	0.4% to 5.4%	Beta	[2]*
	65 – 74 years	4.2%	1.8% to 7.6%	Beta	[2]*
	≥ 75 years	6.0%	2.6% to 10.7%	Beta	[2]*
Suspected ACS gender distribution	Male	43.9%	41.3% to 46.8%	Beta	[1]

Time from symptom onset to presentation (male)	< 4 hours	47.5%	38.9% to 55.4%	Dirichlet	[3, 4]
	4 – 5.5 hours	9.0%	5.2% to 14.5%	Dirichlet	[3, 4]
	5.5 – 7 hours	7.3%	3.4% to 11.8%	Dirichlet	[3, 4]
	7 – 10 hours	9.4%	5.2% to 14.9%	Dirichlet	[3, 4]
	> 10 hours	26.8%	20.0% to 34.6%	Dirichlet	[3, 4]
Time from symptom onset to presentation (female)	< 4 hours	54.8%	47.1% to 62.7%	Dirichlet	[3, 4]
	4 – 5.5 hours	9.4%	5.1% to 13.9%	Dirichlet	[3, 4]
	5.5 – 7 hours	6.5%	3.1% to 10.7%	Dirichlet	[3, 4]
	7 – 10 hours	8.0%	4.6% to 13.1%	Dirichlet	[3, 4]
	> 10 hours	21.2%	15.2% to 27.8%	Dirichlet	[3, 4]
ECG	Performed at GP's office	57.4%	56.0% to 58.7%	Beta	[5]
	Sensitivity for STEMI	50.0%	37.6% to 62.6%	Beta	[6]
Perceived probability of ACS by GP	High probability	14.7%	8.2% to 22.5%	Beta	[7]
Revise referral decision after discordant POCT troponin	Negative troponin	80.6%	73.3% to 87.1%	Beta	[7]
	Positive troponin	83.3%	76.3% to 89.3%	Beta	[7]
Diagnostic performance of GP, without POCT troponin	Sensitivity	88.3%	79.2% to 95.2%	Beta	[8]
	Specificity	72.2%	67.9% to 76.3%	Beta	[8]
Sensitivity of POC troponin at different duration of symptoms	4 – 5.5 hours	66.7%	55.5% to 77.0%	Beta	[9]
	5.5 – 7 hours	79.2%	69.3% to 87.5%	Beta	[9]
	7 – 10 hours	84.7%	75.3% to 91.9%	Beta	[9]
	> 10 hours	87.5%	78.8% to 94.2%	Beta	[9]
Specificity of POC troponin at different duration of symptoms	4 – 5.5 hours	95.9%	94.2% to 97.3%	Beta	[9]
	5.5 – 7 hours	94.4%	92.5% to 96.1%	Beta	[9]
	7 – 10 hours	93.4%	91.3% to 95.3%	Beta	[9]
	> 10 hours	92.6%	90.4% to 94.5%	Beta	[9]
Mortality modifiers	Per 10 years increase in age	1.7	1.6 to 1.9	LogNormal	[10]
	Missed UAP	1.7	0.2 to 15.2	LogNormal	[11]
	Missed NSTEMI	1.9	0.7 to 5.2	LogNormal	[11]
	Missed STEMI	1.9	0.7 to 5.2	LogNormal	[11]
	Per 10 minutes delay due to POCT troponin – NSTEMI	0.2%	0.1% to 0.4%	Beta	[12]
	Per 10 minutes delay due to POCT troponin – STEMI	0.2%	0.1% to 0.4%	Beta	[12]

Occurrence of each type of ACS (among ACS)	UAP or NSTEMI	72.7%	61.5% to 82.5%	Beta	[13]
	UAP (among UAP or NSTEMI)	29.2%	17.3% to 42.5%	Beta	[13]
Death at ED	Diagnosed UAP	2.7%	2.4% to 3.1%	Beta	[14]
	Diagnosed NSTEMI	5.9%	5.4% to 6.4%	Beta	[14]
	Diagnosed STEMI	7.8%	7.2% to 8.4%	Beta	[14]
Heart failure	Diagnosed UAP	9.2%	8.5% to 10.0%	Beta	[14]
	Missed UAP	44.4%	23.0% to 67.0%	Beta	[15]
	Diagnosed NSTEMI	17.9%	16.9% to 18.9%	Beta	[14]
	Missed NSTEMI	44.4%	23.0% to 66.4%	Beta	[15]
	Diagnosed STEMI	18.4%	17.4% to 19.4%	Beta	[14]
	Missed STEMI	44.4%	22.7% to 66.7%	Beta	[15]
Production hours lost due to UAP, NSTEMI, STEMI, HF or death	UAP	91.4	51.6 to 142.1	Gamma	[16]
	NSTEMI	71.7	41.0 to 111.3	Gamma	[16]
	STEMI	71.7	41.0 to 111.3	Gamma	[16]
	UAP/NSTEMI/STEMI + HF	408.9	233.6 to 638.1	Gamma	[17]
	Death	675.1	389.1 to 1045.9	Gamma	[18]
Working hours lost due to GP and possibly ED visit (non ACS patients)	GP, referral	8.0	4.6 to 12.5	Gamma	Assumption, [18]
	GP, no referral, no POCT	0.8	0.5 to 1.3	Gamma	Assumption, [18]
	GP, no referral, POCT	0.7	0.4 to 1.0	Gamma	Assumption, [18]

Table 1b. Quality of life estimates used in the model, showing the parameter, the value used in the model, the 95% confidence interval, the distribution used, and the data source. ACS = acute coronary syndrome, AMI = acute myocardial infarction, CI = confidence interval, HF = heart failure, NSTEMI = non-ST elevation myocardial infarction, STEMI = ST elevation myocardial infarction, UAP = unstable angina pectoris, QALY = quality-adjusted life year. * 95% CI is based on an assumed standard error of 25%.

Parameter	Category	QALYs	95% CI *	Distribution	Reference
Lifetime QALYs AMI (NSTEMI or STEMI)	35 – 44 years	12.2	6.2 to 18.2	Gamma	[19]
	45 – 54 years	9.5	4.8 to 14.1	Gamma	[19]
	55 – 64 years	6.7	3.4 to 10.0	Gamma	[19]
	65 – 74 years	4.7	2.4 to 6.9	Gamma	[19]
	≥ 75 years	2.4	1.2 to 3.6	Gamma	[19]
Lifetime QALYs UAP	35 – 44 years	13.0	6.6 to 19.4	Gamma	[19-21]
	45 – 54 years	10.2	5.2 to 15.2	Gamma	[19-21]
	55 – 64 years	7.3	3.7 to 10.9	Gamma	[19-21]
	65 – 74 years	5.4	2.7 to 8.0	Gamma	[19-21]
	≥ 75 years	3.1	1.6 to 4.6	Gamma	[19-21]
Lifetime QALYs without ACS (males)	35 – 44 years	19.5	9.9 to 29.0	Gamma	[19]
	45 – 54 years	15.8	8.1 to 23.6	Gamma	[19]
	55 – 64 years	11.5	5.9 to 17.1	Gamma	[19]
	65 – 74 years	7.6	3.9 to 11.3	Gamma	[19]
	≥ 75 years	3.3	1.7 to 4.9	Gamma	[19]
Lifetime QALYs without ACS (females)	35 – 44 years	19.0	9.7 to 28.3	Gamma	[19]
	45 – 54 years	15.4	7.9 to 23.0	Gamma	[19]
	55 – 64 years	11.2	5.7 to 16.7	Gamma	[19]
	65 – 74 years	7.4	3.8 to 11.0	Gamma	[19]
	≥ 75 years	3.2	1.6 to 4.7	Gamma	[19]
Lifetime QALYs HF (males)	35 – 44 years	9.9	5.1 to 14.8	Gamma	[22, 23]
	45 – 54 years	7.5	3.8 to 11.2	Gamma	[22, 23]
	55 – 64 years	5.0	2.6 to 7.5	Gamma	[22, 23]
	65 – 74 years	3.4	1.8 to 5.1	Gamma	[22, 23]
	≥ 75 years	2.5	1.3 to 3.8	Gamma	[22, 23]
Lifetime QALYs HF (females)	35 – 44 years	9.8	5.0 to 14.6	Gamma	[22, 23]
	45 – 54 years	7.5	3.8 to 11.1	Gamma	[22, 23]
	55 – 64 years	5.1	2.6 to 7.5	Gamma	[22, 23]
	65 – 74 years	3.4	1.8 to 5.1	Gamma	[22, 23]
	≥ 75 years	2.5	1.3 to 3.7	Gamma	[22, 23]

Table 1c. Cost parameters used in the model, showing the parameter, the value used in the model, the 95% confidence interval, the distribution used, and the data source. ACS = acute coronary syndrome, AMI = acute myocardial infarction, CI = confidence interval, ECG = electrocardiogram, ED = emergency department, GP = general practitioner, HF = heart failure, NSTEMI = non-ST elevation myocardial infarction, POCT = point-of-care test, STEMI = ST elevation myocardial infarction, UAP = unstable angina pectoris. All costs >€1,000 are rounded to zero decimals. *95% CI is based on an assumed standard error of 25%.

Parameter	Category	Costs	95% CI*	Distribution	Reference
Costs at GP office and ambulance transport	GP consultation (10 minutes)	€33.21	€18.84 to €51.36	Gamma	[18]
	ECG at GP office	€9.59	€5.52 to €14.91	Gamma	[24]
	POCT troponin at GP	€15.00	€8.66 to €23.38	Gamma	Assumption
	Ambulance to ED	€518.33	€295.93 to €802.29	Gamma	[18]
Diagnostic and treatment costs	UAP treatment	€3,042	€1,782 to €4,687	Gamma	[25]
	NSTEMI treatment	€4,878	€2,812 to €7,479	Gamma	[25]
	STEMI treatment	€5,771	€3,355 to €8,893	Gamma	[25]
	Diagnostics ED (no ACS)	€717.57	€414.89 to €1,103	Gamma	[25]
Discounted lifetime costs AMI patients	30 – 44 years	€5,185	€2,943 to €7,992	Gamma	[19]
	45 – 54 years	€4,025	€2,276 to €6,248	Gamma	[19]
	55 – 64 years	€2,862	€1,650 to €4,416	Gamma	[19]
	65 – 74 years	€1,977	€1,145 to €3,061	Gamma	[19]
	≥ 75 years	€1,034	€597.70 to €1,603	Gamma	[19]
Discounted lifetime costs UAP patients	30 – 44 years	€5,453	€3,115 to €8,391	Gamma	[19-21]
	45 – 54 years	€4,271	€2,420 to €6,606	Gamma	[19-21]
	55 – 64 years	€3,059	€1,776 to €4,709	Gamma	[19-21]
	65 – 74 years	€2,261	€1,296 to €3,482	Gamma	[19-21]
	≥ 75 years	€1,295	€736.99 to €2,018	Gamma	[19-21]
Discounted lifetime costs HF patients (male)	30 – 44 years	€113,073	€63,447 to €176,649	Gamma	[23, 26, 27]
	45 – 54 years	€89,827	€51,944 to €139,300	Gamma	[23, 26, 27]
	55 – 64 years	€63,619	€35,998 to €98,820	Gamma	[23, 26, 27]
	65 – 74 years	€44,757	€25,580 to €68,797	Gamma	[23, 26, 27]
	≥ 75 years	€33,481	€18,957 to €51,527	Gamma	[23, 26, 27]
Discounted lifetime costs HF patients (female)	30 – 44 years	€112,354	€64,477 to €171,333	Gamma	[23, 26, 27]
	45 – 54 years	€89,624	€50,804 to €138,934	Gamma	[23, 26, 27]
	55 – 64 years	€63,735	€36,442 to €99,045	Gamma	[23, 26, 27]
	65 – 74 years	€44,686	€25,219 to €69,384	Gamma	[23, 26, 27]
	≥ 75 years	€33,083	€18,879 to €51,394	Gamma	[23, 26, 27]

Costs of production loss per hour (male)	35 – 45 years	€33.41	€19.18 to €51.13	Gamma	[18]
	45 – 55 years	€36.56	€20.89 to €56.94	Gamma	[18]
	55 – 65 years	€37.10	€20.81 to €57.74	Gamma	[18]
Costs of production loss per hour (female)	35 – 45 years	€27.55	€15.71 to €42.47	Gamma	[18]
	45 – 55 years	€27.48	€15.97 to €42.56	Gamma	[18]
	55 – 65 years	€27.48	€15.68 to €42.12	Gamma	[18]

REFERENCES

- Bosner, S., et al., Gender differences in presentation and diagnosis of chest pain in primary care. *BMC Fam Pract*, 2009. 10: p. 79.
- Bosner, S., et al., Chest pain in primary care: epidemiology and pre-work-up probabilities. *Eur J Gen Pract*, 2009. 15(3): p. 141-6.
- Bruins Slot, M.H., et al., Gender differences in pre-hospital time delay and symptom presentation in patients suspected of acute coronary syndrome in primary care. *Fam Pract*, 2012. 29(3): p. 332-7.
- R Core Team, R: A language and environment for statistical computing. 2015, R Foundation for Statistical Computing: Vienna, Austria.
- van Hassel, D., et al., De Toekomstvisie Huisartsenzorg 2011, waar staat de huisartsenzorg anno 2014? 2015, NIVEL: Utrecht. p. 1-85.
- Martin, T.N., et al., ST-segment deviation analysis of the admission 12-lead electrocardiogram as an aid to early diagnosis of acute myocardial infarction with a cardiac magnetic resonance imaging gold standard. *J Am Coll Cardiol*, 2007. 50(11): p. 1021-8.
- Kip, M.M.A., et al., Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners. *Prim Health Care Res Dev*, 2017: p. 1-12.
- Nilsson, S., K. Ortoft, and S. Molstad, The accuracy of general practitioners' clinical assessment of chest pain patients. *Eur J Gen Pract*, 2008. 14(2): p. 50-5.
- Diercks, D.B., et al., Diagnostic accuracy of a point-of-care troponin I assay for acute myocardial infarction within 3 hours after presentation in early presenters to the emergency department with chest pain. *Am Heart J*, 2012. 163(1): p. 74-80 e4.
- Granger, C.B., et al., Predictors of hospital mortality in the global registry of acute coronary events. *Arch Intern Med*, 2003. 163(19): p. 2345-53.
- Pope, J.H., et al., Missed diagnoses of acute cardiac ischemia in the emergency department. *N Engl J Med*, 2000. 342(16): p. 1163-70.
- Nallamothu, B., et al., Relationship of treatment delays and mortality in patients undergoing fibrinolysis and primary percutaneous coronary intervention. *The Global Registry of Acute Coronary Events*. *Heart*, 2007. 93(12): p. 1552-5.
- Bruins Slot, M.H., et al., Diagnostic value of a heart-type fatty acid-binding protein (H-FABP) bedside test in suspected acute coronary syndrome in primary care. *Int J Cardiol*, 2013. 168(2): p. 1485-9.
- Goldberg, R.J., et al., Six-month outcomes in a multinational registry of patients hospitalized with an acute coronary syndrome (the Global Registry of Acute Coronary Events [GRACE]). *Am J Cardiol*, 2004. 93(3): p. 288-93.
- Sequist, T.D., et al., Prediction of missed myocardial infarction among symptomatic outpatients without coronary heart disease. *Am Heart J*, 2005. 149(1): p. 74-81.
- Mourad, G., et al., Societal costs of non-cardiac chest pain compared with ischemic heart disease—a longitudinal study. *BMC Health Serv Res*, 2013. 13: p. 403.
- Ghushchyan, V., K.V. Nair, and R.L. Page, 2nd, Indirect and direct costs of acute coronary syndromes with comorbid atrial fibrillation, heart failure, or both. *Vasc Health Risk Manag*, 2015. 11: p. 25-34.
- Hakkaart-van Roijen, L., et al., Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. 2015, Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam: Rotterdam. p. 1-120.
- Goodacre, S., et al., Systematic review, meta-analysis and economic modelling of diagnostic strategies for suspected acute coronary syndrome. *Health Technol Assess*, 2013. 17(1): p. v-vi, 1-188.
- Polanczyk, C.A., et al., Emergency department triage strategies for acute chest pain using creatine kinase-MB and troponin I assays: a cost-effectiveness analysis. *Ann Intern Med*, 1999. 131(12): p. 909-18.

21. Ward, S., et al., A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technol Assess*, 2007. 11(14): p. 1-160, iii-iv.
22. Sullivan, P.W. and V. Ghushchyan, Preference-Based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making*, 2006. 26(4): p. 410-20.
23. Alter, D.A., et al., The average lifespan of patients discharged from hospital with heart failure. *J Gen Intern Med*, 2012. 27(9): p. 1171-9.
24. Nederlandse Zorgautoriteit, Bijlage 1 bij tariefbeschikking TB/CU-7089-02: tarievenlijst huisartsenzorg en multidisciplinaire zorg. 2015.
25. Nederlandse Zorgautoriteit. Open data van de Nederlandse Zorgautoriteit. 2016 March 1, 2016 [cited 2016 March 29]; Available from: <http://www.opendisdata.nl/>.
26. Volksgezondheidszorg.info. Cijfertools Kosten van Ziekten. 2015 [cited 2015 November 20]; Available from: https://kostenvanziektentool.volksgezondheidszorg.info/tool/nederlands/?ref=kvz_v2l1b1p4r3c0i0t-1j0o3y6a-1g0d207s54f0z0w2.
27. Rutten, F.H., M.J.J.C. Poos, and P.M. Engelfriet. Hoe vaak komen hartfalen voor en hoeveel mensen sterven eraan? In: *Volksgezondheid en Toekomst Verkenning*. 2014 [cited 2016 March 31]; Available from: <http://www.naalkompas.nl/gezondheid-en-ziekte/ziekten-en-aandoeningen/hartvaatstelsel/hartfalen/omvang/>.

APPENDIX 2

Overview of model assumptions

Assumptions related to the diagnostic work-up of the general practitioner (GP):

1. In case the GP perceives a patient to be at high risk of acute coronary syndrome (ACS), the patient is immediately referred to the hospital, without performing an electrocardiogram (ECG) or a point-of-care (POC) troponin test.
2. In case a GP has the availability of an ECG in his or her general practice, the GP is also assumed to perform an ECG in all chest pain patients presenting in their general practice, except for patients that are immediately referred to the hospital because the GP perceives a high probability of ACS in those patients.
3. In case the GP diagnoses a ST elevation myocardial infarction (STEMI) based on the ECG, the patient is immediately referred to the hospital, without performing a POC troponin test.
4. It is assumed that the POC troponin test is not used in patients presenting within four hours after symptom onset.
5. All patients that are referred to the hospital are assumed to be transported by ambulance.
6. In case the initial decision of the GP was to refer the patient, but the POC troponin level was negative, the probability that the GP would no longer refer this patient was 81%. For a positive POC troponin level in a patient that was initially not referred, this probability was found to be 83% [1].
7. As a previous study indicates that 10 minutes is considered the maximum acceptable test duration for POC troponin [manuscript submitted, 1], and because literature has shown that POC troponin tests who meet this requirement are currently available [2] it was assumed that the POC troponin test takes 10 minutes to perform.
8. The costs of performing a POC troponin test by the GP are assumed to be €15.

Assumptions related to treatment as well as treatment outcomes:

1. In case an ACS patient is not immediately referred to the hospital, it is assumed that the patient will present to the hospital at a later point in time. However, this leads to a delay in treatment, and consequently to an increased risk of mortality and developing heart failure.
2. A higher mortality rate per 10 minutes treatment delay has been reported for STEMI patients, indicating a slightly higher mortality rate attributable to the time delay caused by POC troponin testing. It is conservatively assumed that this time delay is also applicable to non-ST elevation myocardial infarction (NSTEMI) patients [3].

Assumptions related to productivity loss:

1. All patients are assumed to retire at the age of 65, indicating that no costs of productivity loss are assumed to occur in patients aged 65 years and older.

2. As the distance to the GP is approximately 1.1 kilometer [4], the travel time to the GP is set at 15 minutes. Therefore, a GP consultation is assumed to take 2 x 15 minutes travel time, plus 1 x 10 minutes consultation time = 40 minutes. In case a POC troponin test is performed, an additional time of 10 minutes is assumed.
3. The duration of an emergency department referral for chest pain is estimated by various hospitals to differ between 2 and 24 hours, but 8 hours is often considered as the average duration. Therefore, the duration of an ED referral, including the consultation at the GP, is estimated to take one working day, i.e. 8 hours.

REFERENCES

1. Kip, M.M.A., et al., Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners. *Prim Health Care Res Dev*, 2017: p. 1-12.
2. Yang, Z. and D. Min Zhou, Cardiac markers and their point-of-care testing for diagnosis of acute myocardial infarction. *Clin Biochem*, 2006. 39(8): p. 771-80.
3. Nallamothu, B., et al., Relationship of treatment delays and mortality in patients undergoing fibrinolysis and primary percutaneous coronary intervention. *The Global Registry of Acute Coronary Events. Heart*, 2007. 93(12): p. 1552-5.
4. Hakkaart-van Roijen, L., et al., Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg., 2015, Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam: Rotterdam. p. 1-120.

APPENDIX 3

Results of partial perfect value of information analysis

Parameter group:	Per person EVPPI (€)	Indexed to overall EVPPI	EVPPI for the Netherlands per year (€)
Probabilities regarding population parameters (age, gender, probabilities regarding STEMI, NSTEMI, UAP, death, heart failure, and the increased probability of mortality due to aging, missed ACS, and attributable to the time of POC troponin testing).	0.00	0.00	147.18
Probabilities regarding POC testing and decision-making (time to presentation, probability ECG, sensitivity of GP and POC troponin test, probability of revising referral decision).	0.00	0.01	350.06
QALYs (concerns all quality-of-life estimated used in the model).	0.00	0.01	340.62
Productivity loss (in hours) (productivity loss attributable to GP and hospital visits, as well as due to UAP, NSTEMI, STEMI, heart failure, and mortality).	0.00	0.00	0.00
Cost parameters (concerns all cost parameters used in the model).	0.01	0.02	942.37

PART II

Cost-effectiveness analysis of a
procalcitonin-guided algorithm for
antibiotic discontinuation in intensive
care patients with sepsis

5

A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis

Michelle M.A. Kip
Ron Kusters
Maarten J. IJzerman
Lotte M.G. Steuten

This chapter has been published as:

M.M.A. Kip, R. Kusters, M.J. IJzerman, L.M.G. Steuten. *A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis*. *J Med Econ*, 2015. 18(11): p.944-53.

Reprinted with permission.

ABSTRACT

Objective

Procalcitonin (PCT) is a specific marker for differentiating bacterial from non-infective causes of inflammation. It can be used to guide initiation and duration of antibiotic therapy in intensive care unit (ICU) patients with suspected sepsis, and might reduce the duration of hospital stay. Limiting antibiotic treatment duration is highly important because antibiotic over-use may cause patient harm, prolonged hospital stay, and resistance development. Several systematic reviews show that a PCT algorithm for antibiotic discontinuation is safe, but upfront investment required for PCT remains an important barrier against implementation. The current study investigates to what extent this PCT algorithm is a cost-effective use of scarce healthcare resources in ICU patients with sepsis compared to current practice.

Methods

A decision tree was developed to estimate the health economic consequences of the PCT algorithm for antibiotic discontinuation from a Dutch hospital perspective. Input data were obtained from a systematic literature review. When necessary, additional information was gathered from open interviews with clinical chemists and intensivists. The primary effectiveness measure is defined as number of antibiotic days, and cost-effectiveness is expressed as incremental costs per antibiotic day avoided.

Results

The PCT algorithm for antibiotic discontinuation is expected to reduce hospital spending by circa €3503 per patient, indicating savings of 9.2%. Savings are mainly due to reductions in length of hospital stay, number of blood cultures performed, and, importantly, days on antibiotic therapy. Probabilistic and one-way sensitivity analyses showed the model outcome to be robust against changes in model inputs.

Conclusion

Proven safe, a PCT algorithm for antibiotic discontinuation is a cost-effective means of reducing antibiotic exposure in adult ICU patients with sepsis, compared to current practice. Additional resources required for PCT are more than offset by downstream cost savings. This finding is highly important given the aim of preventing widespread antibiotic resistance.

BACKGROUND

Despite advances in medical technology and clinical care, sepsis remains a common cause of morbidity and mortality among hospitalized patients [1]. Diagnosing patients with sepsis is challenging, due to the often non-specific presentation [2]. Yet, early diagnosis of infection and rapid initiation of adequate antimicrobial therapy are critical for successful treatment outcome [3]. While the use of antibiotics has led to great reductions in mortality and morbidity rates among sepsis patients, antibiotic over-use should be avoided as this may cause patient harm and prolonged hospital stay, and plays a role in the development of widespread antibiotic resistance [4,5].

A biomarker that might improve the efficient and more judicious use of antibiotic therapy by monitoring the progression and prognosis of bacterial infections and sepsis is procalcitonin (PCT), a precursor of calcitonin. PCT elevation occurs within 2–4 h after onset of the inflammatory disorder, typically peaks in the second day, and falls rapidly during clinical recovery. The magnitude and duration of PCT elevation correlate with injury severity and prognosis. While PCT may also be elevated in viral and fungal infections (e.g., candidemia), this is generally much less so than with bacterial infections [6].

Several studies have assessed the added value of using PCT to monitor and manage antibiotic therapy in septic intensive care unit (ICU) patients, as well as in reducing the duration of hospital stay [7-9]. A number of systematic reviews have shown that a PCT discontinuation algorithm is safe and may even improve clinical outcomes [10-12]. Yet, the upfront investment for PCT testing compared to other laboratory assays remains a barrier against implementation. Therefore, analysis of the impact of PCT testing on in-hospital mortality, number of antibiotic days, duration of hospitalization, and total costs of sepsis care is warranted.

The goal of the current study is to investigate to what extent this PCT algorithm is a cost-effective use of scarce healthcare resources in ICU patients with sepsis compared to current practice. A model-based analysis was performed based on a systematic review of the literature published until mid-2014.

In addition to previously published cost-effectiveness analyses in this patient population [13, 14], this study explicitly considers the impact of PCT testing on hospital length of stay and on specific clinical outcomes, and reports costs from a non-US perspective.

METHODS

A decision tree was developed to estimate the health economic consequences of a PCT algorithm for antibiotic discontinuation in a hypothetical population of adult ICU patients with sepsis. The analysis was performed from a Dutch hospital perspective. The time horizon of the model covers the duration of a patient's hospital stay. All relevant health economic impacts

of hospital stay and accompanying treatment were incorporated, and compared to current practice. The primary effectiveness measure was defined as the number of antibiotic days in both the PCT strategy and current practice. In this analysis, the total direct hospital costs were balanced against the number of antibiotic days avoided. The incremental cost-effectiveness ratio (ICER) was expressed as incremental costs per antibiotic day avoided and calculated as the difference in direct healthcare costs, between the PCT strategy and current practice, divided by number of antibiotic days avoided by the PCT strategy.

Literature review

A systematic literature review was performed to determine to what extent a PCT algorithm affects the number of antibiotics days, ICU length of stay, total duration of hospital stay, number of days on mechanical ventilation and/or dialysis, number of blood cultures and other lab analyses performed, as well as patient safety which is expressed as in-hospital mortality rates. The PubMed database was searched for relevant articles that reported outcomes on at least one of those parameters. The following combinations of terms were searched in all fields: (algorithm OR guide OR guided OR based) AND (sepsis OR septic shock OR critically ill) AND (PCT or procalcitonin) AND (antibiotic OR antibiotics). The search was limited to articles published in English or Dutch, and was restricted to randomised controlled trials (RCTs), meta-analyses, and systematic reviews. Articles were excluded when they did not focus on: (1) adult patients, (2) sepsis or critically ill patients on the ICU, and (3) the added value of a PCT algorithm for antibiotic discontinuation. Relevant articles were initially selected based on title and abstract. After that, full texts were reviewed to assess whether the papers met the inclusion criteria. The literature search was performed in July 2014. Mean values and standard deviations (SDs) were obtained from each of the individual studies where possible. For studies in which no mean or SDs could be obtained, estimates of mean and SDs were calculated according to Hozo et al. [15]. Following this, weighted mean differences were calculated using Review Manager version 5.1, combining the sample sizes of the studies included with the mean and standard deviations of each parameter (see Appendix 1) using a random effects model [16].

Resource use

Data concerning the length of a patient's hospital stay (both on the ICU and on the general ward), as well as the duration of antibiotic treatment in both the PCT strategy and in current practice, have been derived from the systematic literature review, as described above. In addition, the change in the duration of mechanical ventilation has also been derived from this review. The percentage of patients with sepsis that are treated with mechanical ventilation and/or dialysis, as well as the duration of each, were derived from a retrospective database analysis of ICU patients (age >16 years) performed by Adrie et al. [17]. The percentage of patients in whom a blood culture is performed and who are finally diagnosed as having sepsis was derived from an observational cohort study by Shapiro et al. [18].

Table 1. Resource use. Overview of resource use in the model, showing the values used in the model, and the values applied in the one-way sensitivity analysis in brackets. The right column shows the references used to obtain an estimate for each parameter.

Parameter	Value		Reference
	Without PCT	With PCT	
Percentage treated with antibiotics*	100.0%	100.0% (75.0–100.0%)	Expert opinion
Percentage treated with mechanical ventilation	77.0%	77.0% (57.8–96.3%)	Adrie et al. [17]
Percentage treated with dialysis	16.0%	16.0% (12–20%)	Adrie et al. [17]
Days on dialysis	5.0	5.0 (3.75–6.25)	Adrie et al. [17]
Percentage of patients with (suspected) sepsis in whom a blood culture is taken	97.5%	61.4% (79.5–43.4%)	Müller et al. [19]
Percentage of patients with blood culture performed, diagnosed as having sepsis	8.2%	8.2% (6.1–10.2%)	Shapiro et al. [18]
Sets of blood cultures taken per patient with (suspected) sepsis	2	2 (1.5–2.5)	Müller et al. [19]
Frequency of laboratory tests ordered per patient*	25.1	21.8 (23.4–20.1)	Expert opinion
Number of PCT measurements performed per patient**	0	5 (2.5–7.5)	Stocker et al. [20], Schuetz et al. [21]

* Expert opinions were obtained via interviews with intensivists ($n = 2$) and clinical chemists ($n = 5$).

** Because these publications both report that four to five PCT measurements are performed in neonates with suspected sepsis and patients with lower respiratory tract infections, respectively, an estimate of five PCT measurements is assumed to be a conservative estimate for adult ICU patients with sepsis.

A prospective cohort study by Müller et al. [19] found that PCT measurement is an accurate parameter for predicting bacteraemia in patients with community acquired pneumonia (CAP), and that it has the potential to reduce the number of blood cultures drawn from hospitalized patients with suspected CAP. Because pneumonia is a common site of infection for sepsis [17], the percentage of patients with suspected sepsis in whom a blood culture was performed (both with and without PCT) as well as the number of sets of blood cultures taken per patient, were derived from Müller et al. [19]. The number of PCT measurements performed in ICU patients with sepsis was estimated based on RCTs by Stocker et al. [20], and Schuetz et al. [21]. For the percentage of patients treated with antibiotics, and the frequency at which laboratory

tests (other than the PCT test) are ordered in ICU patients with sepsis, no single estimate could be obtained from the literature. Therefore, those were estimated based on qualitative interviews with intensivists ($n = 2$) and clinical chemists ($n = 5$). An overview of the resource use parameters, the data sources, and assumptions that served as input for the model is provided in Table 1.

Table 2. Cost parameters. Overview of cost parameters used in the model. The right column shows the references used.

Parameter	Value	Reference
Day on general ward	€ 500.96	Hakkaart-van Roijen et al. [25]
ICU admission	€ 485.23	NZA tariffs [23]
Day on ICU	€ 1811.79	NZA tariffs [23]
Day mechanical ventilation	€ 386.32	NZA tariffs [23]
Day dialysis treatment	€ 290.48	NZA tariffs [23]
Day of intravenous antibiotics	€ 126.66	Vandijck et al. [26]
Set of blood cultures performed	€ 57.84	Müller et al. [19], Van Nieuwkoop et al. [22]
Order tariff for laboratory tests	€ 13.73	NZA tariffs [23]
PCT test	€ 15.00	Expert opinion*
Other laboratory tests	Varying	NZA tariffs [23]

* Expert opinions were obtained via interviews with intensivists ($n = 2$) and clinical chemists ($n = 5$).

Unit costs

The model incorporates costs of diagnostic testing (i.e., blood cultures, PCT testing, and other routinely performed laboratory tests), hospital stay on ICU and general ward, antibiotic therapy, mechanical ventilation, and dialysis. Unit costs of blood cultures performed were derived from publications by Müller et al. [19] and Van Nieuwkoop et al. [22]. Tariffs for laboratory tests were derived from the Dutch Healthcare Authority (Nederlandse Zorgautoriteit, NZA) [23]. Because no such tariff currently exists for PCT, those costs were based on interviews with clinical chemists and intensivists. Unit costs of hospital stay, separate for ICU stay and stay on a general ward, mechanical ventilation, and dialysis were derived from the Dutch Healthcare Authority [24] and the costing manual by Hakkaart-van Roijen et al. [25]. Mean daily costs for antibiotic treatment were obtained from Vandijck et al. [26]. All costs were converted to 2013 Euros, using Dutch consumer price index levels [27]. Because the time horizon of the model concerns the duration of a patient's hospital stay, lasting shorter than 1 year, discounting is not required. A summary of all cost inputs used in the model is provided in Table 2. Direct hospital costs are calculated by multiplying resource use with the accompanying unit costs.

Sensitivity analyses

To determine the joint decision uncertainty, a probabilistic sensitivity analysis (PSA) was performed with 10,000 model runs, in which random samples are drawn from all input parameters simultaneously based on pre-defined parameter distributions. Distributions were parameterized based on the observed parameter mean and on the observed or assumed standard error [28]. Beta distributions were fitted to the probability parameters, and Gamma distributions to the resource use parameters [28].

To identify which individual parameters drive the model outcome we conducted a one-way sensitivity analysis. For each parameter, the impact of a change in the base case value across a pre-determined range on the ICER (i.e., costs per antibiotic day avoided) was analysed. Parameters concerning resource use that were obtained from the literature review (ICU days, general ward days, duration of antibiotics, and duration of mechanical ventilation) were all varied with 1 SD below and above the mean.

Parameters that are not directly affected by PCT implementation but which are to some extent uncertain in both strategies, were varied with 25% below and above the mean in the PCT strategy. In the remaining parameters for which an effect due to the PCT algorithm compared to current practice was expected, this effect was increased and decreased with 50% in the PCT strategy (Table 1). The results of this sensitivity analysis are shown in a tornado diagram, in which the impact of each parameter is sorted by decreasing impact on the ICER (i.e., costs per antibiotic day avoided).

RESULTS

Systematic review

The search strategy initially resulted in 27 articles. Based on the exclusion criteria, five articles were excluded because they did not focus on sepsis or critically ill patients at the ICU, three articles were excluded because they focused on infants or new-borns instead of adults, and two articles were excluded because they either focused specifically on point-of care testing or on the comparison with another laboratory marker instead of focusing on current practice. Finally, this resulted in the identification of 11 reviews (of which eight meta-analyses), five original RCTs and one economic evaluation of PCT. As the summarised or pooled data from the reviews and economic evaluation did not exactly match the data specification as needed for our model, the reviews and the economic evaluation were hand searched to find the relevant original studies (RCTs) to directly obtain the relevant original data. In addition to five RCTs already identified, this hand search yielded one additional RCT, amounting to a total of six unique RCTs that were included [7-9, 29-31]. Figure 1 shows a flow chart of the search strategy. Review Manager was used to calculate pooled estimates of the duration of ICU and general ward stay, duration of mechanical ventilation, duration of antibiotic treatment, and in-hospital

mortality. Mean values and SDs were obtained from each of the individual studies where possible. In one study, no mean and SDs were reported [29]. Because the sample size of this study was sufficiently large, the median was assumed to be the best estimate of the mean according to Hozo et al. [15], who also state that the accompanying SD can be calculated by dividing the interquartile range (IQR) by a factor 1.35 ($SD = IQR/1.35$) [15, 16].

Figure 1. Overview of included and excluded articles. This figure shows an overview of the selection process of the studies included in the systematic literature review.

* Consisting of 11 reviews (of which 8 meta-analyses), and 1 economic evaluation.

Cost-effectiveness

Five studies reported in-hospital mortality rates and showed no statistically significant difference between the PCT strategy and current practice ($OR = 0.83$; $95\% CI = 0.49-1.38$) [7-9, 29, 31]. Therefore, equal in-hospital mortality rates of the PCT algorithm compared to current practice were applied.

The PCT algorithm applied varied between studies, using either a decrease in the peak PCT value (ranging from 20–90%), and/or a decrease in the absolute value of PCT (ranging from $<0.1 \mu g/L$ to $<1 \mu g/L$) [7-9, 29-31]. Regarding the primary effectiveness parameter, a statistically significant reduction in antibiotic days with a mean of 1.71 days ($95\% CI = -2.67, -0.74$) was found [7-9, 29-31]. Implementation of PCT testing showed a trend towards a decrease in overall hospital length of stay of on average 3.34 days ($95\% CI = -9.38, 2.69$) [9, 29-31] as well as a decrease in the duration of ICU stay of 1.08 days ($95\% CI = -3.52, 1.36$) [7-9, 29-31]. Therefore, a decreased length of stay on the general ward of $3.34 - 1.08 = 2.26$ days is expected.

Two studies reported the effect of the PCT algorithm on the duration of mechanical ventilation [30, 31], with a weighted mean increase of 0.71 days ($95\% CI = -1.00, 2.42$) per 12.0 days. Combined with the baseline number of 10.0 mechanical ventilation days as reported by

Adrie et al. [17], the PCT strategy is associated with a weighted increase of 0.59 mechanical ventilation days. None of the included studies reported an effect of a PCT algorithm on the duration of dialysis. A summary of the parameter inputs based on the systematic review, as well as the range applied in the one-way sensitivity analysis (mean $\pm$ 1 SD) is provided in Table 3. The forest plots of these parameters are shown in Appendix 1.

A PCT algorithm to guide antibiotic discontinuation is expected to reduce direct hospital costs per adult ICU patient with sepsis from €37,917 to €34,414, a decrease of €3,503 (-9.2%). This cost reduction is achieved with a 1.7 day reduction in duration of antibiotic use, i.e. from 11.6 to 9.9 days, and this translates into an incremental cost saving of €2043 per antibiotic day avoided. On a national level it is estimated that ~13,000 adult ICU patients in the Netherlands are diagnosed with sepsis each year, indicating a potential cost saving of almost 46 million Euros per year [32]. An overview of those results is shown in Table 4.

Of the €3,503 cost savings per patient, a decrease of €3,132 is attributable to the reduced hospital length of stay. In the conservative scenario where the reduction of length of stay is fully ignored, as is the accompanying decrease in laboratory tests that are assumed to be performed once-daily (i.e. savings of €82), overall direct hospital costs are still estimated to decrease with €289 per patient (i.e. €3,503 – €3,132 – €82 = €289). Those cost savings are mainly achieved by a decrease in the number of blood cultures performed and the duration of antibiotic therapy.

Table 3. Result of systematic literature review, showing the values used in the model, and the values applied in the one-way sensitivity analysis ($\pm$ 1 SD) in brackets. The right column shows the references used.

Parameter	Value		Reference
	Without PCT	With PCT	
General ward days	11.3	9.0 (6.0–12.0)	Nobre et al. [9], Bouadma et al. [30], Annane et al. [31], Deliberato et al. [29]
ICU days	13.8	12.7 (11.5–13.9)	Nobre et al. [9], Hochreiter et al. [8], Schroeder et al. [7], Bouadma et al. [30], Annane et al. [31], Deliberato et al. [29]
Days on antibiotics	11.6	9.9 (9.4–10.3)	Nobre et al. [9], Hochreiter et al. [8], Schroeder et al. [7], Bouadma et al. [30], Annane et al. [31], Deliberato et al. [29]
Days on mechanical ventilation	10.0	10.6 (9.9–11.6)	Annane et al. [31], Bouadma et al. [30], Adrie et al. [17]

Table 4. Model results. The costs per patient with sepsis at the ICU, split up for each of the aspects of the treatment. Overall costs are shown both per patient and for the estimated yearly number of ICU patients with sepsis in the Netherlands ($n = 13,000$) [32]. Numbers may not add up due to rounding.

Parameter	Value		Difference
	Without PCT	With PCT	
Hospital stay	€31,214	€28,083	€-3,132
General ward	€5,666	€4,555	€-1,112
ICU admission and stay	€25,548	€23,528	€-2,020
Treatment	€4,672	€4,637	€-35
Antibiotics	€1,465	€1,248	€-218
Mechanical ventilation	€2,974	€3,157	€182
Dialysis therapy	€232	€232	€0
Laboratory analyses	€2,030	€1,694	€-336
Blood cultures	€1,392	€1,063	€-329
PCT	€0	€75	€75
Other laboratory tests	€638	€556	€-82
Total costs per patient	€37,917	€34,414	€-3,503
Total costs in the Netherlands ($n= 13\ 000$)	€492,916,869	€447,379,610	€-45,537,259

Figure 2. Scatterplot of the results of Probabilistic Sensitivity Analysis (PSA), showing the result of 10,000 simulations and the average value.

The decision uncertainty surrounding the incremental cost-effectiveness ratio (ICER; point estimate €2,043 per antibiotic day avoided) as depicted in Figure 2, is low, as 82% (i.e. 8,167/10,000) of Monte Carlo simulations indicate that PCT reduces the duration of antibiotic treatment while saving costs compared to current practice. Only 18% (1,771/10,000) of the simulations suggest that PCT would be more expensive while being more effective in reducing the number of antibiotic days.

One-way sensitivity analysis was performed to estimate the impact of changes in individual input parameters on the difference in costs of the PCT strategy compared to current practice. Results are summarised in a tornado diagram (Figure 3). This figure illustrates that the net change in ICU and general ward days have the largest impact on costs. An overview of all results of the one-way sensitivity analysis is provided in Appendix 2.

DISCUSSION

This study shows that the upfront investments in PCT testing should not be considered in isolation, but as part of the whole pathway of care a patient receives. Our results indicate that, although PCT requires additional investments, those are more than offset against downstream cost savings due to a reduced duration of hospital stay and accompanying treatment. PCT to guide antibiotic discontinuation in adult ICU patients with sepsis is expected to reduce the number of antibiotic days and save costs without compromising patient outcomes. As such, this study adds new insights to the very recent evidence base regarding cost-effectiveness of PCT testing in different patient populations and settings. For example, Harrison and Collins found that the use of a PCT guided treatment algorithm dominated current practice with improved quality-of-life and decreased overall treatment costs in a US cohort of adult ICU patients with suspected bacterial infection and sepsis [14]. Notably, their analysis does not take the impact of PCT testing on hospital length of stay into account, which our study showed to be a very important driver of cost savings.

A couple of aspects of our analysis warrant further attention. First, none of the RCTs included showed a difference in in-hospital mortality for the PCT algorithm compared to current practice. Actually, as the pooled estimate of the five RCTs reporting in-hospital mortality shows a trend towards decreased in-hospital mortality in the PCT group (odds ratio = 0.83), our model is conservative for using an equal mortality rate. Although one of the RCTs reported an absolute increase in 60-day mortality of 3.8% in the PCT group, which may potentially question the safety of that PCT algorithm, the authors of this study state that ‘no patient in either group who died during days 29-60 had an infection relapse, and most deaths resulted from complications directly related to the severity of underlying disease’ (p. 472). Also, after controlling for potential confounders, the odds ratio for death by day 60 was not significantly different between the study groups [30].

Figure 3. Tornado diagram showing the effect of varying input parameters on model outcome. The lower and upper limits used in the sensitivity analysis for each parameter are shown in the grey boxes. The parameters derived from the systematic review (duration of ICU stay, general ward stay, mechanical ventilation, and antibiotic therapy) are varied with one SD below and above the mean. For parameters that shown an effect in the PCT strategy compared to current practice, the impact of changing this effect with 50% is shown (percentage of patients in whom a blood culture is performed). In all input parameters for which no effect was found due to the PCT strategy, a variation of 25% was used as lower and upper limit.

Second, cost savings associated with a PCT algorithm were estimated to be €3,503 per patient (-9.2%), while avoiding 1.7 antibiotic days (-14.8%), which indicates that the PCT algorithm dominates current practice. Those cost savings are achieved by a reduction in ICU and general ward length of stay, a reduction in the number of blood cultures performed, and, importantly, a reduction in the duration of antibiotic treatment. Although the results of the systematic review indicate a small (and non-significant) increase in mechanical ventilation days, this evidence is only based on two studies. Because one of these studies is very small, this increase cannot be considered conclusive. In addition, a retrospective study by Hohn et al. [33] shows a decrease in the duration of mechanical ventilation. This clearly indicates that further research about the effect of a PCT algorithm on the duration of mechanical ventilation is necessary.

A conservative re-analysis of the model, assuming that the PCT strategy does not reduce ICU and general ward length of stay, further confirms the conclusion that the PCT algorithm to discontinue antibiotic treatment is cost-saving. Compared to current practice, €289 per ICU patient with sepsis are saved (compared to the €3,503 in the base case scenario). These remaining cost savings are mainly attributable to the reduced duration of antibiotic treatment and the reduced number of blood cultures performed. As blood cultures are an important tool to confirm sepsis and because the amount of blood cultures required to do so is assumed to be affected by the PCT test, costs for all blood cultures performed in relation to sepsis were included to give the best estimate of the actual impact of PCT. An ad hoc two-way sensitivity analysis (data not shown) shows that, in fact, when the number of ICU days is kept equal between the strategies, the number of regular ward days may increase with 1 day in the PCT strategy, before this strategy becomes more costly than current practice. Clearly, when the cost-savings due to the length of stay reduction (i.e., €3,132) would not accrue, the decision uncertainty surrounding the model outcomes will increase. The probabilistic sensitivity analysis performed under this conservative scenario shows a 53% probability that the PCT algorithm dominates current practice by saving €289 per patient (compared to 82% under the base case scenario), while there is a 46% probability that the PCT strategy is more expensive while reducing the number of antibiotic days.

Third, the reduction in number of antibiotic days that can be achieved by a PCT algorithm for antibiotic discontinuation is highly important, not only for its impact on total direct healthcare costs, but notably so given the rise in antibiotic resistance. Indeed, prolonged antibiotic duration impacts the incidence of antibiotic resistance and *Clostridium difficile* infections, which in this population amounts to 4.7% and 4.6% per hospital episode, respectively [34-36]. While not the focus of the current paper, one could make a rough estimation of the additional indirect cost savings of PCT testing by considering the excess length of stay due to antibiotic resistance and *C. difficile*, reportedly circa 4.6 days and 0.9 days per patient, respectively [37-39]. As shorter duration of antibiotic therapy is shown to decrease the incidence of antibiotic resistance and *C. difficile* infections to 4.5% and 3.9%, respectively [36, 40, 41], including those

costs in the analyses increases per-patient costs in both strategies, but less so for the PCT algorithm (€39,028 in current practice compared to €35,235 in the PCT strategy). As a result, incremental cost savings of a PCT algorithm compared to current practice increase with €290 (from €3,503 to €3,793) per patient.

Fourth, adherence to the PCT algorithm is shown to affect its cost-effectiveness. For example, Harrison and Collins [14] showed that adherence of at least 42.3% was needed to render their specific PCT testing strategy cost-effective. Although we did not perform a sensitivity analysis on adherence rate, the data used in our model do reflect sub-optimal adherence to some extent, as the studies of Bouadma et al. [30] and Annane et al. [31] report that adherence in their studies was low. Explicit consideration of adherence is recommended for further work in this area.

5

The results of this analysis are in line with other reviews such as Tang et al. [12], Agarwal and Schwartz [11], and Heyland et al. [42], who conclude that PCT guided antibiotic therapy is associated with a reduction in antibiotic usage that may reduce overall costs of care, under certain assumptions. The latter is important for transferring results to a specific country, as one has to consider to what extent the current model assumptions and inputs are representative for that country. Because relative treatment effects are typically more transferable between developed Western countries than costs are (due to large differences in resource use and unit costs), ideally the model should be populated with country-specific data as much as possible to make valid per country estimations. For example, the duration of hospital stay and the duration of antibiotic treatment is relatively short in the Netherlands compared to other European countries [43], thus more benefit might be expected of the PCT assay in other European countries. Analyses for Germany and the UK are currently ongoing, but preliminary results suggest that the conclusions are robust across the different countries.

Furthermore, the reduction in the duration of antibiotic use as found in our review (1.7 days) can be considered as consistent yet conservative compared to other studies. Heyland et al. [42] reported a weighted average decrease in antibiotic duration of 2.14 days, and a study by Wilke et al. [13] reports an average reduction of 4 days. Interestingly, the expected cost savings as reported in our study are higher than those reported by Wilke et al., which can, amongst others, be explained by the fact that Wilke et al. considered the effect of PCT on ICU length of stay and on the duration of antibiotic therapy, while our study considered the effect of PCT on the entire hospitalization episode. Also, in the study by Wilke et al., the costs were derived from the German DRG calculation and applied to a real-life patient population, which might explain the differences in reported cost savings. Despite differences in the magnitude of cost savings, both studies suggest that substantial cost savings can be achieved following PCT implementation [13, 42].

This study considered the use of PCT as a biomarker for antibiotic discontinuation, not as a biomarker that guides initiation of antibiotic therapy. Although other studies, notably the one of Saeed et al. [44], report that PCT testing can support the decision of whether or not to start using antibiotics in situations where there is a clinical suspicion of infection, we conservatively did not consider this option in our model. Interviews with intensivists and clinical chemists in the Netherlands revealed that PCT is unlikely to be accepted as a marker to decide on starting antibiotics, because of the rapid increase in mortality associated with delayed antibiotic therapy in sepsis patients. However, it seems reasonable to argue that PCT might support the decision to withhold antibiotics if a sub-group of patients can be identified that only have a minor suspicion of sepsis. If possible, this may further improve the added value of the PCT test as a means for fighting antibiotic resistance.

An additional recommendation for further cost-effectiveness studies in this field is to consider the costs of implementing a PCT algorithm. Although the test is available for most routinely used laboratory analysers, other additional resources may be needed to implement a PCT algorithm (e.g., costs of educating laboratory staff to perform the test), which will affect its incremental cost-effectiveness, particularly in the early stages of implementation.

CONCLUSIONS

Proven safe, we conclude that a PCT algorithm for antibiotic discontinuation is a cost-effective testing strategy in adult intensive care patients with sepsis compared to current practice. The PCT strategy as studied in this analysis effectively reduces the duration of antibiotic therapy, while the cost of testing is more than recouped by downstream cost-savings that accrue from shorter hospital length of stay, shorter duration of antibiotic therapy, and reduced number of blood cultures. Further research is needed to explore the potential impact of PCT algorithms on reducing antibiotic resistance.

TRANSPARENCY

Declaration of funding

PANAXEA b.v. received funding from BRAHMS GmbH, part of Thermo Fisher Scientific, Hennigsdorf, Germany, for developing and analysing the model. The funding party has reviewed draft and final versions of the manuscript. The views expressed in this paper are solely the authors'.

Declaration of financial/other relationships

LMGS and MJIJ report stock ownership of Panaxea bv, Enschede, The Netherlands, which received funding from BRAHMS GmbH, part of Thermo Fisher Scientific, Hennigsdorf, Germany.

Acknowledgements

The authors thank the clinical chemists and intensivists who participated in the interviews.

REFERENCES

- Dellinger, R.P., et al., Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med*, 2013. 41(2): p. 580-637.
- Marik, P.E., Don't miss the diagnosis of sepsis! *Crit Care*, 2014. 18(5): p. 529.
- Carlet, J., Rapid diagnostic methods in the detection of sepsis. *Infect Dis Clin North Am*, 1999. 13(2): p. 483-94, xi.
- Schuetz, P., et al., Procalcitonin algorithms for antibiotic therapy decisions: a systematic review of randomized controlled trials and recommendations for clinical algorithms. *Arch Intern Med*, 2011. 171(15): p. 1322-31.
- Wenzel, R.P. and M.B. Edmond, Managing antibiotic resistance. *N Engl J Med*, 2000. 343(26): p. 1961-3.
- Becker, K.L., R. Snider, and E.S. Nylen, Procalcitonin assay in systemic inflammation, infection, and sepsis: clinical utility and limitations. *Crit Care Med*, 2008. 36(3): p. 941-52.
- Schroeder, S., et al., Procalcitonin (PCT)-guided algorithm reduces length of antibiotic treatment in surgical intensive care patients with severe sepsis: results of a prospective randomized study. *Langenbecks Arch Surg*, 2009. 394(2): p. 221-6.
- Hochreiter, M., et al., Procalcitonin to guide duration of antibiotic therapy in intensive care patients: a randomized prospective controlled trial. *Crit Care*, 2009. 13(3): p. R83.
- Nobre, V., et al., Use of procalcitonin to shorten antibiotic treatment duration in septic patients: a randomized trial. *Am J Respir Crit Care Med*, 2008. 177(5): p. 498-505.
- Matthaiou, D.K., et al., An ESICM systematic review and meta-analysis of procalcitonin-guided antibiotic therapy algorithms in adult critically ill patients. *Intensive Care Med*, 2012. 38(6): p. 940-9.
- Agarwal, R. and D.N. Schwartz, Procalcitonin to guide duration of antimicrobial therapy in intensive care units: a systematic review. *Clin Infect Dis*, 2011. 53(4): p. 379-87.
- Tang, H., et al., Effect of procalcitonin-guided treatment in patients with infections: a systematic review and meta-analysis. *Infection*, 2009. 37(6): p. 497-507.
- Wilke, M.H., R.F. Grube, and K.F. Bodmann, The use of a standardized PCT-algorithm reduces costs in intensive care in septic patients - a DRG-based simulation model. *Eur J Med Res*, 2011. 16(12): p. 543-8.
- Harrison, M. and C.D. Collins, Is procalcitonin-guided antimicrobial use cost-effective in adult patients with suspected bacterial infection and sepsis? *Infect Control Hosp Epidemiol*, 2015. 36(3): p. 265-72.
- Hozo, S.P., B. Djulbegovic, and I. Hozo, Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*, 2005. 5: p. 13.
- The Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. 2011 Lasted updated March 2011 [cited 2012 June 12]; Available from: <http://handbook.cochrane.org/>.
- Adrie, C., et al., Epidemiology and economic evaluation of severe sepsis in France: age, severity, infection site, and place of acquisition (community, hospital, or intensive care unit) as determinants of workload and cost. *J Crit Care*, 2005. 20(1): p. 46-58.
- Shapiro, N.I., et al., Who needs a blood culture? A prospectively derived and validated prediction rule. *J Emerg Med*, 2008. 35(3): p. 255-64.
- Muller, F., et al., Procalcitonin levels predict bacteremia in patients with community-acquired pneumonia: a prospective cohort trial. *Chest*, 2010. 138(1): p. 121-9.
- Stocker, M., et al., Use of procalcitonin-guided decision-making to shorten antibiotic therapy in suspected neonatal early-onset sepsis: prospective randomized intervention trial. *Neonatology*, 2010. 97(2): p. 165-74.
- Schuetz, P., et al., Effect of procalcitonin-based guidelines vs standard guidelines on antibiotic use in lower respiratory tract infections: the ProHOSP randomized controlled trial. *JAMA*, 2009. 302(10): p. 1059-66.
- van Nieuwkoop, C., et al., Procalcitonin reflects bacteremia and bacterial load in urosepsis syndrome: a prospective observational study. *Crit Care*, 2010. 14(6): p. R206.
- Nederlandse Zorgautoriteit. Bijlage 1 bij tariefbeschikking TB/CU-7041-03. 2013 [cited 2013 November 26]; Available from: <http://www.nza.nl/98174/139255/654366/bijlagemap/TB-CU-7041-03-bijlage1.pdf>.
- DBC Onderhoud. Gebruikersdocument uitlevering deel 2: verantwoording tariefswijzigingen. 2011 [cited 2011 November 29]; 1-47]. Available from: http://www.nza.nl/104107/138040/bijlagemap/Bijlage_2_bij_release_RZ13a_-_Gebruikersdocument_uitlevering_deel_2.pdf.
- Hakkaart-van Roijen, L., S.S. Tan, and C.A.M. Bouwmans, Handleiding voor kostenonderzoek; Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. 2010, Institute for Medical Technology Assessment, Erasmus University Rotterdam. p. 1-127.
- Vandijck, D.M., et al., Daily cost of antimicrobial therapy in patients with Intensive Care Unit-acquired, laboratory-confirmed bloodstream infection. *Int J Antimicrob Agents*,

2008. 31(2): p. 161-5.
27. Consumentenprijzen;prijnsindex 2006 = 100. Centraal Bureau voor de Statistiek, 2014. Available at: <http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71311NED&D1=0&D2=0&D3=64,77,90,103,116,129,142,155,168,181,194,219,232,245&VW=T> [Last accessed 18 February 2014]
 28. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press, 2006
 29. Deliberato RO, Marra AR, Sanches PR, et al. Clinical and economic impact of procalcitonin to shorten antimicrobial therapy in septic patients with proven bacterial infection in an intensive care setting. *Diagn Microbiol Infect Dis* 2013;76:266-71
 30. Bouadma L, Luyt CE, Tubach F, et al. Use of procalcitonin to reduce patients' exposure to antibiotics in intensive care units (PRORATA trial): a multicentre randomised controlled trial. *Lancet* 2010;375:463-74
 31. Annane, D., et al., Procalcitonin levels to guide antibiotic therapy in adults with non-microbiologically proven apparent severe sepsis: a randomised controlled trial. *BMJ Open*, 2013. 3(2).
 32. VMS veiligheidsprogramma. Voorkomen van lijnsepsis en behandeling van ernstige sepsis. 2013 [cited 2014 August 8]; Available from: <http://www.vmszorg.nl/Themas/Sepsis>.
 33. Hohn, A., et al., Procalcitonin-guided algorithm to reduce length of antibiotic therapy in patients with severe sepsis and septic shock. *BMC Infect Dis*, 2013. 13: p. 158.
 34. European Centre for Disease Prevention and Control. Antimicrobial resistance interactive database. 2012; [cited 2014 May 25]; Available from: <http://www.ecdc.europa.eu/en/activities/surveillance/EARS-Net/Pages/index.aspx>.
 35. Vincent, J.L., et al., Sepsis in European intensive care units: results of the SOAP study. *Crit Care Med*, 2006. 34(2): p. 344-53.
 36. Stevens, V., et al., Cumulative antibiotic exposures over time and the risk of *Clostridium difficile* infection. *Clin Infect Dis*, 2011. 53(1): p. 42-8.
 37. Lye, D.C., et al., The impact of multidrug resistance in healthcare-associated and nosocomial Gram-negative bacteraemia on mortality and length of stay: cohort study. *Clin Microbiol Infect*, 2012. 18(5): p. 502-8.
 38. Mitchell, B.G., et al., The prolongation of length of stay because of *Clostridium difficile* infection. *Am J Infect Control*, 2014. 42(2): p. 164-7.
 39. Roberts, R.R., et al., Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. *Clin Infect Dis*, 2009. 49(8): p. 1175-84.
 40. Singh, N., et al., Short-course empiric antibiotic therapy for patients with pulmonary infiltrates in the intensive care unit. A proposed solution for indiscriminate antibiotic prescription. *Am J Respir Crit Care Med*, 2000. 162(2 Pt 1): p. 505-11.
 41. Chastre, J., et al., Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults: a randomized trial. *JAMA*, 2003. 290(19): p. 2588-98.
 42. Heyland, D.K., et al., Procalcitonin for reduced antibiotic exposure in the critical care setting: a systematic review and an economic evaluation. *Crit Care Med*, 2011. 39(7): p. 1792-9.
 43. Health at a Glance, Europ. OECDiLibrary, 2014; [cited 2015 June 5] ; Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-europe_23056088;jsessionid=1wladsmo5vt8l.x-oecd-live-02.
 44. Saeed, K., et al., Reduction in antibiotic use through procalcitonin testing in patients in the medical admission unit or intensive care unit with suspicion of infection. *J Hosp Infect*, 2011. 78(4): p. 289-92.

APPENDIX 1: RESULTS OF LITERATURE REVIEW

The following figures show the results of the literature review concerning the safety and effectiveness of a PCT algorithm for antibiotic discontinuation in adult ICU patients with sepsis. Pooled estimates for each parameter were obtained using Review Manager. A random effects model was applied. An overall estimate of the treatment effect is visualized in a forest plot for each parameter, including the 95% confidence interval.

Figure 1a. In-hospital mortality. This table shows the impact of a PCT algorithm for antibiotic discontinuation on in-hospital mortality.

Figure 1b. Hospital length of stay. Impact of a PCT algorithm for antibiotic discontinuation on hospital length of stay.

Figure 1c. ICU length of stay. Impact of a PCT algorithm for antibiotic discontinuation on ICU length of stay.

Figure 1d. Antibiotic use. Impact of a PCT algorithm for antibiotic discontinuation on duration of antibiotic treatment.

Figure 1e. Mechanical ventilation. Impact of a PCT algorithm for antibiotic discontinuation on duration of mechanical ventilation.

APPENDIX 2: RESULTS OF SENSITIVITY ANALYSIS

Table 1. One-way sensitivity analysis results. This table shows the result of the one-way sensitivity analysis. Both the base case input parameter as well as the lower and upper limit, and the effect on direct hospital costs are shown for each parameter. Results are based on deterministic cost-effectiveness analysis.

Parameter	Current model input with PCT	Lower and upper limit	Effect on costs
Base case		-	€ - 3,462
Effect on regular ward days	9.0	6.0 12.0	€ - 4,975 € - 1,951
Effect on ICU days	12.7	11.5 13.9	€ - 5,672 € - 1,251
Percentage receiving antibiotics	100.0%	75.0%	€ - 3,774
Effect on antibiotic days	9.9	9.4 10.3	€ - 3,523 € - 3,400
Effect on percentage mechanical ventilation	77.0%	57.8% 96.3%	€ - 4,249 € - 2,674
Effect on days mechanical ventilation	10.6	9.9 11.6	€ - 3,681 € - 3,172
Percentage of patients on dialysis therapy	16.0%	12.0% 20.0%	€ - 3,520 € - 3,404
Number of days on dialysis therapy	5.00	3.75 6.25	€ - 3,520 € - 3,404
Effect on percentage of patients in whom a blood culture is performed	61.4%	79.5% 43.4%	€ - 3,207 € - 3,717
Sets of blood cultures performed per patient	2.0	1.5 2.5	€ - 3,679 € - 3,244
Blood cultures performed in patients diagnosed as having sepsis	8.2%	6.1% 10.2%	€ - 3,111 € - 3,672
Number of times laboratory tests ordered	21.8	23.4 20.1	€ - 3,421 € - 3,502
Number of times PCT test performed	5.0	2.5 7.5	€ - 3,499 € - 3,424

6

Procalcitonin to guide antibiotic stewardship in intensive care

Michelle M.A. Kip
Hendrik Koffijberg
Maarten J. IJzerman
Ron Kusters

This chapter has been published as:

M.M.A. Kip, H. Koffijberg, M.J. IJzerman, R. Kusters. *Procalcitonin to guide antibiotic stewardship in intensive care*. Lancet Infect Dis, 2016. 16(8). p. 888.

Reprinted with permission.

We read with great interest the Article by Evelien de Jong and colleagues [1], showing that procalcitonin guidance reduces treatment duration and daily defined doses in critically ill patients with a presumed bacterial infection. Although de Jong and colleagues's data do not immediately allow a cost-effectiveness analysis, they can, however, be used to inform a health economic model. Moreover, such a model has already been published [2] and can thus be updated to immediately assess cost-effectiveness.

We used the reported difference in mean duration of intensive care unit and hospital stay, and in duration and costs of antibiotic treatment, between the procalcitonin group and the control group as input for the cost-effectiveness analysis (table). The corresponding means were estimated from the reported median values [3] (control group) and then combined with the reported difference in means (procalcitonin group). The mean number of procalcitonin measurements was set to seven [1]. Other laboratory analyses were assumed to be performed at intensive care admission, and once daily until hospital discharge or death. All other parameters were assumed equal in both groups. To reflect the combined uncertainty in model input parameters, probabilistic sensitivity analysis was performed using Monte Carlo simulation (10,000 samples). Previously, length of intensive care unit stay was found to have the biggest effect on costs [2]. Therefore, we performed a threshold analysis to determine the decrease in intensive care unit days required to make procalcitonin cost saving. Costs were expressed in euro (2015 value) and based on the recent Dutch costing manual [4].

The procalcitonin strategy costs on average €27,777 per patient, compared with €27,556 per patient in the control group (ie, an increase of 0.80%; difference €221, 95% CI -1,724 to 2,164), and has a probability of 41% to be cost saving. However, the threshold analysis revealed that if the procalcitonin strategy would (for example) decrease intensive care unit stay from 10.0 to 9.0 days, the difference becomes a decrease of €2,231 (95% CI -4,152 to -293), with a 99% probability to be cost saving. The table shows the input parameters and model outcomes.

Our preliminary analysis does not conclusively demonstrate that the procalcitonin assay is cost saving. Because length of intensive care stay is the main cost driver, further research should focus on estimating this parameter more accurately. Finally, a formal cost-effectiveness analysis would also require incorporating evidence on health outcomes such as the effect of procalcitonin on decreasing mortality rates. Iteratively updating a health economic model with new evidence to update cost-effectiveness estimates is straightforward.

We declare no competing interests.

Table 1. Overview of input parameters and outcomes of the health economic model, in both the control group and the procalcitonin (PCT) group.

	Control group	PCT group	Control group mean (95% CI)	PCT group mean (95% CI)	Difference mean (95% CI)
Antibiotic treatment, days	7.33	6.11	€156 (148 to 164)	€122 (116 to 129)	€-34 (-44 to -24)
Intensive care unit stay, days	10.00	10.21	€19,841 (18,509 to 21,203)	€20,285 (19,050 to 21,546)	€444 (-1,361 to 2,273)
Regular ward stay, days	14.67	14.07	€7,028 (6,668 to 7,395)	€6,740 (6,254 to 7,239)	€-288 (-898 to 331)
Procalcitonin measurements	0	7	-	€105 (86 to 126)	€105 (86 to 126)
Other laboratory tests, number of days	24.67	24.28	€531 (445 to 622)	€525 (449 to 613)	€-7 (-125 to 113)
Overall costs per patient			€27,556 (26,182 to 28,969)	€27,777 (26,457 to 29,154)	€221 (-1,724 to 2,164)
Overall yearly costs in the Netherlands			€358,000,000 (340,000,000 to 377,000,000)	€361,000,000 (344,000,000 to 379,000,000)	€3,000,000 (-22,000,000 to 28,000,000)

REFERENCES:

1. de Jong, E., et al. Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: a randomised, controlled, open-label trial. *The Lancet Infect Dis*, 2016.

3. Wan, X. et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*, 2014. 14: p. 135.

2. Kip, M. M., et al. A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis. *J Med Econ*, 2015. 18(11): p. 944-53.

4. Hakkaart-van Roijen, L., et al., *Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg*. 2015, Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam: Rotterdam. p. 1-120.

7

Cost-effectiveness of procalcitonin testing to guide antibiotic treatment duration in critically ill patients: results from a randomized controlled multicenter trial

Michelle M.A. Kip
Jos A. van Oers
Arezoo Shajiei
Albertus Beishuizen
A.M. Sofie Berghuis
Armand R. Girbes

Evelien de Jong
Dylan W. de Lange
Maarten W.N. Nijsten
Maarten J. IJzerman
Hendrik Koffijberg
Ron Kusters

Submitted

ABSTRACT

Background

Procalcitonin (PCT) testing can help in safely reducing antibiotic treatment duration in intensive care patients with sepsis. However, the cost-effectiveness of this PCT-guidance is not yet known.

Methods

A trial-based analysis was performed to estimate the cost-effectiveness of PCT-guidance compared with standard-of-care (without PCT-guidance). Patient-level data from the SAPS-trial, which randomized 1,546 patients, were used. Outcome measures were (1) incremental direct hospital cost and (2) incremental cost per quality-adjusted life year (QALY) gained from a healthcare perspective over a one-year time horizon. Uncertainty in outcomes was assessed with bootstrapping.

Findings

Mean in-hospital costs are €46,146/patient in the standard-of-care group, compared with €46,081/patient in the PCT group (i.e. €-65; -0.1%), while the duration of the first course of antibiotic treatment decreased from 8.2 to 6.9 days (i.e. -1.2 days, -14.8%). Over a one-year time horizon, mean healthcare cost increased from €70,961/patient to €73,665/patient (i.e. +€2,704, +3.8%) and mean QALYs increased from 0.47/patient to 0.52/patient (i.e. +0.05; +10.1%), owing to a decreased mortality in the PCT group. This resulted in an incremental cost-effectiveness ratio of €57,402/QALY gained.

Interpretation

The impact of PCT-guidance on costs during the initial hospitalization episode is minimal, whereas (as previously reported) in-hospital mortality and antibiotic treatment duration are significantly reduced. The decrease in mortality led to an increase in healthcare-related costs, due to high costs among sepsis survivors. PCT-guidance cannot be considered cost-effective within a one-year time frame at a willingness-to-pay threshold of €20,000/QALY. Longer follow-up studies are required to quantify the cost-effectiveness of this PCT-guidance from a long-term, societal perspective.

INTRODUCTION

Procalcitonin (PCT) is a biomarker that can be used in addition to traditional markers (i.e. C-reactive protein [CRP]) for diagnosing and monitoring patients with bacterial infections [1]. Findings from several randomized controlled trials (RCTs) indicate that the use of a PCT-guided antibiotic treatment algorithm (i.e. PCT guidance) is likely to contribute to a reduction of antibiotic exposure in septic patients in the intensive care unit (ICU) [2-10], without an adverse effect on health outcomes [2,3,5-7,9,10]. In addition, a recently published Cochrane review concluded that the use of PCT-guided initiation and duration of antibiotic treatment results in a decreased mortality risk, lower antibiotic consumption, and (consequently) a lower risk for antibiotic-related side effects in patients with acute respiratory tract infections [11]. Yet, the considerable costs of PCT testing compared to other laboratory assays (i.e. CRP) remains an important barrier to broader implementation. This barrier may (partly) result from limited insight in the consequences of this PCT algorithm, as both costs and health outcomes that occur along the diagnostic and treatment pathway have not been analysed in depth. Importantly, previous modelling studies into PCT-guidance have suggested that it has the potential to save costs [12-14]. However, those studies were all based on a hypothetical patient population instead of real-life patient outcome data, and proved to strongly depend upon the input parameters used. In addition, as results from cost-effectiveness studies may not be readily transferable between countries, it is uncertain to what extent the results from those previous studies can be applied to the Netherlands. In particular, as antibiotic prescription and the duration of hospitalization are relatively low in the Netherlands compared to other developed countries [15,16], this will limit the potential impact of PCT-guidance.

Recently, the Stop Antibiotics on Procalcitonin guidance Study (SAPS) was performed in 1,546 patients in 16 hospitals in the Netherlands to investigate the efficacy and safety of PCT-guidance in adult ICU patients with sepsis [10]. This trial found that PCT-guidance reduces the median duration of antibiotic treatment from 7 to 5 days, and also found – unexpectedly – that this strategy may improve survival. PCT-guidance had no impact on the ICU or hospital length of stay. This trial provided extensive real-life data suitable for a cost-effectiveness analysis that allows accurate, patient-level estimates of both costs and health outcomes with respect to PCT use in septic ICU patients. Therefore, we used the SAPS data as the input for a trial-based cost-effectiveness analysis in a Dutch setting, with (on average) low antibiotic use and a short duration of hospital stay [15,16].

METHODS

Study design and patients

The patient data was derived from the SAPS [10]. This was a prospective, multicentre, randomised, open-label intervention trial among patients admitted to the ICU of four university medical centres and 12 teaching hospitals in the Netherlands. The medical ethical committee

of the VU University Medical Centre (Amsterdam, the Netherlands) approved this study for all participating centres. ICU patients were eligible for inclusion if they were aged ≥ 18 years and received their first dose of antibiotics for a presumed or proven infection ≤ 24 h before trial inclusion. Patients could participate only once. Written informed consent was obtained from all patients or from their legal representatives. Patients either received treatment according to PCT-guidance (PCT group, $n = 761$) or standard-of-care (standard-of-care group, $n = 785$), based on random allocation in a 1:1 ratio. A full description of the study design, baseline patient characteristics, and study findings has been published previously [10,17].

In the standard-of-care group PCT was not measured. In the PCT group, PCT was measured daily until ICU discharge or until 3 days after systemic antibiotics were stopped, and the results were made available to the attending physician. In addition, a baseline PCT measurement was performed as close to initiation of antibiotics as possible (at least within 24h), and also made available to the attending physician. The study protocol advised to discontinue antibiotics if the PCT concentration had decreased by 80% or more of its peak value, or when it reached a value of 0.5 $\mu\text{g/L}$ or lower. It was left to the discretion of the attending physician whether to adhere to this stopping advice. In the standard-of-care group, antibiotics were stopped according to local or national guidelines, and this was also left to the discretion of attending physicians. Except for PCT measurements, all monitoring was similar between the PCT and the standard-of-care group. There was no loss to follow-up. All randomized patients were included in this health economic evaluation.

Outcome measures and model design

A trial-based analysis was performed to estimate the cost-effectiveness of the use of PCT-guidance compared with standard-of-care. The main outcome measures were defined as the impact on the duration of antibiotic treatment, in-hospital mortality, and healthcare costs. These main outcome measures reflect the actually recorded data (during the SAPS trial) on resource use and patient outcomes, covering the period from sepsis diagnosis and initiation of systemic antibiotic treatment until hospital discharge. In addition, the secondary outcome measure was defined as the incremental cost per quality-adjusted life year (QALY) gained. This outcome measure is referred to as the incremental cost-utility ratio (ICUR). As the SAPS did not collect data on quality of life (i.e. utilities) among sepsis survivors, these were derived from a Dutch follow-up study [28], and combined with one-year mortality rates from the SAPS study. Short-form 36 (SF-36) scores measured at ICU discharge, hospital discharge, and 3 and 6 months after ICU discharge were converted to mean EuroQol-5D values, and extrapolated to one-year after ICU discharge. An overview of the utilities used (and accompanying references) is provided in Table 1 of Appendix 1. As these utility values (as well as costs among sepsis survivors which are described in the next paragraph) were based on literature and thus involved additional assumptions, determination of the ICUR was considered an additional scenario.

In addition, as the SAPS did not collect data on productivity losses among sepsis patients, incorporating societal costs within this analysis would have required making many assumptions, which would have strongly increased the overall uncertainty in the results. However, as current guidelines for health economic evaluations recommend the use of a societal perspective [18], this analysis was performed as a separate scenario. This analysis included costs of lost productivity (estimating using the friction cost approach) due to the estimated recovery period after hospitalization. This period was estimated at 12 weeks, which is likely a strong underestimation [19]. In addition, these societal costs included productivity losses due to premature mortality and the impact of PCT guidance on QALYs.

Model inputs - Resource use

Data regarding the duration of stay on the ICU and the general ward, antibiotic prescriptions (including the type, duration, dose, and mode of administration) and other treatments administered (i.e. mechanical ventilation, dialysis, selective digestive decontamination [SDD], and selective oral decontamination [SOD]), were derived from the SAPS database. For the scenario that incorporates the estimated costs of productivity losses, data on mean working hours were derived from Statistics Netherlands.

Per patient costs of diagnostic testing were calculated based on the frequency and type of blood cultures, other cultures, PCT tests, and routine laboratory tests (i.e. C-reactive protein [CRP], bilirubin, creatinine, leukocyte, and thrombocyte testing) performed. The frequency of each, during the first 28 days of hospitalization, was obtained from the SAPS database. A detailed overview of all resource use parameters, references, and assumptions used is provided in Appendix 1.

Model inputs – Costs

Tariffs of laboratory tests, blood cultures, and other cultures were derived from the Dutch Healthcare Authority (Nederlandse Zorgautoriteit). Tariffs of hospital stay on the ICU and general ward (which also includes costs of hospital staff and overhead), as well as costs of mechanical ventilation and dialysis, were obtained from the Dutch Healthcare Authority. Costs of antibiotic therapy were calculated by combining data from individual drug administration records from the SAPS database with their accompanying unit costs (as published previously). All healthcare-related costs up to one year after ICU admission were derived from literature [25]. To ensure that these costs involved only healthcare costs after the initial hospitalization period, the costs after 6 months of sepsis onset were used for the purpose of the current analysis. All costs were converted to 2017 Euros, using Dutch consumer price index levels [20]. The friction cost method was applied to calculate productivity losses, using average, gender-specific Dutch wage rates. In the current analysis, a price of €31.71 per PCT measurement was used. A detailed overview of all cost parameters, references, and assumptions used is provided in Appendix 1.

Analysis

The data were analysed using R (version 3.4.1) [21]. Multiple imputation was used to handle missing values, as this technique is known to yield more valid results than complete case analysis [22]. This multiple imputation step was performed using the package “mice” (version 2.46) and 10 imputation sets [23]. In order to reflect the uncertainty in cost-effectiveness outcomes, the potential variance within the sample data should be considered. In this regard, bootstrapping was used to incorporate this uncertainty and to obtain the accompanying confidence intervals [24]. This bootstrap consisted of two stages: (1) random sampling (selection) of individual hospitals from the 16 participating hospitals in the SAPS study 10, (2) random sampling (selection) of patients from the hospitals selected in step 1. This first step was repeated until the minimum intended sample size of 631 patients 10 was reached in both the PCT and the control group. This two-stage bootstrapping procedure was repeated until 10,000 samples (i.e. replicated databases) had been generated.

Results of both the primary and secondary outcome measure were presented in a cost-effectiveness plane. In addition, to evaluate the probability that the PCT strategy is cost-effective, the cost-utility analysis (the secondary outcome measure) was presented in a cost-effectiveness acceptability curve (CEAC). Willingness-to-pay (WTP) thresholds ranging from €0/QALY to €100,000/QALY were used.

Figure 1. Incremental cost-effectiveness plane showing the impact of the use of a PCT-guided antibiotic treatment algorithm, as compared to standard-of-care, on the difference in in-hospital mortality and accompanying costs within this (initial) hospitalization episode. The result is based on 10,000 bootstrap samples.

As data regarding costs and utilities were not collected as part of the SAPS (except for costs of antibiotic treatment), these were based on national tariffs and published literature as described previously. To reflect the uncertainty in these mean cost and utility estimates, a scenario was analysed in which the impact of varying all cost and utility parameters in the cost-effectiveness analysis was investigated using a normal distribution (for costs) and a beta distribution (for utilities). For costs among sepsis survivors, the standard error was based on published literature [25]. All other parameters were varied assuming a standard error of 10% [26].

Figure 2. Incremental cost-effectiveness plane showing the impact of the use of a PCT-guided antibiotic treatment algorithm, as compared to standard-of-care, on the difference in QALYs (until one year after ICU admission) and accompanying healthcare-related costs within this one-year time period. In addition, the willingness-to-pay thresholds of €20,000/QALY and €80,000/QALY are shown. The result is based on 10,000 bootstrap samples. QALY = quality-adjusted life year; WTP = willingness-to-pay.

RESULTS

The results of this trial-based cost-effectiveness analysis indicate that the expected in-hospital costs per patient are €46,081/patient in the PCT group, compared with €46,146/patient in the standard-of-care group, indicating an average decrease of €65/patient (95% CI: -€6,314 to €6,107, a relative decrease of 0.14%). The average in-hospital mortality was found to be 21.8% in the PCT group, compared with 29.8% in the standard-of-care group, indicating a decrease of 7.9% (95% CI: -13.9% to -1.8%, i.e. a relative decrease of 26.6%). In addition, the duration of antibiotic use was found to be 6.9 days in the PCT group, compared with 8.2 days with standard-of-care, i.e. -1.2 days (95% CI: -1.9 to -0.4, i.e. -14.8%). When incorporating healthcare-related costs up to one year after ICU admission, the average per patient costs increase to €73,665 per patient in the PCT group, compared with €70,961 in standard-of-care, i.e. +€2,704

(95% CI: -€4,495 to €10,005, i.e. +3.8%). When considering the impact on QALYs, results indicate an average of 0.52 QALYs in the PCT group, compared with 0.47 QALYs in standard-of-care, i.e. +0.05 QALYs (95% CI: 0.00 to 0.10, i.e. +10.1%), resulting in an ICUR of €57,402/QALY gained. Assuming that annually ~13,000 adult ICU patients in the Netherlands are diagnosed as having sepsis [27], the use of this PCT-guided treatment algorithm could result in an annual increase in total healthcare related costs (up to one year after ICU admission) of approximately €35M, while saving 612 QALYs.

A detailed overview of the outcomes of the analyses (including the 95% CIs) is provided in Table 1. The incremental cost-effectiveness plane for the primary outcome measure (i.e. costs per in-hospital death avoided) is shown in Figure 1, whereas the incremental cost-effectiveness plane for the secondary outcome measure (i.e. costs/QALY one-year after ICU admission) is shown in Figure 2.

Table 1. Overview of model outcomes. This table shows an overview of the model outcomes in terms of effectiveness and costs. The mean model outcomes (and accompanying 95% CIs) are shown for the PCT group and the standard-of-care group. In addition, the difference between these groups (and accompanying 95% CIs) are provided.

Type of parameter	Parameter	PCT, mean (95% CI)	Standard-of-care, mean (95% CI)	Effect, mean (95% CI)
Effectiveness outcomes				
Hospital stay	ICU stay, days (n)	14.6 (12.4 to 16.9)	14.3 (12.5 to 16.1)	0.3 (-2.2 to 2.8)
	General ward stay, days (n)	16.9 (13.9 to 19.6)	17.6 (14.5 to 20.8)	-0.4 (-4.3 to 3.5)
Organ support	Mechanical ventilation, days (n)	4.7 (4.0 to 5.6)	5.4 (4.5 to 6.3)	-0.6 (-1.3 to 0.0)
	Renal replacement therapy, days (n)	0.8 (0.5 to 1.1)	0.9 (0.5 to 1.3)	-0.1 (-0.5 to 0.2)
Medication	Antibiotics, days (n)	6.9 (5.6 to 8.5)	8.2 (7.0 to 9.5)	-1.2 (-1.9 to -0.4)
	SDD and SOD (n)	4.0 (0.7 to 8.2)	4.9 (0.9 to 9.9)	-0.9 (-2.1 to -0.1)
Laboratory tests	Cultures (n)	4.2 (2.8 to 5.9)	4.8 (3.2 to 6.5)	-0.5 (-1.4 to 0.2)
	PCT (n)	6.4 (5.7 to 7.3)	0.0 (0.0 to 0.0)	6.4 (5.7 to 7.3)
	Other tests (including order tariff) (n)	14.3 (11.3 to 17.3)	14.7 (11.8 to 17.7)	-0.4 (-2.8 to 2.1)
In-hospital mortality		21.8% (17.1% to 26.4%)	29.8% (23.5% to 36.4%)	-7.9% (-13.9% to -1.8%)

QALYs		0.52 (0.49 to 0.54)	0.47 (0.43 to 0.51)	0.05 (0.00 to 0.10)
Cost outcomes				
Hospital stay	ICU stay	€32,908 (€28,109 to €38,131)	€32,390 (€28,080 to €36,673)	€519 (-€5,227 to €6,118)
	General ward stay	€9,594 (€6,218 to €12,669)	€9,972 (€6,331 to €13,525)	-€378 (-€2,206 to €1,300)
Organ support	Mechanical ventilation	€1,991 (€1,667 to €2,369)	€2,259 (€1,867 to €2,671)	-€268 (-€555 to €9)
	Renal replacement therapy	€362 (€244 to €500)	€408 (€247 to €592)	-€46 (-€217 to €116)
Medication	Antibiotics	€203 (€131 to €283)	€237 (€168 to €317)	-€35 (-€73 to €6)
	SDD and SOD	€127 (€32 to €226)	€157 (€42 to €267)	-€30 (-€64 to -€1)
Laboratory tests	Cultures	€109 (€72 to €151)	€122 (€81 to €165)	-€13 (-€34 to €6)
	PCT	€204 (€181 to €232)	€0 (€0 to €0)	€204 (€181 to €232)
	Other tests (incl. order tariff)	€584 (€484 to €677)	€602 (€502 to €707)	-€19 (-€96 to €56)
Total hospital costs		€46,081 (€38,242 to €54,120)	€46,146 (€39,383 to €53,042)	-€65 (-€6,314 to €6,107)
Healthcare costs (follow-up)		€27,585 (€26,031 to €29,261)	€24,815 (€22,311 to €27,056)	€2,770 (€136 to €5,550)
Total healthcare costs (up to one year follow-up)		€73,665 (€66,065 to €81,344)	€70,961 (€64,776 to €77,082)	€2,704 (-€4,495 to €10,005)
Lost productivity		€6,982 (€6,582 to €7,370)	€6,923 (€6,570 to €7,276)	€59 (-€364 to €485)
Total societal costs (up to one year follow-up)		€80,647 (€72,918 to €88,401)	€77,884 (€71,604 to €84,116)	€2,763 (-€4,491 to €10,172)

Table 1 indicates that, when only considering the initial hospitalization episode, the main increase in costs in the PCT group as compared with standard-of-care is attributable to an increase in costs of ICU stay as well as due to costs of PCT testing. The main cost savings are achieved through reduced costs of general ward stay and mechanical ventilation. However, when considering all healthcare-related costs up to one year after ICU admission, an increase in costs in the PCT group is expected. In addition, the scenario in which costs of lost productivity are included indicates that the costs in both the PCT and the standard-of-care group will increase with almost €7,000. However, the current results indicate an ICUR of €57,402/QALY gained when considering all healthcare-related costs within one year, and an ICUR of €58,648/QALY gained when also incorporating societal costs. Thus, these outcomes are very similar and

both would not be considered cost-effective at a WTP threshold of €20,000/QALY, whereas both would be considered cost-effective at a WTP threshold of €80,000/QALY.

Figure 3. Cost-effectiveness acceptability curve showing the probability that the use of a PCT-guided antibiotic treatment is cost-effective compared to standard-of-care, for a WTP threshold ranging from €0/QALY to €200,000/QALY. This analysis incorporates all healthcare-related costs over a one-year time horizon. WTP: willingness-to-pay; QALY: quality-adjusted life year.

The results of the CEAC show the probability that PCT-guidance (until one year after ICU discharge) is cost-effective. This probability was found to range from 23.5% to 95.7%, for accompanying WTP thresholds ranging from €0/QALY to €200,000/QALY. For a WTP of €20,000/QALY, which is the most commonly used WTP threshold in the Netherlands, this probability was 30.6% (Figure 3).

In addition, in Appendix 2, the model outcomes are shown for the scenario in which all utilities and costs were varied with their (assumed) standard error, including incremental cost-effectiveness planes and a CEAC.

DISCUSSION

Results indicate that the use of PCT-guidance is expected to involve a minor (and non-significant) decrease in healthcare costs (i.e. -0.14%) during the initial hospitalization episode. This is accompanied by a decrease in the duration of antibiotic use as well as a significant decrease of in-hospital mortality. However, when considering a one-year time horizon, total healthcare-related costs are estimated to increase with 3.8%, due to high healthcare expenditures in survivors of sepsis. This is accompanied by a 0.05 increase in QALYs, i.e. +10.1%

(until one year after ICU admission), resulting in an ICUR of €57,402/QALY gained. Incorporating cost of lost productivity will increase the ICUR to €58,648/QALY gained. Thus, in this context, health gains from reducing mortality will always be accompanied by an increase in healthcare costs, given the higher costs of survivors compared to non-survivors.

Strengths

The main strength of the current analysis is that the majority of model inputs was derived from a large multicenter RCT. This database had no missing values in the effectiveness parameters of interest (i.e. in-hospital mortality, and one-year mortality). In addition, the maximum number of missing values in the parameters used in the analysis concerned five out of 761 (i.e. 0.7%) and four out of 785 (i.e. 0.5%) in the parameter 'length of ICU stay', for the PCT and standard-of-care group respectively. Consequently, the impact of multiple imputation on the cost-effectiveness outcomes was negligible, allowing an accurate estimation of the cost-effectiveness of PCT-guidance. Furthermore, although (inevitably) some assumptions had to be made, it is very unlikely that these assumptions would have affected the conclusions drawn.

In addition, although the SAPS did not collect data on quality of life (i.e. utilities), or on healthcare-related costs up to one year after ICU discharge, the use of previously published studies performed in the Netherlands [25,28] has likely contributed to obtaining an accurate estimate of the long-term impact of PCT-guidance.

Limitations

As data concerning lost productivity were not measured as part of the SAPS, assumptions had to be made to estimate costs of productivity losses due to the recovery period and/or premature mortality. This could be considered as a limitation of the analysis. However, the friction period in the Netherlands is set at 85 calendar days, while the recovery period after hospitalization is set at 84 calendar days, which is likely a strong underestimation [19]. Consequently, every patient will incur the maximum friction period of 85 days regardless of patient outcomes. Therefore, this limitation will not have affected cost outcomes.

In addition, the SAPS did not collect data on resource use on the general ward. This resource use may for example include surgical interventions, imaging, blood transfusions, or the use of medication. However, these costs are assumed to be relatively small compared to the high costs of ICU stay. It is thus unlikely that including these costs would have changed the overall conclusion. Furthermore, as approximately 50% of the infections in the SAPS were hospital or ICU-acquired, it would have been difficult to determine whether these costs on the general ward are indeed related to the sepsis, or related to the initial reason for hospitalization.

All model inputs regarding quality of life were based on published literature [28], which only concerned patients with severe sepsis. As a consequence, the quality of life from this previous

study may be lower than what would have been expected in the current analysis, which included patients with varying sepsis severity. However, similar or lower utilities were reported in previously published studies [29]. This makes it likely that the data from Hofhuis et al. were suitable for the purpose of the current analysis [28].

Results of sensitivity analysis indicate that there is relatively large uncertainty regarding the impact of PCT-guidance on costs (Appendix 2). Although this complicates decision making, this uncertainty is almost inevitable owing to the fact that sepsis is known to be a highly heterogeneous condition. Consequently, the duration of ICU and general ward stay vary strongly between patients. As these parameters are two of the main cost drivers, the uncertainty in total costs remains considerable despite the high number of patients included in the SAPS.

Recommendations for further research

The adherence to the advice to stop antibiotic treatment (based on the PCT value) was only followed-up within 24h in 44% of the patients [10]. This non-adherence may be explained by hesitance of physicians to stop antibiotic use in unstable patients [10]. However, over time, physicians may gain trust in PCT testing which may increase adherence. Further research is therefore warranted to investigate the future impact of PCT testing on antibiotic use, health outcomes and costs.

Since antibiotic prescription and the duration of hospitalization is relatively low in the Netherlands [15,16], the results of the current study may not be readily transferable to other countries. Consequently, it is likely that the benefits shown in the current study represent the minimum expected benefits of PCT-guidance. The within-hospital benefits of PCT-guidance are therefore expected to be larger when performed in other countries, making it also more likely that it may be cost-effective in these countries.

As mentioned previously, the impact of PCT-guidance on long-term survival (i.e. >1 year) and accompanying long-term costs, could not be quantified in the current analysis. In addition, as the number of studies reporting on long-term survival and healthcare costs among sepsis survivors is very limited, extrapolating the results of the current study over a longer time period would make the results highly uncertain. Although the current one-year results cannot yet be considered cost-effective at the most commonly used WTP threshold of €20,000/QALY, the actually applied WTP threshold also depends on disease severity. WTP thresholds up to €80,000/QALY have been used for diseases or conditions with a high burden [30]. With an ICUR of almost €60,000/QALY gained, PCT-guidance would be considered cost-effective at WTP thresholds starting from ~€60,000/QALY. Besides, the impact of a shorter antibiotic treatment duration may decrease the incidence of resistant infections and lead to a relative decrease in costs in the PCT group [14]. Furthermore, the lower mortality rate in the PCT group may incur further benefits to society which cannot be quantified in a health economic evaluation,

including the impact on the quality of life of relatives. Thus, in order to fully quantify the impact of PCT-guidance, long-term follow-up studies, from a societal perspective, are required.

CONCLUSION

PCT-guided antibiotic treatment in ICU patients with sepsis is safe and reduces the duration of antibiotic use, while the impact on hospital costs is minimal. When considering one-year outcomes, the high healthcare-related costs occurring in sepsis survivors are however expected to lead to a (non-significant) increase in costs of the PCT group compared with standard-of-care. Despite the estimated increase in QALYs (achieved by the reduced mortality rate in the PCT group), the use of PCT-guidance cannot be considered cost-effective within a one-year time frame, at a WTP threshold of €20,000/QALY. However, to fully quantify the costs of PCT-guidance from a societal perspective, long-term follow-up studies are required.

AUTHORS' CONTRIBUTIONS

MMAK, HK, JAvO, MWNN and RK were involved in the conception and design of the study. MMAK, JAvO, AS, MWNN, HK, and AMSB performed the data analysis. MMAK, JAvO, AS, MWNN, and HK had full access to all of the study data and take responsibility for their integrity and the accuracy of the analyses. All authors were involved in the interpretation of the results. MMAK drafted the manuscript, and all authors were major contributors in critically reviewing the manuscript for important intellectual content. All authors read and approved the final manuscript.

FUNDING

No funding was received for this study.

DECLARATION OF INTERESTS

DWdL reports grants from ThermoFisher, during the conduct of the SAPS-trial, which data has been used for this study. All other authors declare no competing interests related to this manuscript.

REFERENCES

1. Becker, K.L., R. Snider, and E.S. Nylen, Procalcitonin assay in systemic inflammation, infection, and sepsis: clinical utility and limitations. *Crit Care Med*, 2008. 36(3): p. 941-52.
2. Nobre, V., et al., Use of procalcitonin to shorten antibiotic treatment duration in septic patients: a randomized trial. *Am J Respir Crit Care Med*, 2008. 177(5): p. 498-505.
3. Bouadma, L., et al., Use of procalcitonin to reduce patients' exposure to antibiotics in intensive care units (PRORATA trial): a multicentre randomised controlled trial. *Lancet*, 2010. 375(9713): p. 463-74.
4. Annane, D., et al., Procalcitonin levels to guide antibiotic therapy in adults with non-microbiologically proven apparent severe sepsis: a randomised controlled trial. *BMJ Open*, 2013. 3(2).
5. Deliberato, R.O., et al., Clinical and economic impact of procalcitonin to shorten antimicrobial therapy in septic patients with proven bacterial infection in an intensive care setting.

- Diagn Microbiol Infect Dis, 2013. 76(3): p. 266-71.
6. Hochreiter, M., et al., Procalcitonin to guide duration of antibiotic therapy in intensive care patients: a randomized prospective controlled trial. *Crit Care*, 2009. 13(3): p. R83.
 7. Schroeder, S., et al., Procalcitonin (PCT)-guided algorithm reduces length of antibiotic treatment in surgical intensive care patients with severe sepsis: results of a prospective randomized study. *Langenbecks Arch Surg*, 2009. 394(2): p. 221-6.
 8. Shehabi, Y., et al., Procalcitonin algorithm in critically ill adults with undifferentiated infection or suspected sepsis. A randomized controlled trial. *Am J Respir Crit Care Med*, 2014. 190(10): p. 1102-10.
 9. Xu, X.L., et al., [Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment of sepsis patients]. *Zhonghua Yi Xue Za Zhi*, 2017. 97(5): p. 343-346.
 10. de Jong, E., et al., Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: a randomised, controlled, open-label trial. *Lancet Infect Dis*, 2016.
 11. Schuetz, P., et al., Procalcitonin to initiate or discontinue antibiotics in acute respiratory tract infections. *Cochrane Database Syst Rev*, 2017. 10: p. CD007498.
 12. Harrison, M. and C.D. Collins, Is procalcitonin-guided antimicrobial use cost-effective in adult patients with suspected bacterial infection and sepsis? *Infect Control Hosp Epidemiol*, 2015. 36(3): p. 265-72.
 13. Wilke, M.H., R.F. Grube, and K.F. Bodmann, The use of a standardized PCT-algorithm reduces costs in intensive care in septic patients - a DRG-based simulation model. *Eur J Med Res*, 2011. 16(12): p. 543-8.
 14. Kip, M.M., et al., A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis. *J Med Econ*, 2015. 18(11): p. 944-53.
 15. OECD/EU, Health at a Glance: Europe 2016 - State of Health in the EU Cycle. 2016, OECD Publishing: Paris.
 16. European Centre for Disease Prevention and Control, Surveillance of antimicrobial consumption in Europe 2012. 2014, ECDC: Stockholm.
 17. Assink-de Jong, E., et al., Stop Antibiotics on guidance of Procalcitonin Study (SAPS): a randomised prospective multicenter investigator-initiated trial to analyse whether daily measurements of procalcitonin versus a standard-of-care approach can safely shorten antibiotic duration in intensive care unit patients--calculated sample size: 1816 patients. *BMC Infect Dis*, 2013. 13: p. 178.
 18. Zorginstituut Nederland, Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2015. p. 1-51.
 19. Prescott, H.C. and D.C. Angus, Enhancing Recovery From Sepsis: A Review. *JAMA*, 2018. 319(1): p. 62-75.
 20. Centraal Bureau voor de Statistiek. Consumentenprijzen; prijsindex 2006 = 100. 2017 [cited 2017 November 12]; Available from: <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=83131NED>.
 21. R Core Team, R: A language and environment for statistical computing. 2015, R Foundation for Statistical Computing: Vienna, Austria.
 22. van der Heijden, G.J., et al., Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*, 2006. 59(10): p. 1102-9.
 23. van Buuren, S. and K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 2011. 45(3): p. 1-67.
 24. Campbell, M.K. and D.J. Torgerson, Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. *QJM*, 1999. 92(3): p. 177-82.
 25. Koster-Brouwer, M.E., et al., Chronic health-care expenditure in survivors of sepsis in the intensive care unit. *Intensive Care Med*, 2016. 42(10): p. 1641-1642.
 26. Briggs, A. and K. Claxton, Decision Modelling for Health Economic Evaluation, in NY. 2006, Oxford University Press: New York.
 27. VMS veiligheidsprogramma. Voorkomen van lijnsepsis en behandeling van ernstige sepsis. 2013 [cited 2014 August 8]; Available from: <http://www.vmszorg.nl/Themas/Sepsis>.
 28. Hofhuis, J.G., et al., The impact of severe sepsis on health-related quality of life: a long-term follow-up study. *Anesth Analg*, 2008. 107(6): p. 1957-64.
 29. Alam, N., et al., Long-term health related quality of life in patients with sepsis after intensive care stay: A systematic review. *Acute Med*, 2017. 16(4): p. 164-169.
 30. Zwaap, J., et al., Kosteneffectiviteit in de praktijk. 2015, Zorginstituut Nederland. p. 1-64.

APPENDIX 1

Overview of model input parameters and assumptions used

Table 1. Overview of effectiveness input parameters (from literature). This table shows the model input parameters regarding the quality of life among sepsis patients, and the data source used. ICU = intensive care unit.

Quality of life, at:	Value	Reference
ICU discharge	0.50	[1,2]
Hospital discharge	0.64	[1,2]
3 months after ICU discharge	0.73	[1,2]
6 months after ICU discharge	0.75	[1,2]
12 months after study initiation	Varying (extrapolated per individual patient, depending on duration of ICU stay and total hospital stay)	[1-3]

Table 2. Overview of cost input parameters. This table shows the cost input parameters used in the cost-effectiveness analysis, showing the parameter, the value used in the model, and the data source. CI = confidence interval, SDD = selective digestive decontamination, SOD = selective oral decontamination.

Parameter	Costs	Reference
Intensive care unit day		
- Category 1	€ 1,854.04	[4] assumption
- Category 2	€ 2,024.77	[4], assumption
- Category 3	€ 2,302.53	[4], assumption
Mechanical ventilation day		
- Category 1	€ 304.07	[5,6], assumption
- Category 2	€ 379.93	[5,6], assumption
- Category 3	€ 428.51	[5,6], assumption
Regular ward day		
- Academic	€ 447.09	[6,7]
- General	€ 647.93	[6,7]
Dialysis day	€ 473.16	[4]
Antibiotic day	Varying (obtained from SAPS)	[8]
SOD day	€ 1.00	Elisabeth-TweeSteden Hospital, Tilburg
SDD day	€ 48.00	Elisabeth-TweeSteden Hospital, Tilburg

Order tariff laboratory tests		
- with phlebotomy	€ 12.03	[4]
- without phlebotomy	€ 6.27	[4]
Blood culture	€ 31.82	[4]
Other culture	€ 19.53	[4]
Determination of positive culture	€ 9.89	[4]
Procalcitonin (PCT)	€ 31.71	[4]
C-reactive protein (CRP)	€ 4.19	[4]
Bilirubin	€ 1.54	[4]
Creatinin	€ 1.69	[4]
Leukocytes	€ 1.85	[4]
Thrombocytes	€ 1.83	[4]
Haemoglobin (incl. haematocrit)	€ 1.78	[4]
Differential blood cell count	€ 1.85	[4]
Glucose	€ 1.69	[4]
Potassium	€ 1.68	[4]
Sodium	€ 1.69	[4]
Urea	€ 1.57	[4]
Aspartate aminotransferase (ASAT)	€ 1.86	[4]
Alanine-aminotransferase (ALAT)	€ 1.96	[4]
Lactate dehydrogenase (LD)	€ 1.86	[4]
Gamma-glutamyltransferase (GGT)	€ 1.84	[4]
Alkaline phosphatase (AF)	€ 1.87	[4]
Lactate	€ 12.84	[4]
Blood gas	€ 5.39	[4]
Urinary sediment	€ 2.62	[4]
Productivity losses	Varying (depending on duration of hospital stay and survival)	[9-11]
Long-term care costs	Varying (depending on duration of hospital stay and survival)	[8,12]

Assumptions used in the cost-effectiveness analysis:

Regarding the costs of an ICU-day:

- In the Netherlands, all ICUs have been classified into level 1, 2 and 3, based on ICU size, patient volume, ventilation days, and staffing [13]. This level also determines the tariff that is charged per day that a patient spends on the ICU. Therefore, the costs per ICU day were determined per hospital individually, as based on the classification of the hospitals' ICUs.
- As this ICU tariff also includes costs of imaging procedures (which were not collected in the database), and which were assumed to be independent of the group the patient is assigned to (e.g. PCT or standard-of-care), costs of these procedures were not separately accounted for in this analysis.
- Costs per day mechanical ventilation were based on the add-on tariff which was used until 2013, which was converted to 2017 Euros. However, as the costs of mechanical ventilation are currently incorporated in the tariff of one day ICU stay, costs of mechanical ventilation were subtracted from this tariff of ICU stay, to avoid double counting of costs. From the SAPS database it was calculated that patients receive, on average, mechanical ventilation for 33% of the time they spend on the ICU.

Regarding all laboratory tests:

- It is assumed that laboratory tests are performed once-daily until hospital discharge. The order tariff without phlebotomy is used for patients on the ICU (as the blood sample is commonly taken from the central line), while the tariff with phlebotomy is used for patients on the regular ward.

Regarding laboratory blood tests:

- The number of bilirubin, creatinine, leukocyte, thrombocyte, CRP and PCT tests in the first 28 days of hospitalization was derived from the SAPS database. The assumed set of laboratory tests that was performed upon ICU admission was based on a previous study [14]. For patients hospitalized >28 days, it was assumed that creatinine, leukocyte, thrombocyte and CRP tests were performed once per three days during the hospitalization period.
- Other laboratory tests (haemoglobin, haematocrit, sodium, potassium, urea, and glucose) were assumed to be performed once per three days during the hospitalization period.

Regarding blood cultures and other cultures:

- During the SAPS study it was recorded how many blood cultures were performed in each patient. In addition, it was also recorded how many other cultures were performed in this study, although it was not specified which type of culture this was. Therefore, we assumed that this concerned a bacteriologic culture, at an average cost of €19.53 [4].
- Both for blood cultures and other cultures, it was assumed that 35% was positive [15] and required determination of the micro-organism.

APPENDIX 2

Results of sensitivity analysis

Table 1. Overview of model outcomes. This table shows an overview of the model outcomes of the sensitivity analysis in terms of effectiveness (i.e. in-hospital mortality, antibiotic days, and QALYs), and in terms of hospital costs. These costs are subdivided into direct hospital costs (consisting of costs of hospital stay, organ support, medication, and laboratory tests), total healthcare-related costs within one year follow-up, and total societal costs which includes productivity losses. The mean model outcomes (and accompanying 95% CIs) are shown for the PCT group and the standard-of-care group. In addition, the difference between these groups (and accompanying 95% CIs) are provided.

Type of parameter	Parameter	PCT (95% CI)	Standard-of-care (95% CI)	Effect (95% CI)
In-hospital mortality		21.8% (17.2% to 26.3%)	29.7% (23.3% to 36.5%)	-7.9% (-13.9% to -1.8%)
Antibiotic days		6.97 (5.63 to 8.57)	8.17 (6.97 to 9.51)	-1.20 (-1.93 to -0.39)
QALYs		0.52 (0.49 to 0.54)	0.47 (0.43 to 0.51)	0.05 (0.00 to 0.10)
Hospital stay	ICU stay	€32,859 (€25,744 to €41,010)	€32,322 (€25,567 to €39,790)	€537 (-€5,109 to €6,369)
	General ward stay	€9,574 (€6,193 to €13,268)	€9,952 (€6,290 to €14,138)	-€310 (-€662 to €28)
Organ support	Mechanical ventilation	€1,986 (€1,532 to €2,544)	€2,252 (€1,735 to €2,859)	-€266 (-€574 to €19)
	Renal replacement therapy	€361 (€229 to €524)	€404 (€233 to €616)	-€44 (-€219 to €116)
Medication	Antibiotics	€202 (€131 to €282)	€236 (€168 to €315)	-€34 (-€72 to €6)
	SDD and SOD	€126 (€31 to €231)	€155 (€41 to €274)	-€29 (-€65 to -€1)
Laboratory tests	Cultures	€109 (€70 to €156)	€122 (€79 to €170)	-€13 (-€35 to €6)
	PCT	€204 (€159 to €254)	€0 (€0 to €0)	€204 (€159 to €254)
	Other tests (incl. order tariff)	€582 (€473 to €695)	€601 (€490 to €726)	-€19 (-€96 to €59)
	Total hospital costs	€46,002 (€36,633 to €56,492)	€46,045 (€37,265 to €55,414)	-€42 (-€6,279 to €6,443)

Healthcare costs (follow-up)	€27,606 (€25,415 to €29,797)	€24,855 (€21,997 to €27,574)	€2,751 (€91 to €5,506)
Total healthcare costs (up to one year follow-up)	€73,609 (€64,265 to €84,010)	€70,900 (€62,578 to €79,939)	€2,709 (-€4,442 to €10,271)
Lost productivity	€6,981 (€6,591 to €7,367)	€6,924 (€6,566 to €7,270)	€57 (-€359 to €476)
Total societal costs (up to one year follow-up)	€80,589 (€71,137 to €91,013)	€77,823 (€69,462 to €86,901)	€2,766 (-€4,434 to €10,354)

Figure 1. Incremental cost-effectiveness plane showing the impact of the use of a PCT-guided antibiotic treatment algorithm, as compared to the standard-of-care, on the difference in in-hospital mortality and accompanying costs within this (initial) hospitalization episode. The result is based on 10,000 bootstrap samples, and based on the scenario in which all utilities and costs were varied with their (assumed) standard error.

Figure 2. Incremental cost-effectiveness plane showing the impact of the use of a PCT-guided antibiotic treatment algorithm, as compared to the standard-of-care, on the difference in QALYs (until one year after ICU admission) and accompanying healthcare-related costs within this one-year time period. In addition, the willingness-to-pay thresholds of €20,000/QALY and €80,000/QALY are shown. The result is based on 10,000 bootstrap samples, and based on the scenario in which all utilities and costs were varied with their (assumed) standard error. QALY = quality-adjusted life year; WTP = willingness-to-pay.

Figure 3. Cost-effectiveness acceptability curve showing the probability that the use of a PCT-guided antibiotic treatment is cost-effective compared to the standard-of-care, for a WTP threshold ranging from €0/QALY to €200,000/QALY. This analysis incorporates all healthcare-related costs over a one-year time horizon, and is based on the scenario in which all utilities and costs were varied with their (assumed) standard error. WTP: willingness-to-pay; QALY: quality-adjusted life year.

REFERENCES

1. Hofhuis, J.G., et al., The impact of severe sepsis on health-related quality of life: a long-term follow-up study. *Anesth Analg*, 2008. 107(6): p. 1957-64.
2. Ara, R. and J. Brazier, Deriving an algorithm to convert the eight mean SF-36 dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available). *Value Health*, 2008. 11(7): p. 1131-43.
3. R Core Team, R: A language and environment for statistical computing. 2015, R Foundation for Statistical Computing: Vienna, Austria.
4. Nederlandse Zorgautoriteit, Tarieventabel_dbc_zorgproducten_en_overige_producten_per_1_januari_2017. 2017.
5. Nederlandse Zorgautoriteit, 20131010 Tarieventabel DBC zorgproducten en overige producten per 1 oktober 2013. 2013.
6. Centraal Bureau voor de Statistiek. Consumentenprijzen; prijsindex 2006 = 100. 2017 [cited 2017 November 12]; Available from: <http://statline.cbs.nl/StatWeb/publication/?VW=T&D-M=SLNL&PA=83131NED>.
7. Hakkaart-van Roijen, L., et al., Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. 2015, Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam: Rotterdam. p. 1-120.
8. de Jong, E., et al., Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: a randomised, controlled, open-label trial. *Lancet Infect Dis*, 2016.
9. Zorginstituut Nederland, Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2015. p. 1-51.
10. Centraal Bureau voor de Statistiek. Werkzame beroepsbevolking; arbeidsduur. 2017 [cited 2017 December 12]; Available from: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82647NED/table?ts=1513066092696>.
11. Centraal Bureau voor de Statistiek. Arbeidsdeelname en werkloosheid per maand. 2017 [cited 2017 December 12]; Available from: <http://statline.cbs.nl/Statweb/publication/?D-M=SLNL&PA=80590ned&D1=8-9&D2=0&D3=0&D4=177-191&VW=T>.
12. Koster-Brouwer, M.E., et al., Chronic health-care expenditure in survivors of sepsis in the intensive care unit. *Intensive Care Med*, 2016. 42(10): p. 1641-1642.
13. Kluge, G.H., et al., The association between ICU level of care and mortality in the Netherlands. *Intensive Care Med*, 2015. 41(2): p. 304-11.
14. Kip, M.M., et al., A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis. *J Med Econ*, 2015. 18(11): p. 944-53.
15. Singer, M., et al., The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 2016. 315(8): p. 801-10.

PART III

Evaluation of an extensive laboratory
analysis for the diagnosis and treatment
of anaemia in general practice

8

The effectiveness of a routine versus an extensive laboratory analysis in the diagnosis of anaemia in general practice

Annemarie Schop*
Michelle M.A. Kip*
Karlijn Stouten
Soraya Dekker
Jurgen Riedl
Ron J. van Houten
Joost van Rosmalen

Geert-Jan Dinant
Maarten J. IJzerman
Hendrik Koffijberg
Patrick J.E. Bindels
Ron Kusters
Mark-David Levin

* These authors contributed equally

This chapter has been published as:

A. Schop, M.M.A. Kip, K. Stouten, S. Dekker, J. Riedl, R.J. van Houten, J. van Rosmalen, G-J. Dinant, M.J. IJzerman, H. Koffijberg, P.J.E. Bindels, R. Kusters, M-D. Levin. *The effectiveness of a standardized versus a personalized laboratory analysis in the diagnosis of anaemia in general practice*. Ann Clin Biochem, 2017: p. 4563217748680

Reprinted with permission.

ABSTRACT

Background

We investigated the percentage of patients diagnosed with the correct underlying cause of anaemia by general practitioners when using an extensive versus a routine laboratory work-up.

Methods

An online survey was distributed among 836 general practitioners. The survey consisted of six cases, selected from an existing cohort of anaemia patients ($n = 3,325$). In three cases, general practitioners were asked to select the laboratory tests for further diagnostic examination from a list of 14 parameters (i.e. routine work-up). In the other three cases, general practitioners were presented with all 14 laboratory test results available (i.e. extensive work-up). General practitioners were asked to determine the underlying cause of anaemia in all six cases based on the test results, and these answers were compared with the answers of an expert panel.

Results

A total of 139 general practitioners (partly) responded to the survey (17%). The general practitioners were able to determine the underlying cause of anaemia in 53% of cases based on the routine work-up, whereas 62% of cases could be diagnosed using an extensive work-up ($P = 0.007$). In addition, the probability of a correct diagnosis decreased with the patient's age and was also affected by the underlying cause itself, with anaemia of chronic disease being hardest to diagnose ($P = 0.003$).

Conclusion

The use of an extensive laboratory work-up in patients with newly diagnosed anaemia is expected to increase the percentage of correct underlying causes established by general practitioners. Since the underlying cause can still not be established in 31.3% of anaemia patients, further research is necessary.

INTRODUCTION

Anaemia (i.e. a lowered concentration of haemoglobin) is a common finding among elderly patients (aged 65 years and older) in general practice. Besides the sign of an underlying condition, it has long been considered a benign consequence of aging. However, during the last decade, many studies have been published detailing the relevance of anaemia, such as the associations between anaemia and increased mortality, physical and cognitive decline, cardiovascular events and reduced quality of life have been found [1, 2]. The only way to manage anaemia is through treatment of the underlying cause, which requires an additional diagnostic work up. The most common underlying causes of anaemia in general practice are iron deficiency anemia (IDA) (16.3% – 19.0%), anaemia of chronic disease (ACD) (19.7% – 31.4%) and renal anaemia (8.2% - 12.9%). In addition, a considerable proportion of anaemia cases have no clear cause and are classified as unknown anaemia (31.3% - 44.0%) (see literature [3-6], own data).

Different guidelines are published to help diagnosing the underlying cause of anaemia. In most of them, mean corpuscular volume (MCV) occupies a central position [7, 8]. However, Oosterhuis et al [9] demonstrated that when assigning a major role to ferritin concentration instead of MCV, leads to an increase of the percentage of patients diagnosed with an underlying cause from 48% to 71%. In addition, another study has shown that the percentage of patients diagnosed with an underlying cause of anaemia increases from 14% when general practitioners (GPs) personally order laboratory analysis to 53% when a standard set of 14 laboratory parameters was offered [10].

The high prevalence of anaemia, as well as the need for establishing the underlying aetiology prior to initiating treatment, demands an optimized diagnostic approach. We used an online survey among GPs to establish whether a routine or extensive laboratory approach is more effective in diagnosing the underlying cause of anaemia in general practice.

MATERIALS AND METHODS

Study Design

The percentage of correct diagnoses of the underlying cause of anaemia when using an extensive versus a routine laboratory work-up was investigated through an online survey using *LimeSurvey* [11]. The survey was distributed among 836 GPs, operating in different parts of the Netherlands and was available online for a period of one month (January 2016). The cases used in this survey were selected from a large database of GP patients, included between the 1 February 2007 and the 1 February 2015 ($n = 3,325$) [12]. This prospective cohort study was approved by the internal ethics committee of the Albert Schweitzer Hospital. The database consisted of patients aged 50 years and older (in order to exclude a predominance of iron deficiency due to hypermenorrhea), newly diagnosed with anaemia (i.e. no established

diagnosis of anaemia in the previous two years). Anaemia was defined as haemoglobin below 13.7 g/dL (8.5 mmol/L) for males and below 12.1 g/dL (7.5 mmol/L) for females. An extensive laboratory work-up was performed in all included patients, consisting of haemoglobin (Hb), MCV, C-reactive protein (CRP) and/or erythrocyte sedimentation rate (ESR), vitamin B₁₂, creatinine, ferritin, folic acid, lactate dehydrogenase (LDH), transferrin, reticulocytes, leukocytes, thrombocytes and serum iron. According to the current guidelines about anaemia diagnostics in general practice, these 14 parameters cover all underlying causes of anaemia that can be diagnosed in general practice [7].

Table 1. Distribution of underlying cause of anaemia. The distribution of the underlying causes of anaemia in the larger data-set of newly diagnosed anaemia patients was used to randomly select 201 cases with a similar distribution.

Underlying cause of anaemia	Number of cases before randomization(%)	Number of cases after randomization(%)
Iron deficiency anaemia	389 (16.3%)	33 (16.4%)
Anaemia of chronic disease	751 (31.4%)	63 (31.3%)
Renal anaemia	307 (12.9%)	26 (12.9%)
Other	942 (39.4%)	79 (39.3%)
- Unknown	748 (79.4%)	63 (79.7%)
- Haemoglobinopathy	16 (1.7%)	1 (1.3%)
- Haemolysis	9 (1.0%)	1 (1.3%)
- Possible bone marrow disease	40 (4.2%)	3 (3.8%)
- Vitamin B ₁₂ deficiency	64 (6.8%)	5 (6.3%)
- Folic acid deficiency	8 (0.8%)	1 (1.3%)
- Other	57 (6.1%)	5 (6.3%)
Total	2389	201

An expert panel, consisting of an experienced GP, internist and clinical chemist, determined the underlying cause of anaemia for each patient based on 10 pre-defined causes (i.e. IDA, ACD, renal anaemia, possible bone marrow disease, possible hemolysis, haemoglobinopathy, vitamin B₁₂ deficiency, folic acid deficiency, other and unknown) (Appendix 1). Patients with multiple aetiologies were excluded from the study (n = 293) since those cases were considered too complex for the design of this study. In addition, patients with missing laboratory values were also excluded (n = 643). The remaining patients from the prospective database (n = 2389) were divided into four subgroups: IDA (n = 389, 16.3%), ACD (n = 751, 31.4%), renal anaemia (n = 307, 12.9%) and other (including anaemia with an unknown cause) (n = 942, 39.4%). Based on the prevalence of each of the underlying causes of anaemia, 201 cases were selected from the database and used in the online survey. By selecting 201 cases, three equal sized pools (each consisting of 67 cases) could be used for randomization. For the subgroups IDA, ACD and renal anaemia, stratified randomization was used for selection of cases to ensure that the percentage

of included cases of each of those underlying causes was representative of its occurrence in the database [13]. In the fourth subgroup (i.e. other causes including unknown causes), cases of anaemia with an unknown cause were also selected through stratified randomization. The remaining other causes were selected manually by maximum variation sampling (based on patients' age, gender and the underlying cause of anaemia), which was necessary due to the small numbers of cases per cause [14]. The distribution of underlying causes of anaemia within the four subgroups (i.e. IDA, ACD, renal anaemia and other), before and after randomization is shown in Table 1.

Structure of online survey

For each participating GP, the survey included six patient cases which were randomly selected by LimeSurvey from the set of 201 available cases. For each case, GPs were provided with the age and gender of the patient and were informed that the patient presented with suspicion of anaemia. While filling in the survey, respondents were invited to use guidelines or other tools they also use in daily practice. For the first three cases, respondents were asked to choose the laboratory tests they considered necessary from the predefined list of 14 parameters. This was referred to as the 'routine work-up'. The respondents were then presented with the results of the selected laboratory tests and were invited to determine the underlying cause of anaemia when possible. In cases where they responded that additional information was required, they were given one opportunity to request additional tests from the same list of 14 tests. No limit was set on the number of laboratory tests that could be selected per analysis, but the same test could not be ordered twice. For the second set of three cases, respondents were presented with the results of all 14 tests, and they did not have the ability to request additional tests. This was referred to as the 'extensive work-up'. Again, respondents were asked to determine the underlying cause of anaemia. The survey ended with several personal questions regarding the GPs' gender, age, years active as GP, zip code, daily use of guidelines for the diagnosis/treatment of anaemia patients, and whether the GP had any special affinity with the subject of anaemia. The diagnoses of the respondents were compared with the diagnoses of the expert panel. For the cases with IDA, ACD and renal anaemia as underlying cause, the diagnoses had to be in accordance with the diagnoses of the expert panel. For all other underlying causes (i.e. vitamin B₁₂ deficiency, unknown cause, et cetera), the diagnosis was considered correct if the option IDA, ACD or renal anaemia had not been chosen by the respondent. This classification is a commonly used method in literature and was considered appropriate as diagnosing or excluding of the three most prevalent underlying causes is clinically the most relevant [3-6].

A sample of the survey is shown in Appendix 2. The percentage of correct underlying causes of anaemia as established by GPs was defined as 'effectiveness', and is referred to as such throughout this manuscript.

Statistical analysis

The patient population was described by standard descriptive statistics. An additional analysis was performed to confirm that the choice of laboratory test(s) requested by the GP was in accordance with the diagnosis established by the same GP. The effects of the characteristics of both the patient case and the responding GP on the probability of a correct diagnosis were analyzed using generalized linear mixed models with a logistic link function and a binomial error distribution (i.e. logistic regression analyses with random effects) [15]. These models take into account the correlations between observations due to repeated measurements for both cases and respondents and can handle data with missing observations in the outcome. In the analysis of the effectiveness of the routine versus the extensive laboratory work-up, the dependent variable was a dichotomous variable, indicating whether the case had been diagnosed correctly, and the independent variables were type of work-up (i.e. routine or extensive), age and gender of the case and the cause of anaemia as established by the expert panel. This model was also used for the analysis of the effect of GP characteristics on correct diagnoses, but with the following variables added as independent variables: number of years of working experience as GP, use of guidelines, and whether the respondent had specific affinity with anaemia. GPs with missing information on any of these GP characteristics were excluded from the latter analysis ($n = 16$ [11.5%]). To take into account correlations due to repeated measurements, the relevance of including random effects of both cases and respondents was determined with the Akaike information criterion. Based on this criterion, only random effects of cases were included in the final model. The data were analyzed using IBM SPSS Statistics, version 24 and R version 3.3.1 with the NLME package [16, 17]. All statistical tests were two-sided, and P-values lower than 0.05 were considered significant.

Figure 1. A flow diagram showing the general practitioners who responded to the survey.

RESULTS

Respondents online survey

A total of 125 respondents (15%) completed the survey, resulting in 375 complete cases for both the routine and extensive work-up. In addition, 14 respondents (1.7%) filled in only part of the survey, resulting in an additional 29 cases in the routine work-up (total 404 cases) and three cases in the extensive work-up (total 378 cases) (Figure 1).

Of all 139 respondents, 123 (89.5%) answered the questions regarding personal characteristics. According to the NIVEL institute (Netherlands Institute for Health Services Research) the proportion of males participating in this survey as compared to the overall population of GPs (55.6% versus 55.0%), as well as the median age (48 years versus a median age category of 45-49 years) did not show major differences, suggesting that the respondents are a representative sample of the overall GP population in the Netherlands [18].

Table 2. Routine requested laboratory analysis of GPs. Of the 404 cases assessed with a routine work-up, laboratory tests were ordered once in 216 cases (53.5%) and twice in 183 cases (45.3%). The requested tests are shown as number (percentage). CRP: C-reactive protein; ESR: erythrocyte sedimentation rate; GP: general practitioner; MCV: mean corpuscular volume; LDH: lactate dehydrogenase.

Parameter	1st round (n = 399)	2nd round (n = 183)	Total (n = 404)
Haemoglobin	394 (98%)	0 (0%)	394 (98%)
MCV	370 (92%)	4 (1%)	374 (93%)
Reticulocytes	85 (21%)	65 (16%)	150 (37%)
Leukocytes	160 (40%)	57 (14%)	217 (54%)
Thrombocytes	133 (33%)	48 (12%)	181 (45%)
CRP and/or ESR	267 (66%)	65 (16%)	332 (82%)
Creatinine	235 (58%)	68 (17%)	303 (75%)
Ferritin	184 (46%)	72 (18%)	256 (63%)
Vitamin B ₁₂	126 (31%)	74 (18%)	200 (50%)
Folic acid	110 (27%)	67 (17%)	177 (44%)
LDH	47 (12%)	63 (16%)	110 (27%)
Transferrin	50 (12%)	70 (17%)	120 (30%)
Serum iron	70 (17%)	75 (19%)	145 (36%)
No laboratory analysis	5 (1%)	221 (55%)	5 (1%)

Effect of diagnostic work-up on correctly diagnosing anaemia

Of the 404 cases assessed with a routine work-up, laboratory tests were ordered once in 216 cases (53.5%) and twice in 183 cases (45.3%). No tests were ordered in the remaining five cases (1.2%). The median number of tests ordered during the first and second laboratory analysis

was five (IQR: 3 - 8) and four (IQR: 2 - 6), respectively. A detailed overview of the frequency at which each laboratory analysis was requested by GPs in the routine work-up is shown in Table 2.

Table 3. Diagnosis of underlying cause of anaemia using a routine or extensive work-up. The diagnosis set by the GPs are showed against the diagnosis set by the expert panel for both routine and extensive work-up. The bold values are the correct diagnoses.

Diagnosis GP	Diagnosis expert panel			
	IDA	ACD	Renal anaemia	Other (including unknown)
Routine				
IDA	45 (69.2)	13 (10.3)	2 (3.8)	14 (8.8)
ACD	5 (7.7)	56 (44.4)	17 (32.1)	51 (31.9)
Renal anaemia	0 (0)	4 (3.2)	28 (52.8)	10 (6.3)
Other (including unknown)	15 (23.1)	53 (42.1)	6 (11.3)	85 (53.1)
Total	65	126	53	160
Extensive				
IDA	47 (70.1)	19 (14.4)	0 (0)	11 (8.3)
ACD	6 (9.0)	72 (54.5)	10 (21.7)	28 (21.1)
Renal anaemia	4 (6.0)	8 (6.1)	30 (65.2)	9 (6.8)
Other (including unknown)	10 (14.9)	33 (25.0)	6 (13.0)	85 (63.9)
Total	67	132	46	133

When using a routine work-up, the correct underlying cause of anaemia was established in 214 of 404 cases (53.0%), and with the extensive laboratory work-up, 234 of 378 cases (61.9%) were correctly diagnosed (Table 3). The increase in probability of a successful diagnosis due the extensive laboratory work-up was statistically significant in the multivariate analysis (OR = 1.56 [95% CI 1.12-2.17]) (Table 4).

In addition, patients aged 65 years or older seemed more difficult to diagnose correctly compared to patients aged 50-64 years (65-74 years OR: 0.48 [95% CI: 0.25-0.90], 75 - 84 years OR: 0.45 [95% CI 0.25-0.83], 85+ years OR: 0.47 [95% CI: 0.25-0.87]). Finally, the underlying cause itself affects the probability of a correct diagnosis, with ACD being the most difficult to diagnose compared to IDA, OR = 0.37 (95% CI: 0.19-0.73). When running the analysis again including the GP characteristics (i.e. excluding GPs with missing data in their characteristics), none of the analyzed GP characteristics showed a significant association with the probability of diagnosing the correct underlying cause of anaemia (Table 5).

Table 4. Multivariate analysis of the efficacy of extensive versus routine laboratory work-up.

	Odds ratio (95% CI)	P-value
Extensive work-up	1.56 (1.12-2.17)	0.007
Age of patient		
- 50 - 64 years	Reference category	
- 65 - 74 years	0.48 (0.25-0.90)	0.022
- 75 - 84 years	0.45 (0.25-0.83)	0.010
- 85+ years	0.47 (0.25-0.83)	0.016
Gender of patient (female)	1.28 (0.83-1.97)	0.258
Underlying cause		
- IDA	Reference category	
- ACD	0.37 (0.19-0.72)	0.003
- Renal anaemia	0.69 (0.31-1.56)	0.376
Other incl. unknown	0.58 (0.31-1.10)	0.097

Note: Multivariate analysis using a generalized linear mixed model showed a significant influence of the laboratory work-up, age of patient and the underlying cause of anaemia itself on the correct diagnosis of the underlying cause of anaemia.

Table 5. Multivariate analysis of the effect of GPs characteristics on the correct diagnosis of the underlying cause of anaemia.

	Value	Odds ratio (95% CI)	P-value
Work experience in years (median [IQR])	16 (9-25)	1.01 (0.99-1.02)	0.512
Use national GP guideline	58.5% regular	1.21 (0.84-1.74)	0.306
	41.5% never	Reference category	
Affinity with anaemia	25.2% yes	1.18 (0.78-1.80)	0.437
	74.8% no	Reference category	

Note: results based on a generalized linear mixed model that also included work-up type (i.e. routine or extensive), age and gender of the case, and the cause of anaemia as established by the expert panel as independent variables. Of all 139 respondents, 123 (89.5%) answered the questions regarding personal characteristics. None of the analyzed characteristics showed a significant effect on the probability of diagnosing the correct underlying cause of anaemia.

DISCUSSION

When diagnosing the underlying etiology of anaemia, laboratory tests can be selected by the GP (i.e. routine work-up) or a standard set of tests may be offered by the laboratory (i.e. an extensive work-up). To determine which work-up would be more effective in supporting the GP to diagnose the correct underlying cause of anaemia, we compared both approaches

using an online survey among GPs. Our study included 14 widely available laboratory tests, which are all recommended by the current Dutch anaemia guideline. This extensive laboratory work-up was shown to be more effective than a routine work-up, in which GPs could order laboratory tests twice (from this predefined set of 14 tests). Patient characteristics (aged 65 years and older) and the underlying cause itself (i.e. ACD) may negatively affect the probability of a correct diagnosis.

For many years, anaemia diagnosis was directed by MCV, which divided cases into microcytic, normocytic and macrocytic, with each category including a set list of causes. In recent years, several studies have shown that MCV should not be granted such a central role [19-21]. However, in this study MCV was still the second most often requested test during the first round of the routine work-up (93%), suggesting that GPs still follow the old classification system. This approach may lead to missed causes since, for example, a microcytic anaemia will lead to the exclusion of vitamin B₁₂ deficiency as a possible cause according to the old classification system, while in fact, this deficiency may still be present and contributing to the anaemia [21]. In addition to MCV, the most requested tests during the first round of the routine work-up are CRP/ESR (67%), creatinine (59%) and ferritin (46%). These three tests allow for the exclusion of the three most common causes of anemia, namely ACD (CRP and ferritin), renal anaemia (creatinine) and IDA (ferritin). This suggests that GPs, when ordering tests, first aim to diagnose or exclude the three most common causes of anaemia.

8

This study also demonstrated that increased age (65 years and older), decreases the likelihood of establishing the correct underlying cause of anaemia independently from the laboratory work-up. This may be due to the fact that older patients more often display slightly increased infection parameters [6, 22]. This abnormality in the laboratory results, in combination with consensus-based definitions of underlying causes, might confuse GPs when determining the underlying cause of anaemia. In addition, it is important to realize that the elderly often have multiple diseases or co-morbidities and each of those may (individually) contribute to anaemia. Especially in these patients, the laboratory results might be of less relevance and the clinical presentation (i.e. health condition) would be the guiding principle for treatment. This may have created complications for GPs while attempting to establish the cause in this study, since the survey relied entirely on laboratory results and did not include health condition or co-morbidities of the patient. However, there was no increase in the percentage of cases that were diagnosed as 'unknown' by GPs among the elderly (data not shown). Finally, since the elderly often have multiple diseases or co-morbidities, GPs may be reluctant to commit to an extensive diagnostic process and therefore only perform limited laboratory analyses in this group of patients.

The actual underlying cause of anaemia may also affect the probability of a correct diagnosis by GPs, with ACD being the most difficult to diagnose. The interpretation of laboratory values

can be challenging when ACD is present, especially in multiple causes, for example when an ACD co-exists with an iron deficiency. This combination was not present in the cases selected in our study. However, the cases in this study were classified by an expert panel according to pre-determined definitions of underlying causes of anemia. These definitions were based both on existing guidelines and on the opinion of this expert panel [23]. As a result, we used a strict cut-off value of ferritin (i.e. $>100 \mu\text{g/L}$) in the diagnosis of ACD. A participating GP may have used a different cut-off value or a different definition of ACD, based on his or her own experience. For the purpose of this study, the answer of these GPs may thus have been classified as incorrect and the number of correct diagnoses would, in fact, be higher. Furthermore, some causes have a very low prevalence in general practice. Being less familiar with these causes and their current definitions may make diagnosing them more challenging.

This study showed that the percentage of correct diagnoses of the underlying cause of anaemia increased from 53.0% when using a routine laboratory work-up to 61.9% when using an extensive laboratory work-up. Taking into consideration that annually 57,000 patients aged ≥ 50 years present with a new anaemia in Dutch general practice, this modest absolute difference of almost 9% will benefit a large number of patients (approximately 5130 patients) in whom the underlying etiology of anaemia can be established [24, 25]. Moreover, the extensive laboratory work-up is also expected to be cost-effective compared to the routine work-up, as is shown by the accompanying article by our group, published elsewhere in this journal.

Strengths and limitations

This study aimed to investigate which type of diagnostic work-up (i.e. extensive versus routine) results in the most correct diagnoses of the underlying cause of anaemia. To achieve this, we used real-life patient data from a large transmural project. The participating GPs showed no differences in age and gender compared to the overall population of GPs in the Netherlands, suggesting the participating GPs are an appropriate representative sample of the Dutch GP population. A possible limitation of this study is the lack of clinical information in the cases of the survey, besides age, gender, and the suspicion of anaemia. This clinical information may be of considerable importance in the diagnosis of the underlying cause of anaemia. Adding of this information may have allowed for a higher percentage of correct diagnosis. However, neither approach (routine and extensive work-up) contained this information and thereby this drawback did not affect the difference in percentage of correct diagnoses between both approaches. In addition, the cases were randomly assigned to the GPs, which means that there was a small chance that the same cause could have been the result for all six of the cases presented to a GP. This may cause the GP to perform better or worse, as this study has shown that the underlying cause itself has an effect on correctly diagnosing. However, since this scenario is also possible in daily practice, the survey reflects a real-world situation. In addition, five cases in whom the GP requested no laboratory test(s) were included in the final analysis because it was deemed a conscious choice of the GPs. In this study, we excluded anemia with

multiple etiologies since this is not addressed in the Dutch anaemia guideline and in order to avoid statistical complexity (i.e. more power would be needed against lower prevalence). In the routine laboratory work-up, multiple aetiologies might be missed if additional laboratory tests are withheld once an underlying cause is established. Therefore, it is to be expected that the extensive laboratory work-up performs better in the diagnosis of multiple aetiologies than the routine work-up, since all additional laboratory are simultaneously performed, making multiple aetiologies immediately apparent. Finally, each survey started with three cases for the routine work-up followed with three cases for the extensive work-up. This might have led to an overestimation of the effectiveness of the extensive work-up, since GPs have practiced and possibly learned from the previous cases and apply this to the following cases. However, to ensure that GPs chose the routine laboratory analysis based on their own experience, it was not possible to first show the cases of the extensive laboratory work-up, as this would have led to a larger bias.

CONCLUSION

An extensive laboratory work-up in patients with newly diagnosed anaemia is more effective in finding the underlying etiology than a routine laboratory work-up selected by GPs. Nevertheless, the percentage of incorrect diagnoses remains significant, which should take into account the fact that stand alone laboratory diagnostics is not the gold standard for anaemia diagnostics. Further studies should focus on an extensive laboratory work-up and the added value of multidisciplinary diagnostic approaches in patients with anaemia.

ACKNOWLEDGEMENTS

We thank the participating general practitioners for their generous contribution. In addition, the authors thank Dr. Rebecca Buis for critical reading of the manuscript. This paper was previously presented at the Science day, Albert Schweitzer Hospital, on June 2017.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ETHICAL APPROVAL

The database was approved by the internal ethics committee of the Albert Schweitzer Hospital.

CONTRIBUTORSHIP

AS, MMA, KS, SD, MDL, and RK were involved in the conception and design of the study and performed the data collection. AS, KS, MDL and JVR analyzed the data. All authors were involved in interpreting the results. AS drafted the manuscript, and all author authors were major contributors in critically reviewing the manuscript. All authors read and approved the final manuscript.

REFERENCES

- Lipschitz, D., Medical and functional consequences of anemia in the elderly. *J Am Geriatr Soc*, 2003. 51(3 Suppl): p. S10-3.
- Smith, R.E., Jr, The clinical and economic burden of anemia. *Am J Manag Care*, 2010. 16 Suppl Issues: p. S59-66.
- Patel, K.V., Epidemiology of anemia in older adults. *Semin Hematol*, 2008. 45(4): p. 210-7.
- Eisele, L., et al., Prevalence and incidence of anemia in the German Heinz Nixdorf Recall Study. *Ann Hematol*, 2013. 92(6): p. 731-7.
- Shavelle, R.M., R. MacKenzie, and D.R. Paculdo, Anemia and mortality in older persons: does the type of anemia affect survival? *Int J Hematol*, 2012. 95(3): p. 248-56.
- Ferrucci, L., et al., Proinflammatory state, hepcidin, and anemia in older persons. *Blood*, 2010. 115(18): p. 3810-6.
- van Wijk, M.A.M., et al., Nederlands Huisartsen Genootschap – standaard Anemie. *Huisarts Wet*, 2003. 46(1): p. 21-9 (rectificatie: *Huisarts Wet* 2003;46:147).
- World Health Organization, Iron deficiency anaemia: assessment, prevention, and control: a guide for programme managers. 2001, World Health Organization: Geneva, Switzerland.
- Oosterhuis, W.P., et al., Prospective comparison of the flow chart for laboratory investigations for anaemia from the Dutch College of General Practitioners' guideline 'Anaemia' with a self-developed, substantive and logistical alternative flow chart'. *Ned Tijdschr Geneesk*, 2007. 151: p. 2326-32.
- Droogendijk, J., et al., Adherence to the Dutch general practitioner anemia guideline - a retrospective cohort study. *Perfusion*, 2013. 6(26): p. 192-198.
- LimeSurvey Project Team / Schmitz, C., LimeSurvey: An Open Source survey tool. 2015, LimeSurvey Project: Hamburg, Germany.
- Stouten, K., et al., Prevalence of potential underlying aetiology of macrocytic anaemia in Dutch general practice. *BMC Fam Pract*, 2016. 17(1): p. 113.
- Sedgwick, P., Treatment allocation in trials: stratified randomisation. *BMJ*, 2015. 350: p. h978.
- Palinkas, L.A., et al., Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Health*, 2015. 42(5): p. 533-44.
- Cnaan, A., N.M. Laird, and P. Slasor, Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*, 1997. 16(20): p. 2349-80.
- IBM Corp, IBM SPSS Statistics for Windows. Released 2016, IBM Corp: Armonk, NY.
- Pinheiro, J., et al. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131. 2017; Available from: <https://cran.r-project.org/package=nlme>.
- van Hassel, D.T.P., A. Kasteleijn, and R.J. Kenens. Cijfers uit de registratie van huisartsen. Peiling 2014. 2015; Available from: <http://www.nivel.nl/sites/default/files/bestanden/Cijfers-uit-de-registratie-van-huisartsen-peiling-jan-2014.pdf>.
- Stouten, K., et al., Relevance of mean corpuscular volume for the evaluation of anaemia in general practice. Manuscript submitted, 2017.
- Seward, S.J., et al., Does the mean corpuscular volume help physicians evaluate hospitalized patients with anemia? *J Gen Intern Med*, 1990. 5(3): p. 187-91.
- Petrosyan, I., et al., Anaemia in the elderly: an aetiological profile of a prospective cohort of 95 hospitalised patients. *Eur J Intern Med*, 2012. 23(6): p. 524-8.
- Ershler, W.B., Biological interactions of aging and anemia: a focus on cytokines. *J Am Geriatr Soc*, 2003. 51(3 Suppl): p. S18-21.
- Halwachs-Baumann, G., Diagnosis of anaemia: old things rearranged. *Wien Med Wochenschr*, 2012. 162(21-22): p. 478-88.
- NIVEL. Wie heeft welke gezondheidsproblemen? Incidenties en prevalenties. 2014 [cited 2014 July]; Available from: <http://www.nivel.nl/NZR/wie-heeft-welke-gezondheidsproblemen>
- Centraal Bureau voor de Statistiek. Bevolking; geslacht, leeftijd en burgerlijke staat. 2017; Available from: <http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=7461BEV&D1=0&D2=0&D3=101-130&D4=0,65-66&HDR=G3,T&STB=G1,G2&VW=T>.

APPENDIX 1

Definitions of underlying causes of anaemia

Underlying cause of anaemia	Definition
Anaemia	Haemoglobin <13.7 g/dL (8.5 mmol/L) for males and <12.1 g/dL (7.5 mmol/L) for females
IDA	Ferritin <25 µg/L (males), <20 µg/L (females)
ACD	Ferritin >100 µg/L, transferrin <2.0 g/L and/or serum iron <14 µmol/L (males), <10 µmol/L (females)
Renal anaemia	Estimated creatinine clearance (MDRD) ≤45 mL/min/1.73m ²
Possible bone marrow disease	Abnormal number of leukocytes and platelets, reticulocytes <2.5%
Possible haemolysis	Raised LDH, reticulocytes >2.5% and bilirubin > 17 µmol/L
Vitamin B ₁₂ deficiency	Serum vitamin B ₁₂ <130 pmol/L
Folic acid deficiency	Serum folic acid <5 nmol/L
Other	Reported by the treating physician

APPENDIX 2

Structure online survey

- Based on the information provided, which of the 14 laboratory tests below would you like to request?

Multiple choice: (multiple answers possible) erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), ferritin, folic acid, hemoglobin, creatinine & MDRD, lactate dehydrogenase (LDH), leukocytes, mean corpuscular volume (MCV), reticulocytes, serum iron, transferrin, thrombocytes, vitamin B₁₂.

- Based on the laboratory analysis provided, which type of anaemia does this patient most likely have?

Multiple choice (only one answer possible): iron deficiency anaemia, anaemia of chronic disease, renal anaemia, other type, 'unknown, and I wish more laboratory analyses' or 'unknown but I don't wish more laboratory analyses.'

- In case 'other type' was selected in the previous question: what type of anaemia do you suspect in this patient?

Open question

- Given your chosen answer '_____' as the type of anaemia for this patient: what would be your next step in the treatment of this patient?

Multiple choice (multiple answers possible): end consultation, referral to medical specialist, prescription of medication and/or follow-up, other (open field).

- Based on your previous answer to refer the patient to a medical specialist, to which specialist would you refer the patient?

Open question

- Based on your previous answer to prescribe medication and/or follow-up, which medication would you prescribe and would you like to follow up the patient?

Multiple choice (multiple answers possible): oral iron supplementation, vitamin B₁₂, folic acid, I would like to follow up this patient, other (open field)

- Based on your previous answer to follow up the patient, after how many weeks you decide to make a follow up appointment?

Open question

9

Assessing the cost-effectiveness of a routine versus an extensive laboratory work-up in the diagnosis of anaemia in Dutch general practice

Michelle M.A. Kip*
Annemarie Schop*
Karlijn Stouten
Soraya Dekker
Geert-Jan Dinant

Hendrik Koffijberg
Patrick J.E. Bindels
Maarten J. IJzerman
Mark-David Levin
Ron Kusters

* These authors contributed equally

This chapter has been published as:

M.M.A. Kip, A. Schop, K. Stouten, S. Dekker, G-J. Dinant, H. Koffijberg, P.J.E. Bindels, M.J. IJzerman, M-D. Levin, R. Kusters. *Assessing the cost-effectiveness of a routine versus an extensive laboratory work-up in the diagnosis of anaemia in Dutch general practice*. Ann Clin Biochem, 2017: p. 4563217748984.

Reprinted with permission.

ABSTRACT

Background

Establishing the underlying cause of anaemia in general practice is a diagnostic challenge. Currently, general practitioners individually determine which laboratory tests to request (routine work-up) in order to diagnose the underlying cause. However, an extensive work-up (consisting of 14 tests) increases the proportion of patients correctly diagnosed. This study investigates the cost-effectiveness of this extensive work-up.

Methods

A decision-analytic model was developed, incorporating all societal costs from the moment a patient presents to a general practitioner with symptoms suggestive of anaemia (aged ≥ 50 years), until the patient was (correctly) diagnosed and treated in primary care, or referred to (and diagnosed in) secondary care. Model inputs were derived from an online survey among general practitioners, expert estimates and published data. The primary outcome measure was expressed as incremental cost per additional patient diagnosed with the correct underlying cause of anaemia in either work-up.

Results

The probability of general practitioners diagnosing the correct underlying cause increased from 49.6% (95% CI: 44.8%-54.5%) in the routine work-up to 56.0% (95% CI: 51.2%-60.8%) in the extensive work-up (i.e. +6.4% [95% CI: -0.6%-13.1%]). Costs are expected to increase slightly from €842/patient (95% CI: €704-€994) to €845/patient (95% CI: €711-€994), i.e. +€3/patient (95% CI: €-35-€40) in the extensive work-up, indicating incremental costs of €43 per additional patient correctly diagnosed.

Conclusions

The extensive laboratory work-up is more effective for diagnosing the underlying cause of anaemia by general practitioners, at a minimal increase in costs. As accompanying benefits in terms of quality of life and reduced productivity losses could not be captured in this analysis, the extensive work-up is likely cost-effective.

BACKGROUND

Anaemia is a common medical problem that carries substantial costs to the healthcare system and can be a burden on the health and quality of life of many individuals [1-10]. Therefore, adequate diagnosis and early initiation of correct treatment are essential [11]. However, anaemia is not a disease in itself, but is considered a sign of an underlying condition. Consequently, diagnosing the underlying cause of anaemia is often complex [12-14]. More specifically, previous research estimated that in 14–33% of older persons with anaemia the underlying cause is unknown [1, 12]. In addition, anemia is often under-diagnosed [15] and under-treated, as it is often considered a consequence of aging [12] and not as a specific symptom of disease.

The three most common underlying causes are: iron deficiency anaemia (IDA), anaemia of chronic disease (ACD), and renal anaemia [16]. Anaemia may also have a variety of other causes, including bone marrow diseases or vitamin deficiencies, such as B₁₂ and/or folic acid. Besides anamnesis and physical examination, laboratory analyses are required to identify the different underlying causes of anaemia. However, depending on the laboratory protocol used, the proportion of patients for whom no underlying cause of anaemia can be identified based on their laboratory analyses ranges from 28% to 52% [17]. In order to enhance the effectiveness of laboratory analyses, a guideline has been developed by the Dutch College of General Practitioners (DCGP) [18]. Previous research evaluated all laboratory tests of patients (women >50 years and men ≥18 years) with anaemia newly diagnosed by general practitioners (GPs) within a two-year time period. Unfortunately, 83.9% of those patients could not be diagnosed when applying the DCGP-guideline, because at least one of the required laboratory tests had not been performed [11].

In a recently performed study, we investigated whether an extensive laboratory work-up (consisting of a set of 14 tests) could increase the probability that patients are diagnosed with the correct underlying cause of anemia in Dutch general practice [19]. Although this study has shown that this work-up likely improves the probability that patients are correctly diagnosed with the underlying cause of anaemia, it is unknown whether this approach is cost-effective. In addition, even though most routine laboratory tests for diagnosing anaemia are relatively inexpensive, these test results will likely impact subsequent patient management decisions. As this may involve more expensive diagnostic testing and the referral of patients to secondary care, it is crucial to quantify the impact of such an extensive laboratory work-up further downstream the patient management pathway. Therefore, the current study aims to estimate the cost-effectiveness of this extensive laboratory work-up as compared with the current situation, the routine work-up, in which GPs decide for themselves which tests to request in patients presenting with symptoms of anaemia.

METHODS

Survey

The effectiveness of diagnosing the underlying cause of anaemia in general practice, using either the extensive or a routine work-up, was investigated through an online survey using LimeSurvey [20]. A full description of this survey is provided elsewhere [19]. Details on how this survey was distributed are provided in Appendix 1. In the survey, all participating GPs (139 out of 836, i.e. 16.6%) received six real-world cases of patients presenting with a new anaemia in general practice. In all six cases, the participating GPs were only provided with the age and gender of the patient and were informed that the patient was suspected of anaemia. In the first three cases, GPs were able to choose freely which tests they would request based on a predefined list of 14 common tests (haemoglobin, mean corpuscular volume [MCV], C-reactive protein [CRP] and/or erythrocyte sedimentation rate [ESR], vitamin B₁₂, creatinine, ferritin, folic acid, lactate dehydrogenase [LDH], transferrin, reticulocytes, leukocytes, thrombocytes, and serum iron) and were only given the results of the tests they selected. In the second set of three cases, GPs received the results of all 14 tests. In all six cases, the GPs were asked to choose between IDA, ACD, renal anaemia and 'other'. In cases where the GPs chose 'other', they were asked to specify the underlying cause of anaemia. GPs could also state that the cause was 'unknown', if they did not consider it possible to determine the underlying cause based on the laboratory results provided. Besides, the GPs were also asked which of the subsequent actions they would take in each case: close the consultation (i.e. do nothing), refer the patient to secondary care, prescribe medication (which involved prescribing iron, vitamin B₁₂, folic acid, or antibiotics), or see the patient again in a few weeks (follow-up). During the survey, the GPs were encouraged to use guidelines (e.g. the DCGP guideline) or other tools they use in daily practice.

9

Database used in the survey

A detailed explanation regarding how the cases used in the survey were obtained from this database has been described previously [19].

Health economic model

As health economic models are typically complex, a description of the main aspects regarding model structure and model inputs is provided below. A more extensive description, including the assumptions used, is provided in Appendix 1.

A decision tree was developed to estimate the cost-effectiveness of an extensive laboratory work-up compared with a routine work-up in diagnosing the underlying cause of anaemia in patients presenting with anaemia, aged ≥ 50 years, in Dutch general practice. A simplified version is shown in Figure 1. The correct underlying cause of anaemia was determined according to an expert panel, composed of an internist, a GP, and a clinical chemist. Incremental effect

Figure 1. Simplified decision tree demonstrating both laboratory work-ups (routine and extensive) of patients presenting with new anemia in general practices. *The structure of this decision tree is identical to the routine work-up, but differs in the probabilities that are used. The structure of the entire decision tree could not be shown due to lack of space. † In patients prescribed medication, the GPs chose to prescribe iron, vitamin B₁₂, folic acid, or antibiotics. ‡ Patients whom initially received a treatment that is ineffective (according to the expert panel), either have recover spontaneously, or are assumed to present at the GP again within a few weeks, and undergo a second round of diagnostics and treatment. In this second round, it is assumed that GPs will only make management decisions that are considered effective (medication or referral, depending on the underlying cause of anaemia). ACD: anaemia of chronic disease; GP: general practitioner; IDA: iron deficiency anemia.

was defined as the difference in the percentage of patients for whom the underlying cause of anaemia was correctly determined using a routine laboratory work-up (i.e. current practice) as compared with the extensive work-up. According to Dutch health economic guidelines, a societal perspective was taken in the cost-effectiveness analysis [21]. This means that all costs were included from the moment such a patient presents at the GP, until the patient is (correctly) diagnosed and treated in primary care, or referred to (and diagnosed in) secondary care. As the treatment of anaemia in secondary care strongly varies depending on its underlying cause, quantifying the impact of either work-up on patients' quality of life would require extensive individual patient-level data, which were not available. As such, this was considered outside the scope of this analysis. The time horizon in this study was therefore estimated to be, at most, 200 days. Incremental cost was defined as the difference in average costs per patient for whom the underlying cause of anaemia was determined using the extensive work-up as compared with the routine work-up. Costs were expressed in 2016 Euros. The model outcome was expressed as an incremental cost-effectiveness ratio (ICER), representing the incremental costs per additional patient diagnosed with the correct underlying cause of anaemia. To obtain further insight as to testing for which underlying causes of anaemia (i.e. IDA, ACD, renal anaemia and 'other') potentially has the most room for improvement in terms of cost-effectiveness, subgroup analyses were performed.

Model inputs

The incidence of the different underlying causes of anaemia in Dutch patients aged ≥ 50 years, was based on the abovementioned database [16]. The results of the survey were used to calculate the probability that the right underlying cause of anaemia (according to the expert panel) was established within each of those patient categories, for both work-ups. The likelihoods that GPs chose a certain type of treatment based on the different underlying causes of anaemia (i.e. prescribe medication (including the type of medication), refer the patient to secondary care, perform follow-up of the patient or to close the consultation without performing additional actions) were also derived from this survey. In cases where the GP decided to hold a follow-up appointment with the patient or to close the consultation, the probability of spontaneous recovery of a patient's anaemia was taken into account for each of the different underlying causes. A detailed overview of the model structure, input parameters and the assumptions used is provided in Appendix 1. Values for input parameters that could not be obtained from literature, such as the duration of oral iron supplementation, were derived from expert elicitations with two internist-haematologists (Appendix 1).

Table 1 (on right page): detailed overview of costs and effects of both the routine and the extensive laboratory work-up, for the different underlying causes of anaemia. In addition, costs of initial lab (at GP presentation, maximum of two times), costs of referrals, costs of medication and costs of other laboratory tests (performed during follow-up, to monitor medication, or during referral) are presented. Percentages may not add up to exactly 100% due to rounding. ACD; anaemia of chronic disease; CI: confidence interval, IDA: iron deficiency anaemia.

Table 1	%	Work-up	Probability right cause (95% CI)	Costs initial lab (95% CI)	Costs referral (95% CI)	Costs medication (95% CI)	Costs other lab (95% CI)	Total costs (95% CI)
IDA	19%	Routine work-up	69.26% (57.51% to 80.03%)	€94.51 (€70.70 to €123.36)	€613.52 (€457.00 to €798.41)	€37.64 (€8.16 to €82.85)	€91.28 (€72.84 to €112.21)	€836.95 (€664.11 to €1035.74)
		Extensive work-up	70.13% (58.64% to 80.28%)	€93.63 (€75.88 to €114.36)	€643.31 (€479.02 to €830.42)	€33.28 (€7.26 to €73.48)	€94.75 (€75.95 to €116.15)	€864.98 (€689.65 to €1063.49)
		Effect	0.87% (-14.82% to 16.42%)	-€0.88 (-€10.56 to €7.49)	€29.79 (-€54.22 to €114.20)	-€4.36 (-€12.40 to -€0.11)	€3.47 (-€7.06 to €15.76)	€28.03 (-€61.85 to €119.65)
		Routine work-up	44.48% (35.92% to 53.11%)	€94.51 (€70.70 to €123.36)	€496.59 (€374.31 to €645.54)	€21.93 (€8.02 to €43.18)	€174.65 (€135.57 to €220.56)	€787.67 (€636.73 to €963.29)
ACD	32%	Extensive work-up	54.56% (46.20% to 62.96%)	€93.63 (€75.88 to €114.36)	€543.67 (€420.41 to €691.52)	€14.14 (€3.23 to €30.96)	€170.23 (€134.13 to €213.60)	€821.68 (€676.97 to €989.12)
		Effect	10.08% (-2.01% to 22.12%)	-€0.88 (-€10.56 to €7.49)	€47.08 (€2.75 to €90.54)	-€7.79 (-€14.98 to -€3.15)	-€4.42 (-€15.44 to €5.29)	€34.00 (-€24.82 to €89.92)
		Routine work-up	52.82% (39.64% to 66.27%)	€94.51 (€70.70 to €123.36)	€664.56 (€492.23 to €873.25)	€0.00 (€0.00 to €0.00)	€173.22 (€137.18 to €216.25)	€932.29 (€736.12 to €1158.71)
		Extensive work-up	65.21% (51.06% to 78.12%)	€93.63 (€75.88 to €114.36)	€606.01 (€441.41 to €801.95)	€0.00 (€0.00 to €0.00)	€150.71 (€119.42 to €187.93)	€850.35 (€668.07 to €1062.52)
Renal anaemia	11%	Effect	12.39% (-06.82% to 31.15%)	-€0.88 (-€10.56 to €7.49)	-€58.56 (-€147.27 to €28.53)	€0.00 (€0.00 to €0.00)	-€22.51 (-€45.63 to -€0.60)	-€81.94 (-€192.83 to €26.24)
		Routine work-up	43.50% (36.21% to 51.05%)	€94.51 (€70.70 to €123.36)	€561.78 (€437.78 to €706.01)	€23.53 (€8.03 to €47.14)	€184.25 (€147.26 to €227.85)	€864.06 (€716.99 to €1026.68)
		Extensive work-up	47.80% (39.63% to 55.74%)	€93.63 (€75.88 to €114.36)	€549.54 (€429.11 to €692.74)	€24.84 (€10.98 to €45.13)	€184.66 (€147.42 to €229.00)	€852.67 (€711.57 to €1012.88)
		Effect	4.30% (-6.84% to 15.12%)	-€0.88 (-€10.56 to €7.49)	-€12.24 (-€58.69 to €34.80)	€1.32 (-€3.43 to €5.79)	€0.41 (-€12.13 to €13.34)	-€11.39 (-€70.81 to €48.56)
Other	39%	Routine work-up	49.64% (44.76% to 54.46%)	€94.51 (€70.70 to €123.36)	€561.92 (€445.11 to €694.93)	€23.10 (€7.39 to €47.22)	€162.62 (€131.77 to €198.46)	€842.15 (€704.35 to €993.96)
		Extensive work-up	56.01% (51.20% to 60.75%)	€93.63 (€75.88 to €114.36)	€571.37 (€454.80 to €706.01)	€20.32 (€6.91 to €40.69)	€159.58 (€130.05 to €194.17)	€844.90 (€710.65 to €993.89)
		Effect	6.37% (-0.56% to 13.10%)	-€0.88 (-€10.56 to €7.49)	€9.44 (-€20.03 to €39.19)	-€2.78 (-€7.15 to -€0.06)	-€3.04 (-€9.96 to €3.80)	€2.75 (-€34.72 to €39.86)
Overall								

Probabilistic sensitivity analyses

Random samples were simultaneously drawn for all input parameters based on predefined parameter distributions. Distributions were parameterized based on the observed parameter mean and on the observed or assumed standard error (Appendix 1). To determine the effect of joint uncertainty in all input parameters on model outcomes, a probabilistic sensitivity analysis (PSA) was performed based on a Monte Carlo simulation with 10,000 samples

One-way sensitivity analysis

To identify which individual parameters substantially influence model outcomes, a one-way deterministic sensitivity analysis was conducted. For each parameter, the impact on total costs per patient resulting from a change from the base case value to the lower and to the upper limit for the corresponding 95% confidence interval was analysed. A detailed overview of the inputs used in the one-way sensitivity analysis is provided in Appendix 1.

RESULTS

Overall cost-effectiveness

As shown in Table 1, the routine laboratory work-up costs €842 (95% CI: €704 to €994) per patient, as compared with €845 (95% CI: €711 to €994) for the extensive laboratory work-up (an increase of €3 per patient [95% CI: €-35 to €40], i.e. +0.3%). Compared with the routine work-up, the extensive work-up showed a trend of an increase in the percentage of patients diagnosed with the correct underlying condition of anaemia from 49.6% (95% CI: 44.8% to 54.5%) to 56.0% (95% CI: 51.2% to 60.8%), i.e. +6.4% (95% CI: -0.6% to 13.1%). This resulted in an ICER of €43 per additional patient diagnosed with the correct underlying cause of anaemia. Estimating that 57,000 patients aged ≥50 years present with a new anaemia in Dutch general practices annually (i.e. no anaemia in preceding two years) [22, 23], this can result in an increase of around 3,600 patients who are diagnosed both earlier and with the correct underlying cause of anaemia, at an additional cost of €156,000/year.

Diagnosis specific cost-effectiveness

The results of subgroup analyses indicated that average per patient costs were expected to increase by €28 (95% CI: €-62 to €120) among IDA patients and by €34 (95% CI: €-25 to €90) among ACD patients. In contrast, among renal anaemia patients and patients with 'other' underlying causes of anaemia, those costs are expected to decrease by €82 (95% CI: €-193 to €26) and by €11 (95% CI: €-71 to €49), respectively. The percentage of patients for whom the correct underlying cause is established is expected to increase by 0.9% (95% CI: -14.8% to 16.4%) among IDA patients, by 10.1% (95% CI: -2.0% to 22.1%) among ACD patients, by 12.4% (95% CI: -6.8% to 31.2%) among renal anaemia patients, and by 4.3% (95% CI: -6.8% to 15.1%) among patients with 'other' underlying causes. Table 1 shows a detailed overview of all model outcomes, sorted by underlying cause of anaemia. The incremental cost-effectiveness plane

demonstrating the overall result of 10,000 model simulations is shown in Figure 2. This figure indicates that 44.2% of the model simulations resulted in lower total costs of the extensive work-up as compared with the routine work-up. The incremental cost-effectiveness plane for the subgroups of anaemia patients (i.e. IDA, ACD, renal anaemia, and 'other') is shown in Appendix 1. The impact of separately inserting the inputs of the internist-haematologists on model outcomes as opposed to using their averaged estimates (as in the base case analysis) was considered negligible (Appendix 1). In addition, 56.2% of the model simulations indicated lower costs in the extensive work-up (data not shown), when only considering costs of the initial GP consultation and accompanying phlebotomy and laboratory analyses (with a maximum of requesting laboratory tests twice in the routine work-up).

Figure 2. Incremental cost-effectiveness plane showing the impact of the extensive laboratory work-up as compared with the routine laboratory work-up on the difference in the percentage of correctly diagnosed underlying causes of anaemia, as well as the difference in costs per patient, for 10,000 model simulations.

One-way sensitivity analysis

In Figure 3, the results of one-way sensitivity analyses are shown for the 10 parameters with the highest impact on the difference in costs. The difference in costs between both strategies was found to be most sensitive to changes in the frequency with which GPs correctly diagnose the underlying cause of anaemia, in both the extensive and the routine work-up. The costs per GP consultation and the probability of spontaneous recovery of ACD have the highest impact on the difference in costs between both work-ups.

Figure 3. Tornado diagram showing the impact of changes in the most relevant input parameters on the difference in costs. ACD: anaemia of chronic disease; GP = general practitioner; IDA: iron deficiency anaemia; RA: renal anaemia.

CONCLUSION AND DISCUSSION

The use of an extensive laboratory work-up likely increases the percentage of patients diagnosed with the correct underlying cause of anaemia as compared with the routine work-up. Simultaneously, a very minor increase in costs of ~€3/patient (+0.3%) was expected. Whereas the improvement in diagnosis was quite likely (chance of an increase in correct diagnosis equals 6.4%), the exact effect of this extensive work-up on the changes in costs was quite uncertain (chance that the extensive work-up would actually save costs equals 44.2%). Although higher costs were expected among IDA and ACD patients, most cost savings were expected to be achieved among patients with renal and 'other' types of anaemia.

9

Despite the probably increase of 6.4% in correct diagnoses, results varied considerably between the different underlying causes of anaemia. The largest improvements in this probability were expected in ACD and renal anaemia patients (i.e. +10.1% and +12.4%, respectively). However, costs of referrals were expected to increase by €47/patient among ACD patients, whereas those costs decreased by €59/patient among renal anaemia patients. The increase in costs among ACD patients was explained by the higher immediate referral rate to secondary care (i.e. referral following the initial GP consultation) after the extensive work-up rather than after the routine work-up (38% vs. 26%), while the probability of follow-up was lower (48% vs. 60%). As the probability of spontaneous recovery among ACD patients was estimated to be relatively high (i.e. 38%) higher referral rates lead to higher costs in the extensive work-up in this patient category. Combined with the relatively high frequency of ACD (32% of all newly diagnosed anaemia patients in general practice), this may strongly impact the overall costs. Therefore, effort should be spent on deciding which patient management strategy (i.e. referral or follow-up) results in most improvements in the patient's quality of life, at the lowest costs.

For renal anaemia patients, the results indicate that 42% and 48% of those patients were immediately referred to secondary care in the routine versus the extensive work-up, which was considered the only appropriate management decision according to the expert panel (Appendix 1). Thus, besides expected improvements in diagnosing renal anaemia due to the extensive work-up, this work-up also increased the number of patients who would immediately receive an effective management strategy, which likely decreases costs further downstream the care pathway.

In IDA patients, the probability of a correct diagnosis remained almost unchanged in the extensive as compared with the routine work-up (i.e. +0.9%), while costs increased with €28/patient. When considering the management decisions made in those patients, it was found that this increase in costs was mostly attributable to an expected 6% increase in immediate referrals (43% vs. 37%). However, the DCGP-guideline recommends a colonoscopy and/or gastroscopy in IDA patients aged >50 years, to exclude a gastrointestinal malignancy [24]. As such malignancies can be detected in 6-15% of IDA patients [25-30], this increase in immediate referrals may increase the probability that a gastrointestinal malignancy is diagnosed. Thereby, this increase in costs will likely enhance rapid initiation of adequate treatment, potentially improving treatment effectiveness. However, as such a referral decision should be based on a patient's clinical signs and symptoms (which were unknown in the current analysis), the real-life impact of this extensive work-up on the diagnosis and treatment of gastrointestinal malignancies remains to be investigated.

Although the previously performed effectiveness analysis distinguished only four underlying causes of anaemia (IDA, ACD, renal anaemia, and 'other') [19], the current study further divided the 'other' category into suspected bone marrow disease, vitamin B₁₂ or folic acid deficiency, and unknown causes. Subsequently, only the remaining patients were categorized as 'other'. Although those additional subgroups were too small for any demonstrable results, this subdivision allowed for a more precise calculation on the costs of these diagnoses. However, it is reasonable to assume that GPs are unable to diagnose these less common causes of anaemia based solely on laboratory test results: they often require further diagnostic testing or referral to secondary care to determine these causes. Therefore, in order to allow an accurate cost estimation, the approach taken to analyse the effectiveness in the current study differed slightly from the approach taken in the previously published article. Consequently, the result from the current study indicate a slightly lower percentage of correctly diagnosed underlying causes when compared with the previous study [19].

Strengths

The results are expected to provide a good representation of the Dutch population aged ≥50 years with newly diagnosed anaemia in primary care, because the cases in the survey were based on real-life patient data, the incidence of the various causes of anaemia in the survey

correspond with their occurrence in daily practice [19], and because the participating GPs were representative of the GP population in the Netherlands [19].

As 96.4% of the 10,000 model simulations indicated that the extensive laboratory work-up would increase the percentage of patients correctly diagnosed, this result was robust to uncertainty in input parameters. Of those simulations, 44.2% indicated lower total costs with the use of an extensive work-up, although there was an average increase of ~€3 per patient, overall, indicating that the exact impact on costs is likely very limited but remains uncertain. However, the average number of tests performed in the extensive work-up doubled those performed in the routine work-up (14 vs. 7). That the increase in costs remains so small can easily be explained as the costs of additional diagnostic tests are offset by performing all tests during one GP visit and one phlebotomy, thereby preventing repeated blood sampling (involving additional costs of GP visits, the order tariff for requesting laboratory tests, and lost productivity among patients). This is also confirmed by the results, as 55.4% of 10,000 model simulations indicated that the costs of diagnostic testing at the GP were actually lower in the extensive work-up.

Limitations

This study has certain limitations. First, as described previously [19], the GPs were not provided with the patient's anamnesis, medical history and physical examination. It is therefore likely that the accuracy of the diagnoses of the responding GPs may be higher in real-life. Although this limitation was present in all cases within the study, their potential effect on the differences between the two analysis methods is most likely limited, although it cannot be excluded that the abovementioned patient characteristics may have affected the tests that would have been requested by the GPs. Furthermore, it is uncertain to what extent GPs' diagnostic and treatment decisions in the survey may differ from real life.

9

Secondly, as mentioned previously, the costs of treating anaemia in secondary care have not been included in the model because of large differences in treatments for the different underlying causes and the lack of patient-level data. However, a delayed correct diagnosis likely delays the initiation of adequate treatment, negatively affects quality of life [3] and potentially increases treatment costs owing to an increased severity of anaemia. Therefore, it was conservatively assumed that costs after establishing the correct diagnosis will not differ between patients. Consequently, current results are likely an underestimation of the potential additional benefits provided by the extensive laboratory work-up.

Third, as information for some model input parameters could not be obtained from literature, expert estimates had to be used. Although the number of experts was limited ($n = 2$), the results of probabilistic (Appendix 1) and one-way sensitivity analyses indicated that the impact of changes in model parameters (as based on those expert elicitations) on model outcomes was limited.

Fourth, in the Netherlands, 64% of the clinical chemistry laboratories offer reflex testing [31]. In reflex testing, GPs do not decide themselves which laboratory tests to perform in suspected anaemia patients, but instead request 'anaemia analysis'. The laboratory will then perform a predefined set of tests, and sequentially perform additional tests if the initial tests indicate the presence of anaemia [31]. In addition, 27% of these laboratories provide an interpretative comment along with the test result [31]. As this reflex testing could not be incorporated in the current study, the proportion of patients correctly diagnosed in current clinical practice (reflected by the routine work-up) may be underestimated. However, the set of tests performed in reflex testing differs strongly between laboratories [31] and often involves fewer diagnostic tests compared with the extensive work-up presented in the current study. In addition, although the DCGP-guidelines recommend reflex testing in patients with newly diagnosed anaemia [24], it is not yet offered by all laboratories and can therefore not yet be considered common practice among GPs. Thus, although the added benefit of the extensive work-up may be slightly overestimated in the current analysis, it likely still increases the number of patients for whom the underlying cause is correctly diagnosed by the GP. Furthermore, the insights obtained from this study are likely of added value to decide upon the optimal combination of tests to be used for reflex testing, which may increase similarity in diagnostic work-ups between laboratories.

Implications for practice

As subgroup analyses revealed that the added value of the extensive laboratory work-up depends on the underlying cause of anaemia, further research into this variation is recommended. Which tests contribute most to establishing the correct underlying cause should be investigated, in order to select a laboratory work-up that can correctly diagnose the majority of patients with minimal inconvenience (i.e. multiple venipunctures) and at minimal costs.

Transferability of these results to other countries is not straightforward, as the work-up regarding the diagnosis of anaemia patients and the attributed costs may vary greatly between countries. Populating the model with country-specific data would support reliable country-specific estimations of the cost-effectiveness of both work-ups.

In conclusion, although the extensive laboratory work-up is usually more effective for diagnosing the underlying cause of anaemia by GPs, it is not always more cost-effective than the routine work-up. Nevertheless, the impact on costs was found to be minimal. Given that the extensive work-up results in additional benefits which could not be captured in the current analysis (i.e. in terms of a faster diagnosis which may improve a patient's quality of life, and reduced productivity losses among anaemia patients and their caregivers), using an extensive laboratory work-up is expected to be cost-effective.

ACKNOWLEDGMENTS

We thank all of the GPs for their participation in the survey, and the internist-hematologist for participating in the expert elicitation. In addition, we thank Dr. Rebecca Buis for critical reading of the manuscript.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ETHICAL APPROVAL

The database used in the analysis was approved by the internal ethics committee of the Albert Schweitzer Hospital.

CONTRIBUTORSHIP:

MMAK, AS, KS, SD, MDL and RK were involved in the conception and design of the study and performed the data collection. MMAK and HK designed the health economic model and analysed the data. All authors were involved in interpreting the data. MMAK drafted the manuscript, and all author authors were major contributors in critically reviewing the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Carmel, R., Anemia and aging: an overview of clinical, diagnostic and biological issues. *Blood Rev*, 2001. 15(1): p. 9-18.
2. Riva, E., et al., Association of mild anemia with hospitalization and mortality in the elderly: the Health and Anemia population-based study. *Haematologica*, 2009. 94(1): p. 22-8.
3. Lucca, U., et al., Association of mild anemia with cognitive, functional, mood and quality of life outcomes in the elderly: the "Health and Anemia" study. *PLoS One*, 2008. 3(4): p. e1920.
4. Robinson, B., Cost of anemia in the elderly. *J Am Geriatr Soc*, 2003. 51(3 Suppl): p. S14-7.
5. Smith, R.E., Jr., The clinical and economic burden of anemia. *Am J Manag Care*, 2010. 16 Suppl Issues: p. S59-66.
6. Shavelle, R.M., R. MacKenzie, and D.R. Paculdo, Anemia and mortality in older persons: does the type of anemia affect survival? *Int J Hematol*, 2012. 95(3): p. 248-56.
7. Zakai, N.A., et al., A prospective study of anemia status, hemoglobin concentration, and mortality in an elderly cohort: the Cardiovascular Health Study. *Arch Intern Med*, 2005. 165(19): p. 2214-20.
8. Steensma, D.P. and A. Tefferi, Anemia in the elderly: how should we define it, when does it matter, and what can be done? *Mayo Clin Proc*, 2007. 82(8): p. 958-66.
9. Culleton, B.F., et al., Impact of anemia on hospitalization and mortality in older adults. *Blood*, 2006. 107(10): p. 3841-6.
10. Penninx, B.W., et al., Anemia in old age is associated with increased mortality and hospitalization. *J Gerontol A Biol Sci Med Sci*, 2006. 61(5): p. 474-9.
11. Droogendijk, J., et al., Adherence to the Dutch general practitioner anemia guideline - a retrospective cohort study. *Perfusion*, 2013. 6(26): p. 192-198.
12. Guralnik, J.M., et al., Prevalence of anemia

- in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. *Blood*, 2004. 104(8): p. 2263-8.
13. de Moraes, N.S. and M.S. Figueiredo, Challenges in the diagnosis of iron deficiency anemia in aged people. *Rev Bras Hematol Hemoter*, 2017. 39(3): p. 191-192.
 14. Makipour, S., B. Kanapuru, and W.B. Ershler, Unexplained anemia in the elderly. *Semin Hematol*, 2008. 45(4): p. 250-4.
 15. Fonseca, C., et al., Prevalence of anaemia and iron deficiency in Portugal: the EMPIRE study. *Intern Med J*, 2016. 46(4): p. 470-8.
 16. Stouten, K., et al., Relevance of mean corpuscular volume for the evaluation of anaemia in general practice. Manuscript submitted, 2017.
 17. Oosterhuis, W.P., et al., Prospective comparison of the flow chart for laboratory investigations for anaemia from the Dutch College of General Practitioners' guideline 'Anaemia' with a self-developed, substantive and logistical alternative flow chart'. *Ned Tijdschr Geneesk*, 2007. 151: p. 2326-32.
 18. Kolnaar, B.G., et al., [Summary of the Dutch College of General Practitioners' practice guideline 'Anemia']. *Ned Tijdschr Geneesk*, 2003. 147(40): p. 1956-61.
 19. Schop, A., et al., The effectiveness of a routine versus an extensive laboratory analysis in the diagnosis of anemia in general practice. Manuscript submitted, 2017.
 20. LimeSurvey Project Team / Schmitz, C., LimeSurvey: An Open Source survey tool. 2015, LimeSurvey Project: Hamburg, Germany.
 21. Zorginstituut Nederland, Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2015. p. 1-51.
 22. NIVEL. Wie heeft welke gezondheidsproblemen? Incidenties en prevalenties. 2014 [cited 2014 July]; Available from: <http://www.nivel.nl/NZR/wie-heeft-welke-gezondheidsproblemen>.
 23. Centraal Bureau voor de Statistiek. Bevolking; geslacht, leeftijd en burgerlijke staat, 1 januari. 2017 [cited 2017 March 3]; Available from: <http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=7461BEV&D1=0&D2=0&D3=101-130&D4=0,65-66&HDR=G3,T&ST-B=G1,G2&VW=T>.
 24. Nederlands Huisartsen Genootschap. NHG-Standaard Anemie. 2014; Available from: <https://www.nhg.org/standaarden/volledig/nhg-standaard-anemie>.
 25. Liu, K. and A.J. Kaffes, Iron deficiency anaemia: a review of diagnosis, investigation and management. *Eur J Gastroenterol Hepatol*, 2012. 24(2): p. 109-16.
 26. Droogendijk, J., et al., Screening for gastrointestinal malignancy in patients with iron deficiency anemia by general practitioners: an observational study. *Scand J Gastroenterol*, 2011. 46(9): p. 1105-10.
 27. Ho, C.H., et al., Predictive risk factors and prevalence of malignancy in patients with iron deficiency anemia in Taiwan. *Am J Hematol*, 2005. 78(2): p. 108-12.
 28. James, M.W., et al., Risk factors for gastrointestinal malignancy in patients with iron-deficiency anaemia. *Eur J Gastroenterol Hepatol*, 2005. 17(11): p. 1197-203.
 29. Killip, S., J.M. Bennett, and M.D. Chambers, Iron deficiency anemia. *Am Fam Physician*, 2007. 75(5): p. 671-8.
 30. Rasul, I. and G.P. Kandel, An approach to iron-deficiency anemia. *Can J Gastroenterol*, 2001. 15(11): p. 739-47.
 31. Verboeket - van de Venne, W.P.H.G., et al., Anemieprotocollen voor de eerste lijn in Nederland, in 63e Congres van de Nederlandse Vereniging voor Klinische Chemie en Laboratoriumgeneeskunde. 2010, Ned Tijdschr Klin Chem Labgeneesk: Veldhoven. p. 91-133.

APPENDIX 1

Extensive description of the methods used

Distribution of the questionnaire

In order to obtain a representative sample of Dutch GPs, the survey was distributed among 836 GPs affiliated to one of 13 participating GP associations, located in seven provinces of the Netherlands (i.e. Overijssel, Noord-Brabant, Gelderland, Limburg, Zeeland, Zuid-Holland and Noord-Holland). All GPs were invited by email ($n = 836$). Participation was voluntary and all responses were processed anonymously.

Database used in the questionnaire

The cases used in the questionnaire were derived from a large prospective database of patients with anaemia, in whom the previously mentioned extensive laboratory work-up was performed ($n = 3,325$) [1]. Inclusion criteria of patients were: aged ≥ 50 years, and being diagnosed with new anaemia (i.e. no diagnosis of anaemia in the previous two years) in general practice from February 1st, 2007 until February 1st, 2015. A detailed explanation of this database has been described previously [16]. The probability of each underlying cause of anaemia, as well as the probability that the correct underlying cause of anaemia is established (in both the routine and the extensive laboratory work-up), is shown in Table 1.

Structure of the model

According to Dutch health economic guidelines, a societal perspective was taken in the cost-effectiveness analysis [2]. This means that all costs were included from the moment a patient with anaemia presents to the GP, until the patient has received the correct type of medication in primary care, or has been diagnosed with the correct underlying cause of anaemia (in primary or secondary care). As recommendations regarding treatment prescriptions in primary care as well as when to refer patients to secondary care are provided in the DCGP guideline, this time horizon (roughly ranging from 1 to 200 days) allows a comprehensive evaluation of the costs and effects that are related to the work-up of GPs in anaemia patients.

Table 1 (right page). Input parameters concerning patient pathways used in the model, showing the parameter, the value used in the model, the 95% confidence interval, the distribution used, and the data source. ACD = anaemia of chronic disease, BMD = bone marrow disease, CI = confidence interval, FA = folic acid, GP = general practitioner, IDA = iron deficiency anaemia

Parameter	Category	Probability (95% CI)	Distribution	Reference
Prevalence	IDA	18.7% (17.2% to 20.2%)	Dirichlet	Database [1]
	ACD	31.8% (30.0% to 33.5%)	Dirichlet	Database [1]
	renal anaemia	10.9% (9.7% to 12.1%)	Dirichlet	Database [1]
	unknown	30.9% (29.2% to 32.7%)	Dirichlet	Database [1]
	possible bone marrow disease	1.6% (1.1% to 2.1%)	Dirichlet	Database [1]
	vitamin B ₁₂ /folic acid deficiency	2.7% (2.1% to 3.3%)	Dirichlet	Database [1]
	other	3.4% (2.8% to 4.2%)	Dirichlet	Database [1]
Probability right cause established by GP after routine lab request	IDA	69.2% (57.5% to 80.0%)	Beta	Data questionnaire [3]
	ACD	44.4% (35.9% to 53.1%)	Beta	Data questionnaire [3]
	renal anaemia	52.8% (39.6% to 66.3%)	Beta	Data questionnaire [3]
	unknown	47.7% (39.1% to 56.4%)	Beta	Data questionnaire [3]
	possible bone marrow disease	20.0% (0.6% to 60.4%)	Beta	Data questionnaire [3]
	vitamin B ₁₂ /folic acid deficiency	46.2% (20.8% to 71.7%)	Beta	Data questionnaire [3]
	other	14.3% (1.9% to 36.0%)	Beta	Data questionnaire [3]
Probability right cause established by GP after extensive lab request	IDA	70.1% (58.6% to 80.3%)	Beta	Data questionnaire [3]
	ACD	54.5% (46.2% to 63.0%)	Beta	Data questionnaire [3]
	renal anaemia	65.2% (51.1% to 78.1%)	Beta	Data questionnaire [3]
	unknown	51.8% (42.5% to 61.1%)	Beta	Data questionnaire [3]
	possible bone marrow disease	40.0% (7.0% to 81.1%)	Beta	Data questionnaire [3]
	vitamin B ₁₂ /folic acid deficiency	66.7% (28.2% to 94.6%)	Beta	Data questionnaire [3]
	other	0.0% (0.0% to 0.0%)	Beta	Data questionnaire [3]

Patient pathways

In case the GP decides to follow-up the patient or to close the consultation, a probability of spontaneous recovery of a patient's anaemia was taken into account for the different underlying causes of anaemia. As no evidence regarding those probabilities could be retrieved from literature, these were derived from expert elicitations of two internist-hematologists. The use of such expert elicitations in health economic models has been described previously [4, 5]. Both experts were asked to estimate the mode, and the lower and upper bound of the 95% confidence interval for these probabilities. Following this, a distribution was defined to use the combined estimates of both experts in the cost-effectiveness analysis, and the impact of using each of their estimates was evaluated in separate model analyses.

In all patients in whom the anaemia does not resolve spontaneously, or in all patients in whom inappropriate medication is prescribed (which does not resolve the anaemia), it is assumed that this patient will have a reconsultation at the GP within a few weeks. Following this, the GP requests additional laboratory tests, and subsequently prescribes medication or refers the patient to secondary care. The ratio of 'prescribing medication' and 'referral' remains unchanged, but the options 'close consultation' and 'follow-up' are no longer included, as it is assumed that immediate action is taken when the initial treatment decision proved not effective, or when the anaemia proved not to resolve spontaneously. An overview of correct treatments (specified for the different underlying causes of anaemia, according to the expert panel, is provided in Table 2).

Cost data

Costs of laboratory tests and other types of diagnostic or imaging tests (i.e. colonoscopy, gastroscopy, electrocardiogram, X-thorax, computed tomography (CT) scan, echo, bone marrow biopsy, and/or kidney biopsy) were obtained from the Dutch Healthcare Authority [6], and from a laboratory supporting primary care [7]. Costs of consultations with a GP or a medical specialist, costs of lost productivity, as well as travel costs to the GP and to the hospital were obtained from the Dutch costing manual [8]. Medication costs were obtained from tariffs published by the Dutch National Health Care Institute [9]. A detailed overview of all cost data used is shown in Table 3.

Resource use

No patient level data were available regarding medication prescriptions and regarding the activities performed in patients referred to secondary care. Therefore, the dose, type and frequency of medication administered, were based on the DCGP guideline and combined with expert opinions [11]. Response rates to oral iron and vitamin B₁₂ were based on estimations by the two experts. For all input parameters that were (partly) based on expert opinions, the average of those estimates will be used in the base case model analysis. In addition, the impact of applying each of those expert's inputs separately, as well as the averaged estimates, on

model outcomes will be evaluated. Regarding the activities performed in referred patients, it was assumed that average costs are similar in patients who are correctly or initially incorrectly diagnosed. The types and frequencies of those activities were also derived from expert opinions (see Table 4).

The average return travel time to the GP, combined with a ten minute consultation, was assumed to be 0.8 hours. This time was assumed to be 1.3 hours for a consultation with a specialist in the hospital. The (additional) estimated duration of productivity loss caused by diagnostic or therapeutic interventions performed, were based on patient information folders, assuming an average waiting time of ten minutes [12-20]. An overview of all model input parameters concerning resource use is provided in Table 5.

Table 2. This table provides an overview of the treatments prescribed by the general practitioners, that are considered as correct or incorrect treatment according to the expert panel for the different underlying causes of anaemia. Red indicates incorrect treatment decisions, while green indicates correct treatment decisions. In the column medication, the correct type of medication is specified. Thus, iron is the only correct type of medication in IDA patients, and antibiotics are considered correct treatment when prescribed in ACD patients with a suspected infection. In patients with folic acid deficiency, folic acid should be prescribed, while patients with vitamin B₁₂ deficiency require treatment with vitamin B₁₂. ACD = anaemia of chronic disease, IDA = iron deficiency anaemia.

			Treatment general practitioner			
			Close consultation	Referral to secondary care	Prescribe medication	Follow-up
Underlying cause of anaemia	IDA			Iron		
	ACD			Antibiotics (infection)		
	Renal anaemia					
	Other	Folic acid / vitamin B ₁₂ deficiency			Suitable for deficiency	
	Suspected bone marrow disease					
	Unknown					
	Other types					

Table 3. input parameters concerning costs used in the model, showing the parameter, the value used in the model, the 95% confidence interval, the distribution used, and the data source. CI = confidence interval, CT = computed tomography, NA = not applicable, NZA = Dutch Healthcare Authority (Nederlandse Zorg Autoriteit).

*Are provided along with the haemoglobin result, without additional costs.

Parameter	Category	Probability (95% CI)	Distribution	Reference
Laboratory costs	Order tariff	€9.33 (€5.30 to €14.37)	Gamma	NZA 2016 [6]
	Erythrocyte sedimentation rate (ESR)	€1.75 (€1.00 to €2.72)	Gamma	NZA 2016 [6]
	C-reactive protein (CRP)	€3.75 (€2.15 to €5.77)	Gamma	NZA 2016 [6]
	Haemoglobin (Hb)	€1.51 (€0.87 to €2.33)	Gamma	NZA 2016 [6]
	Haematocrit (Ht)*	NA	NA	-
	Mean corpuscular volume (MCV)*	NA	NA	-
	Reticulocytes	€2.33 (€1.35 to €3.62)	Gamma	NZA 2016 [6]
	Leukocytes	€1.60 (€0.92 to €2.47)	Gamma	NZA 2016 [6]
	Thrombocytes	€1.50 (€0.85 to €2.34)	Gamma	NZA 2016 [6]
	Differential cell count	€1.68 (€0.97 to €2.62)	Gamma	NZA 2016 [6]
	Creatinine and modification of diet in renal disease	€1.53 (€0.88 to €2.37)	Gamma	NZA 2016 [6]
	Lactate dehydrogenase (LDH)	€1.43 (€0.82 to €2.18)	Gamma	NZA 2016 [6]
	Alanine-aminotransferase (ALAT)	€1.68 (€0.96 to €2.58)	Gamma	NZA 2016 [6]
	Calcium	€1.41 (€0.81 to €2.16)	Gamma	NZA 2016 [6]
	Ferritin	€6.58 (€3.76 to €10.13)	Gamma	NZA 2016 [6]
	Serum iron	€1.91 (€1.09 to €2.95)	Gamma	NZA 2016 [6]

Parameter	Category	Probability (95% CI)	Distribution	Reference
Laboratory costs (continued)	Transferrin	€4.23 (€2.41 to €6.49)	Gamma	NZA 2016 [6]
	Folic acid	€7.63 (€4.39 to €11.72)	Gamma	NZA 2016 [6]
	Vitamin B ₁₂	€6.32 (€3.61 to €9.82)	Gamma	NZA 2016 [6]
	Sodium	€1.52 (€0.86 to €2.36)	Gamma	NZA 2016 [6]
	Potassium	€1.51 (€0.85 to €2.31)	Gamma	NZA 2016 [6]
	Urea	€1.32 (€0.76 to €2.03)	Gamma	NZA 2016 [6]
	Bilirubin	€1.53 (€0.89 to €2.37)	Gamma	NZA 2016 [6]
	Gamma-glutamyltransferase (GGT)	€1.60 (€0.91 to €2.48)	Gamma	NZA 2016 [6]
	M-protein	€74.58 (€42.69 to €114.64)	Gamma	Salstro [7]
	Protein spectrum	€16.13 (€9.30 to €25.03)	Gamma	NZA 2016 [6]
	Phosphate	€1.23 (€0.71 to €1.89)	Gamma	NZA 2016 [6]
	Albumin	€1.26 (€0.71 to €1.96)	Gamma	NZA 2016 [6]
	Haptoglobin	€5.82 (€3.31 to €8.96)	Gamma	NZA 2016 [6]
	Creatinine in 24-hour urine	€1.53 (€0.88 to €2.38)	Gamma	NZA 2016 [6]
	Protein in 24-hour urine	€1.48 (€0.84 to €2.27)	Gamma	NZA 2016 [6]
	Anti-parietal cell antibodies	€25.46 (€14.62 to €39.41)	Gamma	NZA 2016 [6]
	Anti-intrinsic factor antibodies	€12.29 (€6.93 to €18.99)	Gamma	NZA 2016 [6]
	Anti-transglutaminase (anti-TTG) + total immunoglobulin A (IgA)	€24.58 (€13.94 to €37.98)	Gamma	NZA 2016 [6]

Parameter	Category	Probability (95% CI)	Distribution	Reference
Consultation and hospital stay	Consultation with general practitioner	€33.30 (€18.84 to €51.21)	Gamma	Dutch cost manual 2015 [8]
	Hospital visit, including consultation with specialist (without admission)	€91.84 (€52.71 to €142.75)	Gamma	Dutch cost manual 2015 [8]
	Cost per hour hospital stay	€138.40 (€79.61 to €213.98)	Gamma	Dutch cost manual 2015 [8]
	Costs of administering vitamin B ₁₂ injection by GP's assistant	€4.53 (€2.57 to €7.00)	Gamma	NZA 2016 [6]
Diagnostics	Colonoscopy	€375.16 (€214.55 to €576.59)	Gamma	NZA 2016 [6]
	Gastroscopy	€261.45 (€149.25 to €404.60)	Gamma	NZA 2016 [6]
	Electrocardiogram (ECG)	€9.59 (€5.44 to €14.78)	Gamma	NZA 2016 [6]
	X-thorax	€41.12 (€23.36 to €63.43)	Gamma	NZA 2016 [6]
	CT thorax	€172.29 (€100.04 to €267.23)	Gamma	NZA 2016 [6]
	CT abdomen	€167.63 (€95.65 to €259.85)	Gamma	NZA 2016 [6]
	Echo abdomen	€85.28 (€48.98 to €131.52)	Gamma	NZA 2016 [6]
	Echo kidneys	€85.28 (€48.76 to €129.82)	Gamma	NZA 2016 [6]
	Kidney biopsy	€134.98 (€77.31 to €207.00)	Gamma	NZA 2016 [6]
	Pathologic investigation of kidney biopsy	€201.33 (€115.50 to €309.39)	Gamma	NZA 2016 [6]
	Bone marrow biopsy	€198.06 (€112.72 to €309.10)	Gamma	Jeroen Bosch Hospital and NZA 2016 [6]
	Cytogenetic testing	€1,459.04 (€835.97 to €2,255.17)	Gamma	NZA 2016 [6]

Parameter	Category	Probability (95% CI)	Distribution	Reference
Medication	Ferrofumarate (per tablet)	€0.02 (€0.01 to €0.03)	Gamma	www.medicijn-kosten.nl [9]
	Ferinject (per injection)	€318.00 (€182.80 to €491.97)	Gamma	www.medicijn-kosten.nl [9]
	Cyanocobalamin (per tablet)	€0.20 (€0.11 to €0.31)	Gamma	www.medicijn-kosten.nl [9]
	Cyanocobalamin (per injection)	€1.72 (€0.99 to €2.65)	Gamma	www.medicijn-kosten.nl [9]
	Folic acid (per tablet)	€0.05 (€0.03 to €0.08)	Gamma	www.medicijn-kosten.nl [9]
	Antibiotic course (21 days)	€4.93 (€2.86 to €7.55)	Gamma	www.medicijn-kosten.nl [9]
Travel costs	To GP	€0.42 (€0.24 to €0.65)	Gamma	Dutch cost manual 2015 [8]
	To hospital (including parking costs)	€5.71 (€3.28 to €8.86)	Gamma	Dutch cost manual 2015 [8]
Productivity costs	Average productivity costs per hour	€10.30 (€5.83 to €15.64)	Gamma	Dutch cost manual 2009 and 2015 [8, 10]

Table 4. Overview of expert estimates ($n = 2$) regarding resource use parameters. The experts were asked to provide the most common value, as well as the 95% confidence interval of these values by providing a high and a low estimate. ACD = anaemia of chronic disease, BMD = bone marrow disease, CT = computed tomography, ECG = electrocardiogram, GP = general practitioner, IDA = iron deficiency anaemia.

Parameter	Description	Low estimate			Most common value			High estimate		
		Expert 1	Expert 2	Average	Expert 1	Expert 2	Average	Expert 1	Expert 2	Average
Medication duration (tablets, in days)	Ferrofumaraat - treatment	30	30	30	70	45	57.5	90	90	90
	Vitamin B ₁₂	30	30	30	100	84	92	365	120	242.5
	Folic acid	30	30	30	90	84	87	365	120	242.5
Medication doses per day	Ferrofumaraat	1	1	1	3	3	3	3	3	3
	Vitamin B ₁₂	1	1	1	1	1	1	1	1	1
	Folic acid	0.5	0.5	0.5	5	1	3	5	5	5
Probability vitamin B ₁₂ oral, and response oral medication (%)	Oral iron	80	85	83	90	90	90	100	95	98
	Vitamin B ₁₂ injections (start with)	90	20	55	95	50	73	99	90	95
	Vitamin B ₁₂ response to oral medication	50	80	65	90	90	90	99	95	97
Medication injections (n) Spontaneous haemoglobin recovery during follow-up (%)	Ferinject	1	1	1	1	1	1	2	3	2.5
	Vitamin B12	6	5	5.5	10	9	9.5	20	10	15
	IDA patients	0	5	3	10	10	10	20	15	18
	ACD patients	20	20	20	40	35	38	60	45	53
	Renal anaemia	0	5	3	10	10	10	20	15	18
	Vitamin B ₁₂ deficiency	0	1	1	10	5	8	20	10	15
	Folic acid deficiency	0	10	5	10	15	13	20	20	20
	Other and unknown	20	15	18	30	20	25	40	30	35
Probability of activities among referrals (%)	Colonoscopy among patients referred for suspected IDA	40	50	45	100	70	85	100	80	90
	Gastroscopy among patients referred for suspected IDA	40	50	45	100	70	85	100	80	90
	Colonoscopy at request GP (no consultation in hospital)	25	10	18	50	25	38	60	35	48

Probability of activities among referrals (continued) (%)	Gastroscopy at request GP (no consultation in hospital)	25	10	18	35	25	30	45	35	40
	X-thorax among suspected ACD patients	50	70	60	75	90	83	90	100	95
	Echo abdomen among suspected ACD patients	50	50	50	75	65	70	90	75	83
	CT thorax among suspected ACD patients	40	10	25	70	25	48	90	35	63
	CT abdomen among suspected ACD patients	40	10	25	70	25	48	90	35	63
	Echo of the kidneys among suspected renal anaemia patients	70	30	50	85	50	68	100	60	80
	ECG among suspected renal anaemia patients	25	0	13	50	5	28	75	10	43
	X-thorax among suspected renal anaemia patients	25	10	18	50	15	33	75	20	48
	Kidney biopsy (using echo) among suspected renal anaemia patients	5	0	3	10	3	7	20	6	13
	Bone marrow biopsy among suspected BMD patients	50	50	50	75	75	75	90	90	90
	Cytogenetic testing in suspected BMD	25	35	30	38	60	49	50	75	63
	Gastroscopy among patients with vitamin B ₁₂ deficiency	20	10	15	40	25	33	50	35	43

Table 5. Input parameters concerning resource use that is used in the model, showing the parameter, the value used in the model, the 95% confidence interval, the distribution used, and the data source. ACD = anaemia of chronic disease, BMD = bone marrow disease, CI = confidence interval, CT = computed tomography, ECG = electrocardiogram, GP = general practitioner, IDA = iron deficiency anaemia, RA = renal anaemia.

Parameter	Category	Probability (95% CI)	Distribution	Reference
Medication duration (days)	Ferrofumarate	57.5 (29.2 to 86.7)	Normal	Expert opinions
	Vitamin B ₁₂	92.0 (32.3 to 152.9)	Normal	Expert opinions
	Folic acid	87.0 (30.3 to 144.8)	Normal	Expert opinions
Medication doses per day (n)	Ferrofumarate	3.0 (1.0 to 4.0)	Normal	Expert opinions
	Vitamin B ₁₂	1.0 (1.0 to 1.0)	NA	Expert opinions
	Folic acid	3.0 (1.0 to 5.1)	Normal	Expert opinions
Vitamin B ₁₂ oral, and response oral medication (%)	Oral iron	90.0% (85.2% to 94.8%)	Beta PERT	Expert opinions
	Vitamin B ₁₂ – start with injections	72.5% (61.2% to 86.1%)	Beta PERT	Expert opinions
	Vitamin B ₁₂ – response oral medication	90.0% (75.6% to 95.4%)	Beta PERT	Expert opinions
Medication infusion/ injection (n)	Ferinject	1.0 (0.3 to 2.5)	Normal	Expert opinions
	Vitamin B ₁₂	9.5 (5.3 to 13.8)	Normal	Expert opinions
Follow-up (n)	Follow-up laboratory tests if medication is prescribed	2.0 (2.0 to 2.0)	NA	Assumption
	Follow-up laboratory tests without further activities	1.0 (1.0 to 1.0)	NA	Assumption
Spontaneous anaemia recovery during follow-up (%)	Iron deficiency anaemia	10.0% (5.3% to 14.7%)	Beta PERT	Expert opinions
	Anaemia of chronic disease	37.5% (26.5% to 47.3%)	Beta PERT	Expert opinions
	Renal anaemia	10.0% (5.4% to 14.7%)	Beta PERT	Expert opinions

Parameter	Category	Probability (95% CI)	Distribution	Reference
Spontaneous anaemia recovery during follow-up (%) (continued)	Vitamin B ₁₂ deficiency	7.5% (3.1% to 12.3%)	Beta PERT	Expert opinions
	Folic acid deficiency	12.5% (7.8% to 17.3%)	Beta PERT	Expert opinions
	Other and unknown	25.0% (20.1% to 31.2%)	Beta PERT	Expert opinions
Frequency of performing lab (n)	Routine lab	1.4 (1.4 to 1.5)	Normal	Data questionnaire [3]
	Extensive lab	1.0 (1.0 to 1.0)	NA	Intervention (all analyses at once)
Activities among referrals (%)	Cytogenetic testing in (suspected) bone marrow disease	48.8% (37.3% to 57.6%)	Beta PERT	Expert opinions
	Colonoscopy among referred patients with (suspected) IDA	85.0% (61.6% to 89.2%)	Beta PERT	Expert opinions
	Gastroscopy among referred patients with (suspected) IDA	85.0% (61.9% to 89.2%)	Beta PERT	Expert opinions
	Colonoscopy at request GP in patients with (suspected) IDA, without specialist consultation	37.5% (25.8% to 44.3%)	Beta PERT	Expert opinions
	Gastroscopy at request GP in patients with (suspected) IDA, without specialist consultation	30.0% (22.2% to 36.4%)	Beta PERT	Expert opinions
	X-thorax among (suspected) ACD patients	82.5% (68.7% to 91.2%)	Beta PERT	Expert opinions
	Echo abdomen among (suspected) ACD patients	70.0% (58.0% to 78.4%)	Beta PERT	Expert opinions
	CT thorax among (suspected) ACD patients	47.5% (33.7% to 57.6%)	Beta PERT	Expert opinions
	CT abdomen among (suspected) ACD patients	47.5% (33.6% to 57.6%)	Beta PERT	Expert opinions
	Echo of kidneys among (suspected) RA patients	67.5% (56.8% to 75.6%)	Beta PERT	Expert opinions
	ECG among (suspected) RA patients	27.5% (18.1% to 37.0%)	Beta PERT	Expert opinions

Parameter	Category	Probability (95% CI)	Distribution	Reference
Activities among referrals (%) (continued)	X-thorax among (suspected) RA patients	32.5% (22.8% to 41.9%)	Beta PERT	Expert opinions
	Kidney biopsy (using echo) among (suspected) RA patients	6.5% (3.8% to 10.4%)	Beta PERT	Expert opinions
	Bone marrow biopsy among (suspected) BMD patients	75.0% (60.0% to 85.2%)	Beta PERT	Expert opinions
	Gastroscopy among patients with (suspected) vitamin B ₁₂ deficiency	32.5% (21.9% to 39.3%)	Beta PERT	Expert opinions
Productivity lost (h)	To hospital	1.3 (0.8 to 2.1)	Gamma	Assumption
	To GP's office	0.8 (0.5 to 1.3)	Gamma	Assumption
	Ferinject infusion	2.0 (1.2 to 3.1)	Gamma	Patient flyer [12]
	ECG	0.3 (0.2 to 0.5)	Gamma	Patient flyer [13]
	Colonoscopy	8.0 (4.6 to 12.2)	Gamma	Patient flyer [15]
	Gastroscopy	8.0 (4.5 to 12.3)	Gamma	Patient flyer [18]
	X-thorax	0.3 (0.2 to 0.5)	Gamma	Patient flyer [19]
	Echo abdomen	0.5 (0.3 to 0.8)	Gamma	Patient flyer [17]
	Echo kidneys	0.5 (0.3 to 0.8)	Gamma	Patient flyer [17]
	CT thorax	0.3 (0.2 to 0.5)	Gamma	Patient flyer [16]
	CT abdomen	0.3 (0.2 to 0.5)	Gamma	Patient flyer [16]
	Kidney biopsy	16.0 (9.2 to 24.8)	Gamma	Patient flyer [20]
	Bone marrow biopsy	1.0 (0.6 to 1.5)	Gamma	Patient flyer [14]

ASSUMPTIONS USED IN THE HEALTH ECONOMIC MODEL

Assumptions related to right and wrong treatment decisions:

The following treatment decisions were considered right for patients with the following underlying causes of anaemia (according to the expert panel):

- Iron deficiency anaemia:
 - › Medication (iron)
 - › Referral to secondary care
- Anaemia of chronic disease:
 - › Referral to secondary care
 - › Follow-up (seeing the patient again in approximately 4 weeks)
- Renal anaemia:
 - › Referral to secondary care
- Vitamin B₁₂ / folic acid deficiency:
 - › Medication (vitamin B₁₂ or folic acid deficiency, depending on the underlying cause)
 - › Referral to secondary care
- Suspected bone marrow disease:
 - › Referral to secondary care
- Unknown:
 - › Referral to secondary care
 - › Follow-up (seeing the patient again in approximately 4 weeks)
- Other:
 - › Referral to secondary care
 - › Follow-up (seeing the patient again in approximately 4 weeks)

Assumptions related to costs used in the health economic model:

- It is assumed that the patient will not pay parking costs when visiting the general practitioner, but will pay parking costs when visiting a medical specialist, or undergoing investigations in a hospital.
- It is assumed that all patients aged 50 to 65 years have a paid job, and that all patients aged >65 years are retired. Therefore, productivity losses are assumed to occur in all patients aged 50 to 65 years, but not in patients aged >65 years.

Assumptions related to resource use applied in the health economic model:

- In case a GP decided to prescribe medication, it was assumed that for iron GPs would start with oral medication, and that injections would only be administered in case the patient would not respond sufficiently to oral therapy.
- For vitamin B₁₂, experts estimated that 72.5% of the GPs would prescribe injections immediately, while the remaining GPs would start with oral therapy. For folic acid, only oral therapy is prescribed.

- Regarding the activities performed among referred patients, it is assumed that those costs are (on average) the same per patient that is correctly or incorrectly diagnosed.
- The waiting time for diagnostic interventions and a consultation with the general practitioner or medical specialist within a hospital was assumed to be ten minutes.
- The travel time to the hospital was assumed to be 30 minutes, while the travel time to the general practitioner was assumed to be 15 minutes.
- The duration of a consultation with a general practitioner or a medical specialist within a hospital was assumed to be 10 minutes.
- It is assumed that blood samples are only taken at one point in time to diagnose patients with suspected bone marrow disease, and that all laboratory analyses are performed based on this single blood withdrawal.
- When a patient is referred to secondary care, this patient is assumed to have on average 1.8 consultations with a medical specialist.

Assumptions related to decisions regarding diagnosing the underlying cause of anaemia:

- It is assumed that the set of laboratory tests that is provided does not enable general practitioners to distinguish between suspected bone marrow disease, vitamin B₁₂ deficiency, and other underlying causes of anaemia. Therefore, those three subgroups, including the patients in whom general practitioners state that they are unable to determine the underlying cause (in the questionnaire), are all categorized in the group 'other'.
- It is assumed that all patients who present with anaemia to their GP are eventually correctly diagnosed. Patients that initially receive the wrong diagnosis are assumed not to be treated properly, and not referred for the appropriate investigations in secondary care, and therefore either present again at the GP within a few weeks, or are referred to the correct specialism after they have initially been referred to the wrong specialism.

Assumptions related to treatment decisions:

- It is assumed that antibiotics, prescribed by the GP, are considered the right treatment ($n = 2$) for a patient with anaemia of chronic disease. For those cases, the costs of a 21-day course of Amoxicillin and Clavulanic acid are incorporated in the model.
- It was assumed that Prednisone could not be correctly prescribed solely based on the laboratory test results that were provided to the GPs. Thus, although two GPs reported to prescribe Prednisone, it was assumed that first a follow-up of those patients was needed. Therefore, those GPs were assumed to follow-up the patient instead of prescribing medication at this point in time.
- It is assumed that 50% of the colonoscopies and gastroscopies performed were self-handled by the general practitioner.
- In case a patient presents (again) to the general practitioner because the anaemia did not

resolve spontaneously, it is assumed that the patient will undergo the same laboratory investigations as initially, in order for the general practitioner to determine a new diagnostic and therapeutic strategy.

- It is assumed that all patients that are prescribed medication undergo laboratory testing to check the response to therapy.
- In case a patient does not recover spontaneously after the general practitioner decided not to take further action (i.e. close consultation), it is assumed that the patient will come back at the general practitioner's office, and that the same laboratory tests will be performed as have been performed initially. Following this, it is assumed that those additional diagnostic test will lead to a referral to the correct specialism, or to prescribing the correct type of medication. The probabilities of prescribing medication and referring the patient to secondary care are proportional to the initial probabilities of those two decisions in this patient category.

Assumptions related to treatment decisions in patients diagnosed with an incorrect underlying cause of anaemia:

- It is assumed that the referral decision is based on the diagnosis that is set by the general practitioner. Consequently, if the general practitioner decides to refer a patient to secondary care, but the suspicion of the underlying cause of the anaemia in this particular patient is incorrect, the specialism to which this patient is referred will also be incorrect.
- As a consequence of this referral to the incorrect specialism, it is assumed that this particular patient is not diagnosed with the correct underlying cause of anaemia during this referral. Instead, they are subsequently referred to the correct specialism, where they undergo additional investigations which will result in a correct diagnosis.
- In case a general practitioner first makes the wrong decision regarding which type of medication to prescribe, the patient is expected to return to the general practitioner (unless the anaemia resolves spontaneously), and that this patient will then be prescribed the right treatment (which may consist of medication or referral, depending on the underlying cause of anaemia).

SUPPLEMENTARY RESULTS

Results of subgroup analysis

Figure 1. Incremental cost-effectiveness plane, showing the 95% confidence ellipses of the impact of the extensive laboratory work-up as compared to the routine laboratory work-up for the different underlying causes of anaemia, on the percentage of underlying causes that are correctly established, as well as the difference in costs per patient, for 10,000 model simulations. ACD = anaemia of chronic disease, IDA = iron deficiency anaemia.

Results of expert elicitations

As shown in Figure 2, the impact on model outcomes of inserting each expert’s inputs separately, when compared to using the averaged estimates (as used in the base case analysis), was considered negligible.

Figure 2. Incremental cost-effectiveness plane, showing the impact of the extensive laboratory work-up as compared to the routine laboratory work-up on the difference in the percentage of correctly diagnosed underlying causes of anaemia, as well as the difference in costs per patient, for 10,000 model simulations, while using each of the expert’s estimates in the model.

REFERENCES

1. Stouten, K., et al., Prevalence of potential underlying aetiology of macrocytic anaemia in Dutch general practice. *BMC Fam Pract*, 2016. 17(1): p. 113.
2. Zorginstituut Nederland, Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2015. p. 1-51.
3. Schop, A., et al., ANNALS EXPRESS: The effectiveness of a routine versus an extensive laboratory analysis in the diagnosis of anemia in general practice. *Ann Clin Biochem*, 2017: p. 4563217748680.
4. Kip, M.M., et al., Using expert elicitation to estimate the potential impact of improved diagnostic performance of laboratory tests: a case study on rapid discharge of suspected non-ST elevation myocardial infarction patients. *J Eval Clin Pract*, 2016.
5. Haakma, W., et al., Belief elicitation to populate health economic models of medical diagnostic devices in development. *Appl Health Econ Health Policy*, 2014. 12(3): p. 327-34.
6. Nederlandse Zorgautoriteit, Tarieventabel DBC zorgproducten en overige producten per 1 januari 2016 (bijlage bij TB-CU-2066-01). 2016.
7. Saltro, D.i.d.z. Tarievenoverzicht. 2016 [cited 2016 December 6]; Available from: <https://www.saltro.nl/tarieven/tarievenoverzicht/>.
8. Hakkaart-van Roijen, L., et al., Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. 2015, Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam: Rotterdam. p. 1-120.
9. Zorginstituut Nederland. Medicijnkosten. 2016 [cited 2016 July 4]; Available from: <https://www.medicijnkosten.nl/>.
10. Hakkaart-van Roijen, L., S.S. Tan, and C.A.M. Bouwmans, Handleiding voor kostenonderzoek; Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. 2010, Institute for Medical Technology Assessment, Erasmus University Rotterdam. p. 1-127.
11. Nederlands Huisartsen Genootschap. NHG-Standaard Anemie. 2014; Available from: <https://www.nhg.org/standaarden/volledig/nhg-standaard-anemie>.
12. Albert Schweitzer Ziekenhuis, Behandeling van ijzertekort: met een Ferinject infuus. 2016: p. 8.
13. Bernhoven, Elektrocardiogram. 2016. p. 2.
14. Albert Schweitzer Ziekenhuis, Een beenmergpunctie/botbiopsie. 2015. p. 4.
15. Albert Schweitzer Ziekenhuis, Colonoscopie: voorbereiding met Picoprep/Picolax. 2016. p. 20.
16. Albert Schweitzer Ziekenhuis, CT-scan: zonder contrast via de bloedbaan. 2016. p. 8.
17. Albert Schweitzer Ziekenhuis, Echografie. 2016. p. 8.
18. Albert Schweitzer Ziekenhuis, Gastroscoopie. 2016. p. 16.
19. Maasstad ziekenhuis, Patiënteninformatie hart / longfoto (thoraxfoto). 2015. p. 1.
20. Maasstad ziekenhuis, Patiënten informatie: nierbiopsie. 2015: p. 8.

PART IV

Expert elicitation, stakeholder preferences
and checklist development concerning
the evaluation of diagnostic tests
and biomarkers

10

Using expert elicitation to estimate the potential impact of improved diagnostic performance of laboratory tests: a case study on rapid discharge of suspected non-ST elevation myocardial infarction patients

Michelle M.A. Kip
Lotte M.G. Steuten
Hendrik Koffijberg
Maarten J. IJzerman
Ron Kusters

This chapter has been published as:

M.M.A. Kip, L.M.G. Steuten, H. Koffijberg, M.J. IJzerman, R. Kusters. *Using expert elicitation to estimate the potential impact of improved diagnostic performance of laboratory tests: a case study on rapid discharge of suspected non-ST elevation myocardial infarction patients.* J Eval Clin Pract, 2016.

Reprinted with permission.

ABSTRACT

Early health technology assessment can provide insight in the potential cost-effectiveness of new tests to guide further development decisions. This can increase their potential benefit but often requires evidence which is lacking in early test development stages. Then, expert elicitation may be used to generate evidence on the impact of tests on patient management. This is illustrated in a case study on a new triple biomarker test (copeptin, heart-type fatty acid binding protein, and high-sensitive troponin [HsTn]) at hospital admission. The elicited evidence enables estimation of the impact of using the triple biomarker on time to exclusion of non ST-elevation myocardial infarction compared with current serial HsTn measurement (performed 0, 2 and 6 h after admission). Cardiologists were asked to estimate the effect of the triple biomarker on patient's discharge rates and interventions performed, depending on its diagnostic performance. This elicited evidence was combined with Dutch reimbursement data and published evidence into a decision analytic model. Direct hospital costs and patients' discharge rates were assessed for 3 testing strategies including this triple biomarker (ie, only at admission or combined with HsTn measurements after 2 and 6 h). Direct hospital costs of suspected non-ST elevation myocardial infarction patients using serial HsTn measurements are estimated at €1825 per patient. Combining this triple biomarker with HsTn measurements after 2 and 6 hours is expected to be the most cost-effective strategy. Depending on the diagnostic performance of the triple biomarker, this strategy is estimated to reduce costs with €66 to €205 per patient (ie, 3.6%-11.3% reduction). Expert elicitation can be a valuable tool for early health technology assessment to provide an initial estimate of the cost-effectiveness of new tests prior to their implementation in clinical practice. As demonstrated in our case study, improved diagnostic performance of the triple biomarker may have benefits that should be further explored.

INTRODUCTION

Over the last several decades, there has been a major increase in both the utilization of existing laboratory tests and the development of new tests [1, 2]. Although new laboratory diagnostics may provide additional information for clinicians and patients, their development also raises concerns about overutilization, potential negative effects of overtesting on patient outcomes, and rising health care costs [3, 4]. Therefore, the potential benefits of new tests, in terms of health outcomes and costs, should be thoroughly investigated prior to their implementation in clinical practice [5-12].

When the evaluation of treatment strategies is compared with diagnostic testing, the most important difference is that tests influence health outcomes indirectly, by changing the clinical decisions made and management of the disease guided by the test results [9, 12]. Consequently, to estimate the potential added value of a new test, the role of the new test in the existing diagnostic pathway first needs to be assessed. Bossuyt et al proposed 3 roles of the new test: (1) replacing an existing test in the diagnostic pathway (eg, because the new test is more accurate); (2) preceding the pathway as a triage test, to safely rule out patients for further testing; or (3) succeeding the pathway as an add-on (eg, to reduce the proportion of false-positive test results) [6]. The diagnostic performance of a new test will depend to a large extent on its intended role, and on the performance of other tests it is combined with [6]. For example, if it is used as an add-on, the combined performance of all tests should be determined, recognizing that the performance of the new test depends on the selection of patients it is actually used in (ie, patients with positive preceding tests).

Evidence concerning the effect of new tests (alone or in combination) on a patient's diagnostic pathway is commonly very limited in early stages of product development. A randomized controlled trial is accepted to provide the highest quality of evidence to quantify the consequences of laboratory testing strategies, but is often not feasible for ethical, financial, or other reasons, in particular in early stages of product development [8, 13-15]. In those situations, Van den Bruel et al suggest that evidence on changes in diagnosis or intended management by the physician, before and after the test result has been disclosed, may serve as a proxy [9]. In addition, such evidence may be used to inform the potential cost-effectiveness of a new test and the drivers of diagnostic performance. One way to deal with this would be to perform a model-based early health technology assessment where part of the evidence is determined by expert elicitations [16-18]. Elicitation involves a process by which experts formulate a quantitative judgement based on their own beliefs, independent of the quality of such knowledge, for an uncertain quantity [19]. Those elicitations can be used as input for a decision analytic model. Such models are particularly helpful in early health economic modeling, where uncertain and unknown priors are used to populate a model. Such early assessments can guide decisions about whether or not to continue test development, and ideally to optimize this process [20, 21].

In this paper, we will therefore illustrate how an expert's judgment about improved test performance can be incorporated into an early model for estimating cost-effectiveness. The case is about adding a copeptin and heart-type fatty acid binding protein (H-FABP) test to conventional serial high-sensitivity troponin (HsTn) measurement, to allow rapid exclusion of acute myocardial infarction (specifically non-ST elevation myocardial infarction [NSTEMI]) in the coronary pain unit (CPU), which is a specialized emergency unit for patients presenting with chest pain. The described approach can further guide test development and optimization and can serve as initial evidence on the likely added value of combinations of tests. In addition, this approach can be used to estimate the likely impact of different implementation strategies of this triple biomarker.

Case description

Patients with chest pain represent a large proportion of acute hospitalizations [22, 23], yet distinguishing patients with acute coronary syndrome (ACS) remains a diagnostic challenge. Acute coronary syndrome encompasses unstable angina, ST elevation myocardial infarction (STEMI), and NSTEMI. Although STEMI can be diagnosed on the basis of persistent ST elevation on the electrocardiogram (ECG), in vitro laboratory tests have great diagnostic value to distinguish NSTEMI from both unstable angina and other (noncardiac) causes of chest pain [22, 23].

Currently, HsTn is the preferred laboratory test for diagnosing NSTEMI because it allows both earlier and more sensitive detection of elevated troponin levels compared with the conventional troponin tests [22]. Although this new generation of tests has improved the early diagnosis of NSTEMI, current guidelines still recommend monitoring of the majority of patients for several hours and to perform serial troponin measurements [22, 24]. Although recommended in the guidelines, such prolonged duration of hospital stay accounts for about two-thirds of direct hospital costs in chest pain patients [25]. Early exclusion of NSTEMI can therefore greatly enhance efficient use of hospital resources.

As reported in previous studies, combining a HsTn measurement with other early cardiac markers might contribute to an early and safe discharge [26-35]. Two potential markers that are elevated rapidly after an acute myocardial infarction are copeptin and H-FABP [36-39]. Hence, adding those markers to the conventional HsTn measurement is expected to improve the overall diagnostic performance in the first hours after chest pain onset. This may then increase the diagnostic certainty of cardiologists and facilitate early discharge of those patients. The combination of 3 markers will hereafter be referred to as "triple biomarker."

METHODS

Model description

A decision tree which was populated with a hypothetical sample of 1,000 patients presenting

with symptoms suggestive of ACS at the CPU was constructed. The population eligible for the current simulation was defined as patients (1) with an onset of chest pain in the last 12 hours, (2) not having persistent ST elevation on the ECG, (3) not presenting at the CPU with cardiac arrest requiring immediate reperfusion.

The current diagnostic pathway consists of serial HsTn measurements at regular intervals, ie, at the time of a patient’s presentation at the CPU (t0), and repeated after 2 and 6 hours (t2 and t6, respectively). To analyze the most cost-effective strategy of implementing the triple biomarker, three alternative diagnostic strategies were evaluated in the model: (A) the triple biomarker only at t0 and without follow-up HsTn measurements, (B) the triple biomarker at t0 followed by repeat testing of HsTn at t2, and (C) the triple biomarker at t0 followed by repeat testing of HsTn at t2 and t6 (Figure 1). The decision tree for each of the 3 strategies is provided in Appendix 1.

Figure 1. Diagnostic strategies evaluated in the model. This figure shows the current diagnostic work-up in suspected non-ST elevation myocardial infarction patients, as well as the three alternative diagnostic strategies evaluated. The current work-up consists of serial high-sensitive troponin (HsTn) measurements at 0, 2 and 6 h after hospital admission (t0, t2, and t6, respectively). The three alternative diagnostic strategies involve (A) the triple biomarker only at t0 and without follow-up HsTn measurements, (B) the triple biomarker at t0 followed by repeat testing of HsTn at t2, and (C) the triple biomarker at t0 followed by repeat testing of HsTn at t2 and t6.

Because the diagnostic performance of the triple biomarker in any of these combinations is unknown, the cost-effectiveness of the 3 strategies was evaluated for three potential diagnostic performances of the triple biomarker. This allows prediction of the minimum diagnostic performance that is required for this triple biomarker to be considered cost-effective.

The primary effectiveness measure in the model was defined as the percentage of patients correctly discharged from hospital at each time point (t0, t2, t6, or after overnight admission). Incremental effectiveness was expressed as the difference in the number of patients discharged (per 1,000 patients) when using the triple biomarker compared with serial HsTn measurement

alone. A direct hospital cost perspective was taken, including all costs that occur from the moment a patient presents at the CPU until hospital discharge. Costs were expressed in 2015 euros. Incremental cost was defined as the difference in direct hospital costs per patient. The incremental cost-effectiveness ratio was obtained by dividing the incremental costs by the incremental effectiveness. Because the length of a patient's hospital stay was shorter than 1 year, discounting of costs and effects was not required.

Estimation of diagnostic performance of the markers

We performed a literature review to estimate the diagnostic performance of HsTn measurement at t0. Details of this literature review are described in Appendix 2. Keller et al reported a sensitivity of 98.3% and negative predictive value (NPV) of 99.0% for the combination of HsTn and copeptin [27]. Based on this study, the maximum performance of the triple biomarker that can realistically be achieved after the addition of a third marker was estimated at a sensitivity and NPV of both 99.0%. Furthermore, an increase in sensitivity of the triple biomarker of at least 5% was considered necessary to change clinical management. Therefore, from those assumptions and the results of the literature review, the sensitivity and NPV of HsTn at t0 in patients with suspected NSTEMI were set at 85% and 95%, respectively [27, 40-42]. The three potential diagnostic performances of the triple biomarker that were evaluated in the model were set at a sensitivity of 90%, 95%, and 99%, with a corresponding NPV of 96%, 98% and 99% (Table 1).

Table 1. Sensitivities and negative predictive values (NPV) used in the questionnaire, for both serial high-sensitive troponin measurement (current practice) as well as for the three scenarios for the performance of the triple biomarker. NPV = negative predictive value

Work-up		Sensitivity at t0	NPV at t0
Serial high-sensitive troponin measurement		85%	95%
Triple biomarker	Scenario 1	90%	96%
Triple biomarker	Scenario 2	95%	98%
Triple biomarker	Scenario 3	99%	99%

The specificity of HsTn was derived from a review by Goodacre et al. [41-44]. The specificity of the triple biomarker at t0 (49.8%) was calculated by multiplying the individual specificities of H-FABP (80.0%) [43, 45-52], copeptin (70.0%) [27, 43], and HsTn (89.0%) [41-44, 53]. Those diagnostic performances were used to determine the flow of patients through the model (Table 2). Furthermore, although the specificity of HsTn is expected to decrease during follow-up measurements (at t2 and t6), as caused by the discharge of patients without NSTEMI, the specificity of HsTn was assumed not to change during those follow-up measurements. As a higher specificity at t2 and t6 would improve model outcomes more for the serial HsTn

measurement than for the triple biomarker strategy, because fewer individuals are discharged at t0 and thus more individuals are tested at t2 and t6, this assumption prevents overestimating the potential impact of this triple biomarker.

Table 2. Specificities used to calculate the flow of patients through the model. The specificity of high-sensitivity troponin is assumed not to change at t2 and t6. Abbreviation: H-FABP, heart-type fatty acid binding protein.

Test performance at t0	Specificity	Source
High-sensitive troponin*	89.0%	Review Goodacre et al, 2013 [41-44]
H-FABP	80.0%	Review Goodacre et al, 2013 [43, 45-52]
Copeptin	70.0%	Review Goodacre et al, 2013 [27, 43]
Triple biomarker	49.8%	Calculated [53]

Questionnaire

The improved diagnostic performance as achieved by the triple biomarker might impact the patient’s diagnostic pathway as well as the treatment received. However, the clinical utility of the marker in suspected NSTEMI patients is unknown. First, patients might be discharged earlier, and secondly, they might have to undergo fewer diagnostic and therapeutic interventions. Besides laboratory tests, the most commonly performed interventions considered in this study involve a stress ECG, a catheterization, and drug treatment. We developed a closed-ended paper-and-pencil questionnaire to evaluate the clinical utility of the triple biomarker to affect clinical decisions, as well as how an improved diagnostic performance may change clinical management. In this questionnaire, we presented cardiologists the (hypothetical) result of a HsTn assay at t0, t2, and t6, in patients with suspected ACS. They were then asked to estimate the patient’s discharge rates at each point in time, as well as the interventions to be given. Following this, they estimated to what extent an improvement in diagnostic performance, as achieved by the triple biomarker, would affect those decisions. Answer options to both questions consisted of percentages with a range of 25%, and including 0% and 100% (i.e. 0%, 0%-25%, 25%-50%, 50%-75%, 75%-100%, and 100%). The results from this questionnaire served as input for the health economic model.

The participating cardiologists (n = 10) had on average 17 years of experience in the field of cardiology. Respondents were derived from the Jeroen Bosch Hospital (Den Bosch, the Netherlands), and from Medisch Spectrum Twente (Enschede, the Netherlands). Both are large (730 and 1,070 beds, respectively) teaching hospitals offering acute coronary interventions. As cardiologists were asked for their opinion and experiences, they were told that no wrong answers could be given and that the answers would be treated anonymously. When filling out the questionnaires, no interaction between experts was possible. All percentages were averaged. The full questionnaire is included in Appendix 3.

Assumptions

The decision analytical model contains a few important assumptions. First, it is assumed that STEMI is diagnosed primarily on the basis of the result of the ECG, indicating that the triple biomarker will not be of added value in those patients. Those patients will undergo a coronary intervention immediately and are therefore excluded from this analysis. Second, from the qualitative interviews with four cardiologists, it was assumed that no diagnosis of NSTEMI will be missed at the CPU. Third, it is assumed that none of the patients with a negative HsTn at t0, but a positive H-FABP and/or copeptin will be discharged at this time point. An extensive description of all model assumptions is included in Appendix 4.

Cost data

The direct hospital costs of NSTEMI patients are based on a study by Polder et al. [54]. Unit costs of hospital stay and a patient's therapy were obtained from the Dutch Healthcare Authority and from the cost manual by Hakkaart-van Roijen et al. [55, 56]. Unit costs for medication were obtained from www.medicijnkosten.nl (carbaspirin calcium, heparin, metoprolol, and clopidogrel) [57]. Tariffs for laboratory tests, including HsTn measurement, were derived from the Dutch Healthcare Authority [56]. Costs for the H-FABP test were based on published literature [58], while costs of the copeptin test were based on the laboratory tariff of a Dutch university medical center [59]. Direct hospital costs were estimated by summing all different elements of a patient's hospital stay and the probability of each element occurring. All costs were converted to 2015 euros.

Analysis of the results

Descriptive statistics were used to report means and percentages. Ninety-five percent confidence intervals were determined for the probabilities obtained from the questionnaire using Microsoft Excel 2013. Deterministic sensitivity analysis is applied to test the robustness of the model to changes in one variable. The minimum performance at which the additional upfront costs of the triple biomarker are offset by savings further downstream the diagnostic pathway and in subsequent treatment (ie, break-even point) was estimated. Similarly, the maximum allowable total costs of the two additional tests (ie, H-FABP and copeptin), which could be justified by their expected increased performance as compared with current practice, were assessed. The results were presented for both a negative HsTn level and a slightly elevated HsTn level.

Sensitivity analysis

All probability parameters were assigned a β distribution utilizing a standard error derived from the results of the questionnaire or from literature [60], while all cost parameters were assigned a γ distribution and an assumed standard error of 25%. A probabilistic sensitivity analysis was performed by means of Monte Carlo simulations, using 10,000 model iterations for each of the nine scenarios. Mean outcomes of the probabilistic sensitivity analysis were

used as our base case outcomes.

One-way sensitivity analyses were performed to test the robustness of the model outcomes to changes in input parameters. All model input parameters were varied along their 95% confidence interval. As Keller et al reported a sensitivity of 98.3% and NPV of 99.0% for the combination of HsTn and copeptin [27], this one-way sensitivity analysis was only performed for the scenario concerning a sensitivity and NPV of both 99.0%, and only for the strategy (A, B, or C) that is found to be most cost-effective. The results are visualized in a tornado diagram, showing the effect of changes in input parameters both on costs and on outcomes (ie, the incremental discharge rate).

RESULTS

Figure 2 shows the impact of the triple biomarker as compared with serial HsTn measurement on the effectiveness (expressed as patients’ discharge rates), as elicited by the cardiologists at each point in time (t0, t2, and t6). Results from the questionnaire indicate that in the current diagnostic pathway, cardiologists estimate to discharge 12% of all suspected NSTEMI patients from the CPU following a negative HsTn result at t0. An increase in sensitivity and NPV of the triple biomarker is expected to increase the percentage of patients correctly discharged (Figure 2). In addition, if the performance of the triple biomarker improves, cardiologists may decide not to perform further interventions as shown in Figure 3.

Figure 2. Expected discharge rates according to experts as a function of diagnostic performance. This figure shows the impact of the triple biomarker as compared with serial HsTn measurement on the percentage of patients discharged as indicated by the cardiologists at each point in time (t0, t2, and t6), including the 95% confidence interval. NPV = negative predictive value, sens = sensitivity.

Figure 3. Interventions performed according to experts as a function of diagnostic performance. This figure shows the percentage of patients in whom diagnostic or therapeutic interventions (stress ECG, catheterization, and administration of medication) are performed during their hospital admission, as indicated by the cardiologists, including the 95% confidence interval. ECG = electrocardiogram, NPV = negative predictive value, sens = sensitivity.

If sensitivity and NPV of the triple biomarker would be 99% at t0, cardiologists estimate the discharge of 30% of suspected NSTEMI patients following the laboratory results at t0. If a triple biomarker with this diagnostic performance is combined with HsTn measurements at t2 and t6 (strategy C), the percentage of patients without NSTEMI that is admitted overnight might decrease from 38% in current practice to 27%.

Regardless of the performance of the triple biomarker, this third strategy incurs the highest patients' discharge rates at the lowest cost. In current practice, direct hospital cost per suspected NSTEMI patient at the CPU is estimated to be €1,825. Strategy C could decrease those costs to €1,758 (-3.6%), €1,703 (-6.6%), or €1,618 (-11.3%), depending on the performance of the triple biomarker at t0 (Table 3). In this strategy, a 5% increase in sensitivity of the triple biomarker (at t0) from 85% to 90%, and an accompanying increase in NPV from 95% to 96% is already expected to result in higher patient discharge rates and lower costs compared to serial HsTn measurement. In this same strategy, a sensitivity and NPV of 99% of the triple biomarker with additional HsTn measurements at t2 and t6 (strategy C), are expected to result in an additional discharge of 120 patients per 1,000 following the laboratory results at t6, with accompanying cost savings of €205 per patient. As it is estimated that 137,000 people in the Netherlands each year present at the hospital with chest pain, this strategy could involve costs savings ranging from 9 up to 28 million euros per year, depending on the performance of the triple biomarker [61, 62]. As shown in Table 3, an increase in the performance of the triple biomarker also increases cost savings.

Table 3. Costs and effects of the alternative strategies compared to current serial HsTn analysis.

Table 3. Costs and effects of the alternative strategies compared to current serial HsTn analysis.			Effectiveness (patients discharged)	Incremental effectiveness	Costs (per patient)	Incremental costs	ICER
All patients, n = 1,000 (NSTEMI, n = 84)							
Serial HsTn analysis			616-	-	€ 1,825	-	-
Triple biomarker	Strategy I	Sens. 90%	101	-512	€ 2,217	€ 393	- 768
		Sens. 95%	177	-435	€ 2,119	€ 296	- 679
		Sens. 99%	298	-314	€ 1,972	€ 148	- 473
	Strategy II	Sens. 90%	433	-179	€ 1,944	€ 120	- 672
		Sens. 95%	477	-135	€ 1,873	€ 49	- 365
		Sens. 99%	549	-63	€ 1,764	- € 60	957
	Strategy III	Sens. 90%	671	58	€ 1,758	- € 66	- 1,132
		Sens. 95%	694	81	€ 1,703	- € 121	- 1,488
		Sens. 99%	733	120	€ 1,618	- € 205	- 1,706
All patients with negative HsTn result at t0, n = 815							
Serial HsTn analysis			604	-	€ 1,036	-	-
Triple biomarker	Strategy I	Sens. 90%	97	-507	€ 1,520	€ 484	- 955
		Sens. 95%	166	-438	€ 1,416	€ 380	- 868
		Sens. 99%	275	-328	€ 1,257	€ 221	- 674
	Strategy II	Sens. 90%	421	-183	€ 1,209	€ 173	- 943
		Sens. 95%	458	-146	€ 1,137	€ 101	- 690
		Sens. 99%	520	-84	€ 1,024	- € 12	149
	Strategy III	Sens. 90%	632	28	€ 994	- € 42	- 1,475
		Sens. 95%	651	48	€ 941	- € 95	- 1,990
		Sens. 99%	684	80	€ 859	- € 177	- 2,215
All patients with a slightly elevated HsTn result at t0 without NSTEMI, n = 101							
Serial HsTn analysis			12	-	€ 1,850	-	-
Triple biomarker	Strategy I	Sens. 90%	3	-8	€ 1,776	- € 74	9,042
		Sens. 95%	11	-1	€ 1,651	- € 200	220,990
		Sens. 99%	23	11	€ 1,471	- € 379	- 34,251
	Strategy II	Sens. 90%	12	1	€ 1,589	- € 262	- 321,508
		Sens. 95%	19	7	€ 1,468	- € 383	- 55,597
		Sens. 99%	29	17	€ 1,291	- € 559	- 32,211
	Strategy III	Sens. 90%	38	27	€ 1,477	- € 373	- 13,999
		Sens. 95%	42	30	€ 1,356	- € 494	- 16,207
		Sens. 99%	49	37	€ 1,180	- € 671	- 17,981

A sensitivity of 90%, 95%, and 99% correspond with a negative predictive value of 96%, 985 and 99%, respectively. Abbreviations: HsTn, high-sensitivity troponin; ICER, incremental cost-effectiveness ratio; NSTEMI, non-ST elevation myocardial infarction; Sens, sensitivity.

Figure 4a. Result of one-way sensitivity analysis, showing the difference in the effect on the probability of being discharged more early when using the triple biomarker (strategy C), with a sensitivity and negative predictive value of both 99%, as compared with serial HsTn measurement. All probabilities were varied according to the 95% confidence interval. ECG = electrocardiogram.

Figure 4b. Result of one-way sensitivity analysis, showing the difference in costs of triple biomarker (strategy C), with a sensitivity and negative predictive value of both 99%, as compared with serial HsTn measurement. All probabilities were varied according the 95% confidence interval; all costs were varied 25% below and above the mean. ECG = electrocardiogram, H-FABP = heart-type fatty acid binding protein, HsTn = high-sensitivity troponin, NSTEMI = non-ST elevation myocardial infarction.

After excluding the estimated number of NSTEMI patients in the hypothetical cohort of 1000 patients, the remaining patients are divided into two subgroups, ie, with and without a slightly elevated HsTn level. These subgroup results indicate that the greatest cost savings per patient can be achieved in patients with a slightly elevated HsTn level (Table 3). In both subgroups, strategy C is expected to incur the most cost savings and involve the highest incremental patient discharge rates as compared with the other two strategies.

As the results indicate that strategy C is the most cost-effective, a one-way sensitivity analysis was performed for this strategy, assuming a sensitivity and NPV of the triple biomarker of both 99%. As shown in Figure 4a and b, results indicate that the costs per patient are mainly driven by the specificity of HsTn at t2, the costs of overnight hospital stay, and the specificity of HsTn at t0, while the incremental effect on time to discharge was mainly driven by the specificity of HsTn at t0 and t6. However, those changes in model outcomes regarding both costs and discharge rates would not have changed the overall conclusion.

Figure 5. Ellipse plot. The ellipses cover 95% of the cost-effectiveness outcomes of the triple biomarker strategy as compared with current serial high-sensitivity troponin measurement, for each of the nine scenarios evaluated in the model. Each ellipse is based on 10,000 model iterations using probabilistic sensitivity analysis. All probability parameters were assigned a β distribution utilizing a standard error derived from the results of the questionnaire or from literature, while all cost parameters were assigned a γ distribution and an assumed standard error of 25%. sens = sensitivity.

Figure 5 shows the results of the probabilistic sensitivity analysis for each of the nine scenarios evaluated in the model. For each scenario, an ellipse which covers 95% of the cost-effectiveness outcomes of the triple biomarker strategy as compared with current serial HsTn measurement is shown. Each ellipse is based on 10 000 model iterations of a hypothetical cohort of 1000 patients. Results indicate that, for strategy C, the probability that the triple biomarker strategy dominates current serial HsTn measurement increases from 61.7% to 84.1% and to 96.0%, depending on the performance of the triple biomarker.

To decrease uncertainty around cost estimates, the maximum costs of both H-FABP and copeptin are determined, for the triple biomarker to break even compared with current practice. For a triple biomarker with a sensitivity of 90% and accompanying NPV of 96%, costs of copeptin and H-FABP together may rise to €109 when combined with two follow-up HsTn measurements (strategy C). In case of a sensitivity of 95% and an NPV of 98% of the triple biomarker, those costs may rise to €164. In this same strategy, costs may rise to €250 in case the triple biomarker increases both the sensitivity and NPV to 99%. Thus, when the diagnostic performance would increase, this would in turn enhance rapid discharge rates and lead to cost savings. Therefore, the addition of those two tests would still be cost-effective even when their costs would be higher than assumed in our analysis.

DISCUSSION

In this paper, we illustrate the value of using expert elicitations in the early assessment of new medical tests, using a case of a triple biomarker to inform early exclusion of NSTEMI at the CPU. Previous studies already investigated the diagnostic accuracy in excluding acute myocardial infarction using combinations of those markers [26-35, 49, 63, 64]. However, this study is the first study that provides an initial estimate of the cost-effectiveness of this triple biomarker in suspected NSTEMI patients on referral rates and on subsequent interventions.

In our case study, the results of the expert elicitations indicate that an increase in diagnostic performance achieved by the use of a triple biomarker as compared with conventional serial HsTn measurement alone may lead to earlier patient discharge, a reduction in the interventions performed in those patients and, consequently, a decrease in direct hospital costs. Results indicate that the triple biomarker with additional HsTn measurements after two and six hours (strategy C), is expected to be the most cost-effective strategy compared with serial HsTn measurement. Regardless of the diagnostic performance of the triple biomarker (ie, a sensitivity of 90%, 95%, and 99%, combined with an NPV of 96%, 98%, and 99%, respectively), time until hospital discharge, interventions performed, and hospital costs are expected to be reduced in this strategy. In addition, results of the sensitivity analyses indicate that the uncertainty in model input parameters is unlikely to change those conclusions. Also, as long as the costs of copeptin and H-FABP together do not exceed €109, this strategy is cost saving or cost neutral

from the perspective of the hospital. For reference, one HsTn measurement costs about €9.55, indicating there is large headroom to achieve a cost-saving outcome.

Limitations of the case study analysis

This study has certain limitations. First, we assumed that all patients with NSTEMI are diagnosed as such at the CPU. This assumption was based on open interviews with cardiologists ($n = 5$), who stated that the risk-averse approach in excluding NSTEMI ensures that no patient with this diagnosis is inadvertently discharged from the hospital. In addition, they stated that using this triple biomarker would not affect this because regardless of the performance of a laboratory test, a patient's signs and symptoms are always considered in the decision to discharge a patient. However, further research is necessary to test the validity of those assumptions in clinical practice.

Second, the number of experts who completed the questionnaire was small ($n = 10$). However, a previous systematic review by Johnson et al reports a median sample size of 11 participants included in expert elicitations [65], indicating that our sample is not unusually small. Also, the variation in the probabilities of patients' discharge rates and interventions performed, as derived from the cardiologists' elicitations, was substantial. This variation is often inevitable in early health technology assessment, and incorporating this uncertainty in the model's input parameters ensures that the real effect of an intervention is not systematically overestimated or underestimated. As the model outcome was shown to be robust against changes in input parameters, this uncertainty is unlikely to change our conclusions.

Third, from our literature review, we applied a sensitivity of 85% and an NPV of 95% for the HsTn measurement at t_0 in our questionnaire. However, a more recently published review reported a lower sensitivity of HsTn of 77%, while the authors mention that the diagnostic performance varied strongly because of substantial heterogeneity between studies [43]. Although this may limit the generalizability of the results, this lower sensitivity also increases the potential benefit of the triple biomarker.

Fourth, although this add-on strategy of the triple biomarker increases sensitivity, this comes at the cost of the specificity [6, 63]. In the current study, the disjunctive positivity criterion was used (as described by Felder and Mayrhofer), which implies that the composite test result is positive when at least one of the tests of the triple biomarker is positive [66]. In our case study, this results in a larger proportion of patients with a positive test result at t_0 in case the triple biomarker is used, as compared with when a single HsTn test is used. Consequently, this will prevent discharge in patients with a positive H-FABP and/or copeptin and may simultaneously increase discharge rates in individuals with a triple-negative result. The decrease in specificity, however, may be smaller than assumed in the current study as the real-life specificity of the triple biomarker likely exceeds the assumed specificity as based on multiplying the specificities

of the three individual tests. For example, Lippi et al. [63] found a combined specificity of troponin and H-FABP of 82%, while multiplying the individual specificities resulted in a combined specificity of 78%.

The rapid increase in the number of available laboratory tests complicates decisions regarding test implementation in clinical practice. When high-grade evidence studies (for example, randomized controlled trials) are not available in early stages of laboratory test development, model-based early health technology assessment can be recommended to guide decisions on whether or not to continue test development, for which diagnostic setting and/or in which patient (sub)group the test should be used, and at what cost. The major advantage of such a model-based analyses is that the results can be updated when new evidence becomes available, for example, about real-world effects of the new test on a patient's diagnostic pathway. In our study, we illustrated such an assessment by developing a decision analytic model combining evidence on the expected (combined) diagnostic accuracy of new tests with cardiologists' elicitations on the impact of this test on a patient's diagnostic pathway and subsequent treatment. This study shows the added value of expert elicitations in guiding decisions about whether or not to continue product development, if possible, to optimize this process and to estimate the cost-effectiveness of those tests prior to their implementation in daily practice [20]. However, because guidelines concerning the process of expert elicitation are currently lacking, different approaches are used, which may lead to different results. Further research into the pros and cons of different methods of expert elicitation is therefore recommended to allow standardization of this process.

Our study illustrates how model-based early health technology assessment combined with expert elicitations can be valuable when estimating the potential cost-effectiveness of a laboratory test, at different diagnostic performances. This was achieved by combining evidence on test performance retrieved from literature, with elicitations from clinicians on the impact of new tests on a patient's diagnostic pathway and subsequent treatment. Our case study suggests that a triple biomarker, consisting of a copeptin, H-FAPB, and a HsTn measurement at the time of a patient's admission at the CPU, when combined with follow-up HsTn measurements at t2 and t6, may have the potential to safely inform the decision for early discharge and may reduce the number of interventions performed, resulting in lower costs.

ACKNOWLEDGEMENTS

We thank the cardiologists who participated in the interviews.

REFERENCES

1. Danzon, P.M., W.G. Manning, and M.S. Marquis, Factors affecting laboratory test use and prices. *Health Care Financing Review*, 1984. 5(4): p. 23-32.
2. Mindemark, M. and A. Larsson, Longitudinal trends in laboratory test utilization at a large tertiary care university hospital in Sweden. *Ups J Med Sci*, 2011. 116(1): p. 34-8.
3. van Walraven, C. and C.D. Naylor, Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA*, 1998. 280(6): p. 550-8.
4. Hickner, J., et al., Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. *J Am Board Fam Med*, 2014. 27(2): p. 268-74.
5. Horvath, A.R., et al., From biomarkers to medical tests: the changing landscape of test evaluation. *Clin Chim Acta*, 2014. 427: p. 49-57.
6. Bossuyt, P.M., et al., Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*, 2006. 332(7549): p. 1089-92.
7. Bossuyt, P.M., et al., Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem*, 2012. 58(12): p. 1636-43.
8. Moons, K.G., Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem*, 2010. 56(4): p. 537-41.
9. Van den Bruel, A., et al., The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *J Clin Epidemiol*, 2007. 60(11): p. 1116-22.
10. Ferrante di Ruffano, L., et al., Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*, 2012. 344: p. e686.
11. Lijmer, J.G., M. Leeflang, and P.M.M. Bossuyt, Proposals for a Phased Evaluation of Medical Tests, in *Medical Tests-White Paper Series*. 2009: Rockville (MD).
12. Bossuyt, P.M.M. and K. McCaffery, Additional Patient Outcomes and Pathways in Evaluations of Testing, in *Medical Tests-White Paper Series*. 2009: Rockville (MD).
13. Schunemann, H.J., et al., Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*, 2008. 336(7653): p. 1106-10.
14. Biesheuvel, C.J., D.E. Grobbee, and K.G. Moons, Distraction from randomization in diagnostic research. *Ann Epidemiol*, 2006. 16(7): p. 540-4.
15. Bossuyt, P.M., J.G. Lijmer, and B.W. Mol, Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*, 2000. 356(9244): p. 1844-7.
16. Haakma, W., et al., Belief elicitation to populate health economic models of medical diagnostic devices in development. *Appl Health Econ Health Policy*, 2014. 12(3): p. 327-34.
17. Cao, Q., et al., Probability elicitation to inform early health economic evaluations of new medical technologies: a case study in heart failure disease management. *Value Health*, 2013. 16(4): p. 529-35.
18. Soares, M.O., et al., Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. *Stat Med*, 2011. 30(19): p. 2363-80.
19. O'Hagan, A., et al., *Uncertain Judgements: eliciting experts' probabilities*. 2006, Chichester, UK: John Wiley & Sons.
20. Ijzerman, M.J. and L.M. Steuten, Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Health Econ Health Policy*, 2011. 9(5): p. 331-47.
21. Pietzsch, J.B. and M.E. Pate-Cornell, Early technology assessment of new medical devices. *Int J Technol Assess Health Care*, 2008. 24(1): p. 36-44.
22. Hamm, C.W., et al., ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J*, 2011. 32(23): p. 2999-3054.
23. Kontos, M.C. and R.L. Jesse, Evaluation of the emergency department chest pain patient. *Am J Cardiol*, 2000. 85(5A): p. 32B-39B.
24. SA Heart Association Task Force, Consensus statement on the use of high sensitivity cardiac troponins. *SA Heart*, 2012. 9(3): p. 210-215.
25. Forberg, J.L., et al., Direct hospital costs of chest pain patients attending the emergency department: a retrospective study. *BMC Emerg Med*, 2006. 6: p. 6.
26. Reichlin, T., et al., Incremental value of copeptin for rapid rule out of acute myocardial infarction. *J Am Coll Cardiol*, 2009. 54(1): p. 60-8.
27. Keller, T., et al., Copeptin improves early diagnosis of acute myocardial infarction. *J Am Coll Cardiol*, 2010. 55(19): p. 2096-106.
28. Lipinski, M.J., et al., A systematic review and collaborative meta-analysis to determine the incremental value of copeptin for rapid rule-out of acute myocardial infarction. *Am J Cardiol*, 2014. 113(9): p. 1581-91.
29. McCann, C.J., et al., Investigation of a multi-

- marker approach to the initial assessment of patients with acute chest pain. *Adv Ther*, 2009. 26(5): p. 531-4.
30. Mockel, M., et al., Early discharge using single cardiac troponin and copeptin testing in patients with suspected acute coronary syndrome (ACS): a randomized, controlled clinical process study. *Eur Heart J*, 2015. 36(6): p. 369-76.
 31. Jacobs, L.H., et al., Rapidly rule out acute myocardial infarction by combining copeptin and heart-type fatty acid-binding protein with cardiac troponin. *Ann Clin Biochem*, 2015.
 32. Thelin, J., et al., The combination of high sensitivity troponin T and copeptin facilitates early rule-out of ACS: a prospective observational study. *BMC Cardiovasc Disord*, 2013. 13: p. 42.
 33. Ray, P., et al., Combined copeptin and troponin to rule out myocardial infarction in patients with chest pain and a history of coronary artery disease. *Am J Emerg Med*, 2012. 30(3): p. 440-8.
 34. Meune, C., et al., Combination of copeptin and high-sensitivity cardiac troponin T assay in unstable angina and non-ST-segment elevation myocardial infarction: a pilot study. *Arch Cardiovasc Dis*, 2011. 104(1): p. 4-10.
 35. Reddy, L.L., et al., Troponin T and Heart Type Fatty Acid Binding Protein (h-Fabp) as biomarkers in patients presenting with chest pain. *Indian J Clin Biochem*, 2015: p. 6.
 36. Kleine, A.H., et al., Release of heart fatty acid-binding protein into plasma after acute myocardial infarction in man. *Mol Cell Biochem*, 1992. 116(1-2): p. 155-62.
 37. Gururajan, P., et al., Heart fatty acid binding protein (H-FABP) as a diagnostic biomarker in patients with acute coronary syndrome. *Heart Lung Circ*, 2010. 19(11): p. 660-4.
 38. Gu, Y.L., et al., Comparison of the temporal release pattern of copeptin with conventional biomarkers in acute myocardial infarction. *Clin Res Cardiol*, 2011. 100(12): p. 1069-76.
 39. Maisel, A., et al., Copeptin helps in the early detection of patients with acute myocardial infarction: primary results of the CHOPIN trial (Copeptin Helps in the early detection Of Patients with acute myocardial INfarction). *J Am Coll Cardiol*, 2013. 62(2): p. 150-60.
 40. Kelly, A.M., Performance of a sensitive troponin assay in the early diagnosis of acute myocardial infarction in the emergency department. *Emerg Med Australas*, 2011. 23(2): p. 181-5.
 41. Reichlin, T., et al., Early diagnosis of myocardial infarction with sensitive cardiac troponin assays. *N Engl J Med*, 2009. 361(9): p. 858-67.
 42. Keller, T., et al., Sensitive troponin I assay in early diagnosis of acute myocardial infarction. *N Engl J Med*, 2009. 361(9): p. 868-77.
 43. Goodacre, S., et al., Systematic review, meta-analysis and economic modelling of diagnostic strategies for suspected acute coronary syndrome. *Health Technol Assess*, 2013. 17(1): p. v-vi, 1-188.
 44. Apple, F.S., et al., Use of the Centaur Tnl-Ultra assay for detection of myocardial infarction and adverse events in patients presenting with symptoms suggestive of acute coronary syndrome. *Clin Chem*, 2008. 54(4): p. 723-8.
 45. Mion, M.M., et al., Analytical and clinical performance of a fully automated cardiac multi-markers strategy based on protein biochip microarray technology. *Clin Biochem*, 2007. 40(16-17): p. 1245-51.
 46. McCann, C.J., et al., Novel biomarkers in early diagnosis of acute myocardial infarction compared with cardiac troponin T. *Eur Heart J*, 2008. 29(23): p. 2843-50.
 47. Ilva, T., et al., Early markers of myocardial injury: cTnl is enough. *Clin Chim Acta*, 2009. 400(1-2): p. 82-5.
 48. Hjortshoj, S., S.R. Kristensen, and J. Ravkilde, Diagnostic value of ischemia-modified albumin in patients with suspected acute coronary syndrome. *Am J Emerg Med*, 2010. 28(2): p. 170-6.
 49. Haltern, G., et al., Comparison of usefulness of heart-type fatty acid binding protein versus cardiac troponin T for diagnosis of acute myocardial infarction. *Am J Cardiol*, 2010. 105(1): p. 1-9.
 50. Body, R., et al., A FABP-ulous 'rule out' strategy? Heart fatty acid binding protein and troponin for rapid exclusion of acute myocardial infarction. *Resuscitation*, 2011. 82(8): p. 1041-6.
 51. Di Serio, F., et al., Proteomic approach to the diagnosis of acute coronary syndrome: preliminary results. *Clin Chim Acta*, 2005. 357(2): p. 226-35.
 52. Garcia-Valdecasas, S., et al., Diagnostic and prognostic value of heart-type fatty acid-binding protein in the early hours of acute myocardial infarction. *Acta Cardiol*, 2011. 66(3): p. 315-21.
 53. Kanchanaraks, S., Evaluation of diagnostic and screening tests: validity and reliability, J.H. University, Editor. 2008.
 54. Polder, J.J., J.J. Barendregt, and H. van Oers, Health care costs in the last year of life—the Dutch experience. *Soc Sci Med*, 2006. 63(7): p. 1720-31.
 55. Hakkaart-van Roijen, L., S.S. Tan, and C.A.M. Bouwmans, Handleiding voor kostenonderzoek; Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. 2010, Institute for Medical Technology Assessment, Erasmus University Rotterdam. p. 1-127.
 56. Nederlandse Zorgautoriteit. Tarievenlijst Eerstelijnsdiagnostiek, Bijlage 1 bij tariefbeschikking TB/CU-7041-03 van 1 april 2013. 2013 [cited 2014 March 3]; Available from: <http://>

www.nza.nl/1048076/1048155/Bijlage_bij_TB_CU_7041_03_Tarieflijst_eerstelijnsdiagnostiek.pdf.

57. Zorginstituut Nederland. Medicijnkosten. 2013 [cited 2013 September 8]; Available from: <http://www.medicijnkosten.nl/>.
58. Vaidya, A., et al., High-sensitive troponin T assay for the diagnosis of acute myocardial infarction: an economic evaluation. *BMC Cardiovasc Disord*, 2014. 14: p. 77.
59. Laboratoriumgeneeskunde, N.V.v.K.C.e. Wie doet wat database. [cited 2016 June 7]; Available from: <https://www.nvkc.nl/professional/wie-doet-wat-database?action=showdetail&id=6331>.
60. Briggs, A. and K. Claxton, *Decision Modelling for Health Economic Evaluation*, in NY. 2006, Oxford University Press: New York.
61. Westwood, M., et al., High-sensitivity troponin assays for the early rule-out or diagnosis of acute myocardial infarction in people with acute chest pain: a systematic review and cost-effectiveness analysis. *Health Technol Assess*, 2015. 19(44): p. 1-234.
62. Vaartjes, I., et al., Hart- en vaatziekten in Nederland. , in *Hart- en vaatziekten in Nederland 2013, cijfers over leefstijl, risicofactoren, ziekte en sterfte*, Hartstichting, Editor. 2013, Hartstichting: Den Haag. p. 18.
63. Lippi, G., C. Mattiuzzi, and G. Cervellin, Critical review and meta-analysis on the combination of heart-type fatty acid binding protein (H-FABP) and troponin for early diagnosis of acute myocardial infarction. *Clin Biochem*, 2013. 46(1-2): p. 26-30.
64. Duchenne, J., et al., Diagnostic accuracy of copeptin sensitivity and specificity in patients with suspected non-ST-elevation myocardial infarction with troponin I below the 99th centile at presentation. *BMJ Open*, 2014. 4(3): p. e004449.
65. Johnson, S.R., et al., Methods to elicit beliefs for Bayesian priors: a systematic review. *J Clin Epidemiol*, 2010. 63(4): p. 355-69.
66. Felder, S. and T. Mayrhofer, *Medical Decision Making - A health economics primer*. 2011, Berlin Heidelberg: Springer Science & Business Media.

APPENDIX 1: DECISION TREE

a. Decision tree representing conventional serial high-sensitive troponin (HsTn) measurement and the triple biomarker at t0 without additional HsTn measurements (strategy I).

b. Decision tree representing conventional serial high-sensitive troponin (HsTn) measurement and the triple biomarker with an additional HsTn measurement at t2 (strategy II).

c. Decision tree representing conventional serial high-sensitive troponin (HsTn) measurement and the triple biomarker with additional HsTn measurements at t2 and t6 (strategy III).

APPENDIX 2: SYSTEMATIC LITERATURE REVIEW

Literature review

The literature review was performed in the PubMed database. The following combinations of terms were searched in title or abstract of published papers: (hsTn OR troponin) AND (sensitive OR highly sensitive OR high-sensitive) AND (acute coronary syndrome OR ACS OR acute myocardial infarction OR AMI) AND (sensitivity OR specificity). The search was limited to articles published after 2006 and in English or Dutch. Exclusion criteria were: 1) articles not reporting both sensitivity, specificity, and negative predictive value (NPV) of HsTnl; 2) articles not reporting the performance of HsTnl specifically to diagnose AMI; 3) articles not reporting the performance of the Siemens ADVIA Centaur Ultra I assay; 4) articles focusing on the prognostic value of HsTnl instead of the diagnostic value; and 5) articles which cannot be obtained from the university library or open access databases. The literature search initially resulted in 139 articles. Potentially relevant articles were selected based on title and abstract, and the full text of the remaining articles was assessed to identify relevant articles. The main reason for exclusion involved the use of a HsTn assay other than the Siemens Advia Centaur I Ultra. Finally three relevant articles were included, and one article was added based on the reference lists of one of the other articles (Table 1) [1-4]. All four the studies had comparable inclusion criteria and used the same cut-off value (0.04 ng/mL). Also, the variation in sensitivity, specificity, and diagnostic odds ratio (DOR), was visualized in a plot, to detect any outliers (see Figure 1-3). Because the studies had comparable inclusion criteria, used the same cut-off value, and did not show any outliers, a fixed effects model was assumed, as recommended by Devillé et al. 2002 [5]. A weighted average of the analytical performance was obtained by multiplying the reported performance of HsTnl with the sample size of each of the four studies. This resulted in an estimated average sensitivity and NPV of 86.6% (95% CI: 85.6% - 87.5%), 91.8% (95% CI: 91.0% - 92.5%) and 96.2% (95.7% - 96.8%), respectively at the time of a patient's presentation at the coronary pain unit.

Table 1: result of systematic review. NPV, negative predictive value; HsTnl, high sensitive troponin I [1-4].

Study	Marker	Sample size (n)	Sensitivity (%)	Specificity (%)	NPV (%)
Keller et al, 2009	HsTnl	1,818	90.7%	90.2%	94.6%
Keller et al, 2010	HsTnl	1,386	86.7%	92.4%	95.0%
Kelly et al, 2011	HsTnl	952	76.7%	93.6%	96.3%
Reichlin et al, 2009	HsTnl	718	89%	92%	98%

Figure 1. Pooled sensitivity

Figure 2. Pooled specificity

Figure 3. Pooled diagnostic odds ratio (DOR)

REFERENCES

1. Reichlin, T., et al., Early diagnosis of myocardial infarction with sensitive cardiac troponin assays. *N Engl J Med*, 2009. 361(9): p. 858-67.
2. Keller, T., et al., Copeptin improves early diagnosis of acute myocardial infarction. *J Am Coll Cardiol*, 2010. 55(19): p. 2096-106.
3. Keller, T., et al., Sensitive troponin I assay in early diagnosis of acute myocardial infarction. *N Engl J Med*, 2009. 361(9): p. 868-77.
4. Kelly, A.M., Performance of a sensitive troponin assay in the early diagnosis of acute myocardial infarction in the emergency department. *Emerg Med Australas*, 2011. 23(2): p. 181-5.
5. Deville, W.L., et al., Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol*, 2002. 2: p. 9.

APPENDIX 3: QUESTIONNAIRE AMONG CARDIOLOGISTS

1. What is your current position within the cardiology department of this hospital?

☐ Cardiologist	Continue with question 3.
☐ Doctor in specialist training	Continue with question 2.
☐ Other, namely:	_____

2. How many months experience do you have with the specialism cardiology? (including time in training)
 _____ months

3. Please give your opinion about the following statement:
If a patient presents at the Coronary Pain Unit with symptoms of ACS, and ST-elevation on the ECG, I consider it unnecessary to wait for the laboratory test results prior to initiating treatment.
 - ☐ Fully disagree
 - ☐ Somewhat disagree
 - ☐ Neutral
 - ☐ Somewhat agree
 - ☐ Fully agree

All following questions consider patients without ST-elevation on the ECG.

4. Please give your opinion about the following statement:
If a patient presents at the Coronary Pain Unit with symptoms of ACS, without ST-elevation on the ECG, I consider it unnecessary to wait for the laboratory results prior to initiating treatment.
 - ☐ Fully disagree
 - ☐ Somewhat disagree
 - ☐ Neutral
 - ☐ Somewhat agree
 - ☐ Fully agree

5. Please give your opinion about the following statement:
The current troponin measurement regularly gives slightly elevated test results (0.045-0.099 ng/mL) in patients who do not have acute myocardial infarction.
 - ☐ Fully disagree
 - ☐ Somewhat disagree
 - ☐ Neutral
 - ☐ Somewhat agree
 - ☐ Fully agree

6. In case a patient has a slightly elevated troponin result (0.045-0.099 ng/mL) always reported as having an acute myocardial infarction?
 - ☐ Yes. Continue with question 8.
 - ☐ No. Continue with question 7.

7. In what percentage of all patients who present with symptoms of ACS at the Coronary Pain Unit, with at least one slightly elevated troponin result (0.045 – 0.099 ng/mL), do you estimate that the final diagnosis is that the patient does not have acute myocardial infarction?
 _____%

8. The table below shows troponin results at t = 0, t = 2, and t = 6 hours, in patients with symptoms of ACS, without ST-elevation. Please try to estimate in which percentage of each of those patients you would decide to discharge from hospital, based on the findings described below:

Please note:

- Patients with laboratory results at t = 2 and t = 6 are patients you decided **not** to discharge based on previous laboratory results (at t = 0).
- Slightly elevated troponin result: 0.045 – 0.099 ng/mL

For example: in patients in whom you did not perform a stress ECG, without ST-elevation, and with a negative troponin result at t = 0 and t = 2, you expect to discharge about 16% of those patients. In this case, please tick the box '0-25%':

		Time	Result	0%	0-25%	25-50%	50-75%	75-100%	100%
No ST-elevation	No stress ECG	t = 0 and t = 2	Negative	☐	☒	☐	☐	☐	☐

Symptoms of ACS – no ST elevation

Result troponin measurement			Patients that are discharged following the test results shown below (%). Please tick the appropriate box.					
	Time (hour)	Result	0%	0-25%	25-50%	50-75%	75-100%	100%
No stress ECG performed	t = 0	Slightly elevated	☐	☐	☐	☐	☐	☐
	t = 0	Negative	☐	☐	☐	☐	☐	☐
	t = 0 and t = 2	Negative	☐	☐	☐	☐	☐	☐
	t = 0, t = 2 and t = 6	Negative	☐	☐	☐	☐	☐	☐
Result stress ECG negative	t = 0	Slightly elevated	☐	☐	☐	☐	☐	☐
	t = 0	Negative	☐	☐	☐	☐	☐	☐
	t = 0 and t = 2	Negative	☐	☐	☐	☐	☐	☐
	t = 0, t = 2 and t = 6	Negative	☐	☐	☐	☐	☐	☐

9. Please answer the previous question again, but in this case consider the following information: in the table below the performance of the current troponin test is shown (in grey):
- Of 100 people who have an acute myocardial infarction (AMI), the current troponin measurement detects 85 of those people at t = 0 (see table).
 - However, in case of 100 negative test results, 5 patients do have an AMI.

Please copy the chosen percentages from the previous question (shown in grey), in case of a negative troponin result at t0.

Imagine that two markers would be added to the current troponin test at t0. Try to estimate what the effect of the improved performances achieved by this combination of three markers would be on the percentage of patients that you would estimate to discharge based on the findings described below. The result of the three markers at t = 0 is negative.

For comparison: at t = 3, the current troponin test detects 98 out of 100 patients who have a myocardial infarction, and in case of 100 negative test results, 1 patient does have an AMI.

Symptoms of ACS – no ST elevation

		Performance markers at t = 0		Patients that are discharged (%). Please tick the appropriate box.					
		Marker positive per 100 AMI patients*	Number of false-negative results per 100 patients**	0%	0-25%	25-50%	50-75%	75-100%	100%
No stress ECG performed	Troponin	85	5	☐	☐	☐	☐	☐	☐
	Three markers	90	4	☐	☐	☐	☐	☐	☐
		95	2	☐	☐	☐	☐	☐	☐
		99	1	☐	☐	☐	☐	☐	☐
Result stress ECG negative	Troponin	85	5	☐	☐	☐	☐	☐	☐
	Three markers	90	4	☐	☐	☐	☐	☐	☐
		95	2	☐	☐	☐	☐	☐	☐
		99	1	☐	☐	☐	☐	☐	☐

*If this analysis is performed in 100 patients with AMI, the laboratory marker in this column will detect this number of patients (sensitivity).

** If this analysis is performed in 100 patients, this will result in the numbers as shown in the column with a negative test result, while those patients do have AMI (1 – negative predictive value).

10. Please answer the previous question again, but in this case consider the following information: possibly, the two additional markers could contribute to excluding an acute myocardial infarction in case of a slightly elevated troponin result. The possible performances of the two additional markers are shown in the table below.

Imagine: the results at $t = 0$ are: troponin 0.045 – 0.099 ng/mL, and the two additional markers are negative.

Symptoms of ACS – no ST elevation

Performance <u>two additional</u> markers at $t = 0$			Patients that are discharged (%). (troponin 0.045 – 0.099 ng/mL, additional markers negative). Please tick the appropriate box.					
	Marker positive per 100 AMI patients*	Number of false-negative results per 100 patients**	0%	0-25%	25-50%	50-75%	75-100%	100%
No stress ECG performed	80	8	☐	☐	☐	☐	☐	☐
	90	3	☐	☐	☐	☐	☐	☐
	99	1	☐	☐	☐	☐	☐	☐
Result stress ECG negative	80	8	☐	☐	☐	☐	☐	☐
	90	3	☐	☐	☐	☐	☐	☐
	99	1	☐	☐	☐	☐	☐	☐

*If this analysis is performed in 100 patients with AMI, the laboratory marker in this column will detect this number of patients (sensitivity).

** If this analysis is performed in 100 patients, this will result in the numbers as shown in the column with a negative test result, while those patients do have AMI (1 – negative predictive value).

11. The table below shows troponin results, negative or slightly elevated at t = 0 (0.045 – 0.099 ng/mL), in patients with symptoms of ACS, and without ST-elevation. Based on this information, please try to estimate in what percentage of those patients you estimate to perform the diagnostic and therapeutic interventions as mentioned below, from the moment the patient enters the Coronary Pain Unit, until the patient is discharged from hospital.

Symptoms of ACS – no ST elevation

		Diagnostic and therapeutic interventions performed per patient (%). Please tick the appropriate box.					
		t = 0					
		0%	0-25%	25-50%	50-75%	75-100%	100%
Stress ECG	Negative	☐	☐	☐	☐	☐	☐
	Slightly elevated	☐	☐	☐	☐	☐	☐
Cardiac catheterization	Negative	☐	☐	☐	☐	☐	☐
	Slightly elevated	☐	☐	☐	☐	☐	☐
Administering medication	Negative	☐	☐	☐	☐	☐	☐
	Slightly elevated	☐	☐	☐	☐	☐	☐

12. The table below shows the performance of a troponin test at $t = 0$, and of the combination of the three markers at $t = 0$, in patients with symptoms of ACS, without ST-elevation. Based on this information, please try to estimate in what percentage of those patients you estimate to perform the diagnostic and therapeutic interventions as mentioned below, from the moment the patient enters the Coronary Pain Unit, until the patient is discharged from hospital, in case those markers are negative at $t = 0$.

Please copy the chosen percentages from the previous question (shown in the three grey rows), in case of a negative troponin result at t_0 .

For comparison: at $t = 3$, the current troponin test detects 98 out of 100 patients who have a myocardial infarction, and in case of 100 negative test results, 1 patient does have an AMI.

Symptoms of ACS – no ST elevation

		Performance markers at $t = 0$		Diagnostic and therapeutic interventions performed per patient (%). Please tick the appropriate box.					
		Marker positive per 100 AMI patients*	Number of false-negative results per 100 patients**	0%	0-25%	25-50%	50-75%	75-100%	100%
Stress ECG	Troponin	85	5	☐	☐	☐	☐	☐	☐
	Three markers	90	4	☐	☐	☐	☐	☐	☐
		95	2	☐	☐	☐	☐	☐	☐
		99	1	☐	☐	☐	☐	☐	☐
Cardiac catheterization	Troponin	85	5	☐	☐	☐	☐	☐	☐
	Three markers	90	4	☐	☐	☐	☐	☐	☐
		95	2	☐	☐	☐	☐	☐	☐
		99	1	☐	☐	☐	☐	☐	☐
Administering medication	Troponin	85	5	☐	☐	☐	☐	☐	☐
	Three markers	90	4	☐	☐	☐	☐	☐	☐
		95	2	☐	☐	☐	☐	☐	☐
		99	1	☐	☐	☐	☐	☐	☐

*If this analysis is performed in 100 patients with AMI, the laboratory marker in this column will detect this number of patients (sensitivity).

** If this analysis is performed in 100 patients, this will result in the numbers as shown in the column with a negative test result, while those patients do have AMI (1 – negative predictive value).

13. Please answer the previous question again, but now with the following information:
Imagine: the troponin at $t = 0$ is 0.045 – 0.099 ng/mL. The two additional markers are negative, with below shown the possible diagnostic performances of only those two markers.

Symptoms of ACS – no ST-elevation

	Performance markers at $t = 0$		Diagnostic and therapeutic interventions performed per patient (%) (troponin 0.045 – 0.099 ng/mL, extra markers negative). Please tick the appropriate box.					
	Marker positive per 100 AMI patients*	Number of false-negative results per 100 patients**	0%	0-25%	25-50%	50-75%	75-100%	100%
Stress ECG	80	8	☐	☐	☐	☐	☐	☐
	90	3	☐	☐	☐	☐	☐	☐
	99	1	☐	☐	☐	☐	☐	☐
Cardiac catheterization	80	8	☐	☐	☐	☐	☐	☐
	90	3	☐	☐	☐	☐	☐	☐
	99	1	☐	☐	☐	☐	☐	☐
Administering medication	80	8	☐	☐	☐	☐	☐	☐
	90	3	☐	☐	☐	☐	☐	☐
	99	1	☐	☐	☐	☐	☐	☐

*If this analysis is performed in 100 patients with AMI, the laboratory marker in this column will detect this number of patients (sensitivity).

** If this analysis is performed in 100 patients, this will result in the numbers as shown in the column with a negative test result, while those patients do have AMI (1 – negative predictive value).

14. In case you have any remarks or suggestions, please write them below.

APPENDIX 4: MODEL ASSUMPTIONS

Several assumptions are made in the decision analytical model. First, it is assumed that STEMI is diagnosed based on the ECG result. Those patients will undergo a coronary intervention immediately, and are therefore excluded from this analysis. The percentage of patients with NSTEMI is based on a study by McManus et al. 2011 [1]. Patients without NSTEMI are assumed to be discharged after the laboratory results at either t0, t2, or t6, or after overnight admission.

Only negative and slightly elevated HsTn levels (0.045 – 0.099 ng/mL) are considered in the questionnaire, because levels >0.099 are assumed to occur only in patients that actually have an NSTEMI.

For the triple biomarker, three outcomes are distinguished:

- a negative result on all three assays;
- a negative result for the two additional biomarkers (e.g. H-FABP and copeptin) combined with a slightly elevated HsTn (specifically, HsTn);
- any other outcome than the previous two.

For all patients in this third category, it is assumed that no patients will be discharged at t0. Diagnostic and therapeutic interventions performed in this patient category are assumed to depend on the HsTn result. Furthermore, a slightly elevated HsTn at t0 or t2, is also expected to occur in any subsequent measurements. Discharge rates and activities performed are assumed to be equal to those estimated by cardiologists in patients with slightly elevated HsTn levels at t0. Also, it is assumed that independent of the strategy applied, one ECG is assessed for each patient.

REFERENCES

1. McManus, D.D., et al., Recent trends in the incidence, treatment, and outcomes of patients with STEMI and NSTEMI. *Am J Med*, 2011. 124(1): p. 40-7.

11

Understanding the adoption and use of point-of-care tests in Dutch general practices using multi-criteria decision analysis

Michelle M.A. Kip
J. Marjan Hummel
Elra B. Eppink
Hendrik Koffijberg
Rogier M. Hopstaken
Maarten J. IJzerman
Ron Kusters

Submitted

ABSTRACT

Background

The increasing number of available point-of-care (POC) tests challenges clinicians regarding decisions on which tests to use, how to efficiently use them, and how to interpret the results. Although POC tests may offer benefits in terms of low turn-around-time, improved patient's satisfaction, and health outcomes, only few are actually used in clinical practice. Therefore, this study aims to identify which criteria are, in general, important in the decision to implement a POC test, and to determine their weight. Two POC tests available for use in Dutch general practices (i.e. the C-reactive protein [CRP] test and the glycated haemoglobin [HbA_{1c}] test) serve as case studies. The information obtained from this study can be used to guide POC test development and their introduction in clinical practice.

Methods

Relevant criteria were identified based on a literature review and semi-structured interviews with twelve experts in the field. Subsequently, the criteria were clustered in four groups (i.e. user, organization, clinical value, and socio-political context) and the relative importance of each criterion was determined using Multi-Criteria Decision Analysis (MCDA), with the Analytic Hierarchy Process. A facilitated group session was organized to elicit priorities of ten experts with regard to the two POC tests, compared to central laboratory testing.

Results

Of 20 criteria in four clusters, the test's clinical utility, its technical performance, and risks (associated with the treatment decision based on the test result) were considered most important for using a POC test, with relative weights of 22.2%, 12.6% and 8.5%, respectively. Overall, the experts preferred the POC CRP test over its laboratory equivalent, whereas they did not prefer the POC HbA_{1c} test. This difference was mainly explained by their strong preference for the POC CRP test with regard to the subcriterion 'clinical utility'.

Conclusions

The list of identified criteria, and the insights in their relative impact on the successful implementation of POC tests, may facilitate implementation and use of existing POC tests in clinical practice. In addition, having experts score new POC tests on these criteria, provides developers with specific recommendations on how to increase the probability of successful implementation and use.

BACKGROUND

In the last decades, the use of diagnostic tests has increased rapidly. This number is expected to rise even further given the increasing number of people with multiple (chronic) conditions, and the availability of a wide range of biomarkers for monitoring disease and treatment response [1, 2]. Rapid, accurate diagnostic tests have the potential to improve targeted treatments or referrals and thereby improve the overall quality and efficiency of care delivery. However, the availability of new diagnostic tests also challenges physicians to effectively use and interpret the (combination of) test results [3]. This particularly applies to general practitioners (GPs), as they encounter a wide variety of medical conditions among patients presenting in their general practice, and have to decide on whether or not (and which) diagnostic test(s), if any, to perform in those patients. Although most of the laboratory tests requested by GPs are still performed in central laboratories, advances in technology increasingly allow to perform some of these tests directly during consultation, i.e. at the point-of-care [4], offering large benefits in terms of timely and targeted treatment [5].

More specifically, POC tests have been shown to improve patient's satisfaction, treatment adherence, and health outcomes, while also saving time and costs [4, 6-10]. In addition, previous research indicated that GPs express the desire to have POC tests available to help them diagnose a range of acute and chronic conditions [11]. Despite these clear advantages, only a limited number of POC tests are actually used in clinical practice [11, 12]. Reasons for non-use include concerns about test accuracy, over-reliance on test results, undermining clinical judgment, limited added value on top of clinical judgement, as well as higher average costs of POC testing compared to central laboratory testing, and test reimbursement [12, 13].

The decision to actually implement a POC test instead of continuing to request central laboratory tests is a tradeoff between multiple, often conflicting, objectives [14-16]. The use of structured, explicit approaches to decision-making involving multiple criteria can improve the quality of decisions and identify factors for improvement [16]. A set of techniques known as multiple criteria decision analysis (MCDA) can be used for this purpose [16]. In health technology assessment, MCDA can be used to obtain clarity on the relevance of decision criteria, and on the performance of alternative technologies (e.g. diagnostic tests) on these criteria, thereby increasing the consistency, transparency, and legitimacy of ensuing implementation decisions [16]. One of the most frequently applied techniques for MCDA, which is often applied in group decision making, is the Analytic Hierarchy Process (AHP) [17, 18]. AHP uses pairwise comparisons of criteria which are considered meaningful to decision makers in comparing their preferences for the alternatives [18].

The current study uses AHP to estimate the relative preference for a POC test as compared to the same test when performed in a central laboratory. This will give insights in the criteria which are considered most important in the decision on whether or not to implement and use

a POC test. We selected two cases that impact care delivery yet face implementation difficulty, i.e. the POC C-reactive protein (CRP) test, and the POC glycosylated haemoglobin (HbA_{1c}) test. CRP is an acute-phase protein with levels rising quickly during inflammatory processes. Since 2011, the use of this test is recommended in the practice guideline 'Acute cough' from the Dutch College of General Practitioners [19]. As symptoms of respiratory tract infections are one of the most common reasons to consult the GP, the POC CRP test is of added value in reducing antibiotic prescription rates by helping the GP to differentiate between patients with bronchitis from those with community-acquired pneumonia in routine practice [10, 20-24]. Previous research has shown that the POC CRP test was used by 48% of Dutch GPs in 2013, whereas a more recent study showed that this percentage increased to 80% in 2015 [11, 25]. In addition, previous research has shown that the POC CRP test has similar diagnostic performance as its central laboratory test [26], that it reduces antibiotic prescription rates in primary care [10, 20-23], and that it is cost-effective [27, 28].

The HbA_{1c} test is used among patients diagnosed with diabetes, whose HbA_{1c} levels are usually measured every 3-6 months [29]. Those patients typically have to visit a nurse or phlebotomist for a venipuncture, 1-2 weeks prior to their appointment [29]. A POC HbA_{1c} test offers immediate test results, thereby allowing immediate therapeutic decision making, and consequently, reducing patient visits for phlebotomies [29, 30]. The POC HbA_{1c} test was used by only 6% of Dutch GPs in 2013, which increased to 19% in 2015 [11, 25]. Although the analytical performance of the POC HbA_{1c} tests has increased strongly over the last years, not all available devices met the generally accepted performance criteria [31, 32], and cost-effectiveness analyses were inconclusive [30, 33].

Although the abovementioned examples indicate the potential added value of both POC tests, they differ strongly in their degree of implementation. Therefore, these tests were selected as case studies to get insight into the relative preference of experts regarding the use of POC tests in primary care. This preference is estimated using AHP, by identifying and weighting the criteria that are relevant to the decision to use the POC test. Thereby, this study aims to determine which criteria are, in general, important in the decision to implement and use a POC test in clinical practice and this information can be used to guide further POC test development.

11

METHODS

Multi-criteria decision analysis with the Analytic Hierarchy Process

The implementation of an MCDA study follows several separate steps, and guidance is provided by the taskforce of the International Society For Pharmacoeconomics and Outcomes Research [34]. In the current study we used the AHP method, a facilitated group decision approach that has been described by Hummel et al [18]. This technique reduces complex decisions to multiple pairwise comparisons, and captures both subjective and objective criteria in the

decision-making process [17]. For sets of two decision criteria, respondents are asked to select the most important criterion and to compare their relevance on a 9-point scale. AHP thus relies on the respondents to value the relative importance of the criteria, and their preferences for the options provided, to finally derive the most promising alternative [17, 35].

More specifically, the AHP method consists of several steps. Firstly, the framework for the analysis is determined by setting the goal of the analysis, identifying the alternatives that are compared (which in this study are POC CRP and POC HbA_{1c} as alternatives to their equivalent central laboratory tests), and defining the relevant criteria. Secondly, importance weights are obtained for each of the criteria. Thirdly, the performance of both POC tests on each decision criterion was valued relatively to the performance of their laboratory equivalents. The final step is the aggregation of the scores and weights to determine the relative value for the POC CRP and the POC HbA_{1c} test, as alternative to their equivalent central laboratory test. We assume that the technique with the highest value is the preferred one, and is therefore most likely to be implemented (and used) in clinical practice.

Identification of the alternatives

For this AHP, it is assumed that the use of a POC CRP test or a POC HbA_{1c} test will replace the use of the equivalent test performed in a central laboratory, instead of being used as an additional test in the diagnostic track. This means that the criteria regarding the (potential) use of a POC test within the GP's office will be compared to the criteria of the equivalent test when performed in a central laboratory.

Because of the differences in the degree of diffusion of the POC CRP and the POC HbA_{1c} test (as mentioned in the introduction), these tests were chosen as case studies in the current analysis. In addition, as both cases (i.e. CRP and HbA_{1c}) have very different areas of application, this increases the probability that the results obtained from this AHP are also relevant to the wide range of POC tests that can be used in Dutch general practices.

Identification and structuring of the criteria

Relevant criteria for the AHP analysis were identified from a literature search, of which the details are described in Appendix 1. The literature search resulted in the inclusion of seven journal articles, one guideline, and three reports [9, 11, 12, 36-43]. Each paper was independently assessed by two reviewers (EE and MK) to identify all relevant criteria. The criteria obtained were compared to a previously published study in implementation science [44]. However, the criteria identified from the literature search were very different from those reported in this previous study, which can be explained by the fact that this previous study specifically focused on guidelines and intervention programs in healthcare [44]. Still, the four categories of criteria which were identified from this previous study, could be used to group the criteria obtained from literature. Those four categories included criteria related to: 1) the user, 2) the

organization, 3) the clinical value, and 4) the socio-political context [44]. Both the identification of the criteria, as well as categorizing those criteria into the four main groups of criteria were performed by the same two reviewers. Differences were discussed until consensus was reached on the selection and categorization of the criteria.

Expert panel

In order to judge both the validity and completeness of this list of criteria, the (potential) presence of redundant or overlapping criteria, as well as the preferential independence of the criteria [34], the initial list and suggested structure were validated by means of individual interviews with 12 experts. In addition, these experts also judged the clarity, completeness, and unambiguity of the definitions accompanying each of the criteria. The panel of experts was selected in such a way to represent all stakeholders who are either involved in the process of POC test development, or who affect the eventual decision on whether or not to implement and use the POC CRP and the POC HbA_{1c} test in clinical practice. Therefore, the experts participating were three GPs (who differed in the use of the POC HbA_{1c} test within their general practice, as shown in Table 1), a GP’s assistant, a diabetes patient (a biology teacher who is familiar with both the HbA_{1c} and the CRP test), two clinical chemists (specialized in the use of POC tests in primary care), a technology developer (director of a company developing a POC test), a policy maker (with a focus on primary care), a POC specialist (who formerly worked as a GP), an insurer in the field of primary health care, and a former director of a health insurance company who also is a professor in the field of healthcare and medical technology. All participating experts were informed on the goal of the study as well as on the duration of the interview and the AHP session. As the current study focuses on the Dutch setting, all selected participants were living and employed in the Netherlands. The participants were initially approached via email. Participation was voluntary and the experts had the opportunity to quit their participation at any time. This study was approved by the Ethics Committee of the faculty of Behavioral, Management and Social Sciences of the University of Twente, Enschede, the Netherlands. A full overview of the participating experts and their professional backgrounds is presented in Table 1.

Figure 1. Hierarchical structure of the AHP. This figure shows an overview of the hierarchical structured used for the analytical hierarchy process (AHP). NPV = negative predictive value, PPV = positive predictive value, POC = point-of-care, QC = quality control

Validation of the list of criteria obtained from literature was performed by means of individual, semi-structured interviews with each of the 12 experts (by EE), either via phone or at the participant's workplace. After each session, the suggested adaptations to the identification and structuring of the criteria, were discussed by MK and EE until consensus was reached. The updated list was then used as input for the next semi-structured interview. This repetitive process resulted in a final list of 20 criteria (categorized in four groups of five subcriteria), which were expected to affect the implementation and use of POC tests in general practices. These criteria were used to compose the hierarchical evaluation structure (Figure 1). A full description of the 20 subcriteria, and the accompanying range applied in the AHP session, is included in Appendix 2.

Table 1. Composition of expert team.

No.	Profession	Core relation to POC testing in general practices
1	GP	User of CRP and HbA _{1c} test
2	GP	User of CRP test, non-user HbA _{1c} test
3	GP	User of CRP test, non-user HbA _{1c} test
4	GP's assistant	User of CRP test, non-user HbA _{1c} test, former nurse
5	Diabetic patient	User of HbA _{1c} test (as a patient), biology teacher (familiar with CRP)
6	Clinical chemist	Laboratory professional, specialized in POC tests in primary care
7‡	Clinical chemist	Laboratory professional, specialized in POC tests in primary care
8	Technology developer	Director lab-on-a-chip company
9	Policy maker	Concerned with the quality of care provided in primary care
10*	POC specialist	Expert in POC testing, GP
11*	Payer	Insurer in primary health care, former GP
12‡	Payer	Former director of health insurance company, and professor in healthcare

**did not participate in the group AHP session but completed the AHP session afterwards. ‡did not participate in the AHP session. CRP = C-reactive protein, GP = general practitioner, HbA_{1c} = glycated haemoglobin, POC = point-of-care.*

All 12 experts who participated in the initial interviews (Table 1), were also invited to participate in the AHP session (at the University of Twente). A group setting was chosen for this AHP analysis because it allows panel members to share information about their attitudes, beliefs, as well as knowledge, which underlie the priorities they assign to the outcome measures [45]. One clinical chemist (expert no. 7) and the former director of a health insurance company (expert no. 12) were however unable to participate in the AHP session. All remaining ten experts agreed to participate in the AHP session, although two of those experts (expert no. 10

and expert no. 11) were unable to attend the actual group session, and therefore completed the AHP session individually afterwards.

AHP session

The AHP session was performed in accordance with previously published literature and/or guidance documents [18, 46]. A three-hour AHP session was organized (by EE, and with attendance of JH and MK) during which the expert team discussed the relative importance of the four main criteria as well as each of the five (sub)criteria within each main criterion. This session was performed using Team Expert Choice software (Expert Choice, Arlington, VA), a group decision support system which incorporates the mathematical procedures of AHP, and all conversations during this AHP session were audio recorded.

During this session, all members of the expert team were asked to provide their judgments on pairwise comparisons of the importance of the four main criteria, as well as the 20 subcriteria, and for their preference for the selected alternatives (i.e. the POC test or the central laboratory test for both CRP and HbA_{1c}). The performance of both POC tests and the central laboratory tests on the selected criteria, was left to the experts' judgement. Responses were collected individually using hand-held remote controlled keypads. All individual judgments were projected on a screen, allowing the expert team members to provide their motivation for their individual scores, and to share their expertise on this topic. To allow the other team members to incorporate the expertise and motivations that were shared, they had the opportunity to alter their judgements during those dialogs. For each pairwise comparison, the final individual judgments were aggregated, based on the geometric mean. This resulted in group weighing factors which represented the relative importance of each of the criteria, and these weighing factors were used to calculate the relative preference for the alternatives (i.e., the POC test or the central laboratory test). In addition, the inconsistency of the expert's judgments was calculated after the pair-wise comparisons of each of the main criteria, and after each of the 20 subcriteria. In accordance with literature, an inconsistency below 0.1 was considered acceptable [17]. In case of a too high inconsistency, the experts were asked to reconsider the pair-wise comparison which caused this inconsistency (to make sure that the criteria compared were well-understood), or they were asked to fill in an additional comparison.

RESULTS

The results of the pairwise comparisons of the four main criteria and the 20 subcriteria, as well as the preferences regarding the POC CRP and POC HbA_{1c} test (as compared to their laboratory equivalents) are shown in Table 2.

Table 2. Result of pairwise comparisons.

Determinant	Weights*		Performance scores			
	Weight of criterion	Overall weight	POC CRP	CRP central lab	POC HbA _{1c}	HbA _{1c} central lab
Determinants in relation to the user	13.6%	NA	81.9%	18.1%	68.2%	31.8%
1. Satisfaction patient	25.4%	3.5%	89.9%	10.1%	85.3%	14.7%
2. Clarity of procedure	24.8%	3.4%	71.3%	28.7%	56.1%	43.9%
3. User-friendliness	21.0%	2.9%	84.2%	15.8%	78.1%	21.9%
4. Test interpretation	6.4%	0.9%	69.0%	31.0%	50.0%	50.0%
5. Turn-around-time	22.3%	3.0%	90.0%	10.0%	89.8%	10.2%
Determinants in relation to the organisation	20.9%	NA	65.4%	34.6%	48.8%	51.2%
6. Frequency of use	14.6%	3.1%	88.6%	11.4%	49.5%	50.5%
7. Room for innovation	33.5%	7.0%	89.7%	10.3%	85.0%	15.0%
8. Workload	22.1%	4.6%	77.1%	22.9%	29.6%	70.4%
9. Support, training and quality control	25.4%	5.3%	21.3%	78.7%	24.4%	75.6%
10. Connectivity	4.4%	0.9%	35.7%	64.3%	46.5%	53.5%
Determinants in relation to the clinical value	51.8%	NA	55.1%	44.9%	45.8%	54.2%
11. Clinical utility	42.8%	22.2%	89.4%	10.6%	53.5%	46.5%
12. Technical performance	24.4%	12.6%	28.1%	71.9%	33.5%	66.5%
13. Negative Predictive Value	13.7%	7.1%	48.3%	51.7%	50.0%	50.0%
14. Positive Predictive Value	2.8%	1.5%	48.3%	51.7%	50.0%	50.0%
15. Risks	16.4%	8.5%	34.4%	65.6%	33.5%	66.5%
Determinants in relation to the socio-political context	13.8%	NA	72.9%	27.1%	49.1%	50.9%
16. Clinical guidelines	34.1%	4.7%	82.1%	17.9%	47.3%	52.7%
17. Scientific evidence	23.7%	3.3%	85.0%	15.0%	60.5%	39.5%
18. Reimbursement	28.5%	3.9%	60.7%	39.3%	42.2%	57.8%
19. Overall costs	8.3%	1.1%	88.8%	11.2%	55.5%	44.5%
20. Legislations	5.5%	0.8%	42.3%	57.7%	40.5%	59.5%
Overall preference for POC or central laboratory test			62.9%	37.1%	49.4%	50.6%

This table shows the results of the pairwise comparisons of the four main criteria and the 20 subcriteria, as well as the preferences regarding the POC CRP and POC HbA_{1c} test (as compared to their central laboratory equivalents). The overall weight is obtained by multiplying the weight of the main criterion which each of the subcriteria. The definition of each of the criteria is provided in Appendix 2. NA = not applicable, POC = point-of-care. * The sum of columns may not add up to 100.0% due to rounding.

Table 2 indicates that a change in the main criterion 'clinical value', as caused by the use of a POC test, is estimated to be most important for comparing preferences for a POC test with its central laboratory test. The other main criteria in order of decreasing importance are 'organisation', 'socio-political context' and 'user'. An overall group inconsistency within those four main criteria of 0.05 was found, which is acceptable considering the 0.1 threshold (Appendix 2). Among the 20 subcriteria, the results indicate that the subcriterion 'clinical utility' (i.e. the extent to which a correct (treatment) decision, as based on the POC test result, has added value in terms of clinical outcomes), with a relative impact of 22.2% is expected to have the highest impact on the relative preference for the POC test as compared to the central laboratory test. The second most important subcriterion was found to be 'technical performance', with a relative weight of 12.6%. The subcriterion 'risks' is the third most important with an expected relative weight of 8.5%. The high weights of those three subcriteria can also be explained by the high impact of the main criterion 'clinical value' (with a weight of 51.8%), to which those three subcriteria belong. Also, results indicate that the subcriteria 'legislations', 'connectivity', and 'test interpretation', have little impact on the preference for the use of a POC test as compared to a central laboratory test, with relative weights of 0.8%, 0.9%, and 0.9% respectively.

Figure 2. Result of AHP analysis on the POC CRP test as compared to central laboratory testing. This figure shows the result of the analytical hierarchy process (AHP) analysis on the POC CRP test as compared to using the central laboratory test, on the four main criteria and the overall result, as well as the performance of the two tests on each criterion. The grey bars represent the relative weights of the four main criteria. The figure represents the performance of the POC test (square) and lab-test (triangle) on each criterion. CRP = C-reactive protein, POC = point-of-care.

Figure 3. Result of AHP analysis on the POC HbA_{1c} test as compared to central laboratory testing. This figure shows the result of the analytical hierarchy process (AHP) analysis on the POC HbA_{1c} test as compared to using the central laboratory test, the four main criteria and the overall result, as well as the performance of the two tests on each criterion. The grey bars represent the relative weights of the four main criteria. The grey bars represent the relative weights of the four main criteria. The figure represents the performance of the POC test (square) and lab-test (triangle) on each criterion. HbA_{1c} = glycated hemoglobin, POC = point-of-care.

The results of the comparison of the preferences of using the POC CRP test and the POC HbA_{1c} test as compared to using the equivalent central laboratory tests, are shown in Figures 2 and 3, respectively.

Results indicate that the experts strongly prefer the POC CRP test as compared to its central laboratory test (i.e. 62.9% vs. 37.1%). In addition, the POC test scores higher in all four main criteria. This strong preference for performing the CRP test on a POC analyzer is mainly caused by its expected improvement in 'clinical utility', which was previously identified as be the subcriterion with the highest relative weight. Besides 'clinical utility', the POC CRP test is mainly preferred with regard to the expected shorter 'turn-around time', an expected increase in 'patient satisfaction' and 'user-friendliness', the 'room for innovation' that is experienced and its high 'frequency of use', its 'uptake in current clinical guidelines' and the 'scientific evidence' available, as well as its impact on 'overall costs' (Table 2 and Appendix 2).

However, the overall preference for the POC CRP test (over its central laboratory test) is dependent on the weights assigned to each of the subcriteria. Results indicate that the subcriteria 'clinical utility' and 'room for innovation' are most influential in favoring the POC

test, whereas the subcriterion 'technical performance' was most influential in favoring the laboratory test.

In contrast, for the HbA_{1c} test, the experts displayed (almost) equal preference for the POC and the central laboratory test (i.e. 49.4% vs. 50.6%). More specifically, with regard to the groups of main criteria, the POC test is only preferred over its central laboratory test for the main criterion 'user', caused by the preference assigned to the POC test with regard to the subcriteria 'satisfaction patient', 'user-friendliness', and 'turn-around-time' (Table 2 and Appendix 2). When taking into account the weight of the subcriteria, the subcriterion 'room for innovation' was found to be most influential in favor of the POC HbA_{1c} test, whereas (as for the CRP test) the subcriterion 'technical performance' was most influential in favor of the central laboratory test. However, in terms of 'clinical value', which is the main criterion with the highest relative weight, the central laboratory test is slightly preferred over the POC HbA_{1c} test. The preferences of the subcriteria within this main criterion indicate that this is mainly caused by the expected decrease in 'technical performance' and expected increase in 'risks' related to making management decision based on the POC test result (Table 2 and Appendix 2). Although the POC HbA_{1c} test was slightly preferred over the central laboratory test for the subcriterion 'clinical utility' (i.e. 53.5% vs. 46.5%), which is the subcriterion with the highest relative weight, this slightly higher score could not offset the lower scores on the other subcriteria within this main criterion. In addition, the POC test is not preferred over the central laboratory test for the main criteria 'organisation' (48.8% vs. 51.2%) and 'socio-political context' (49.1% vs. 50.6%). Regarding the main criterion 'organisation', results indicate that (for HbA_{1c}) the central laboratory test is mainly preferred over the POC test with respect to the subcriteria 'workload', and 'support, training and quality control'.

When considering the differences in preferences assigned to the 20 subcriteria for the POC and the central laboratory test (for both CRP and HbA_{1c}), the main differences were found in the subcriteria 'clinical utility', the user's 'workload', its expected 'frequency of use', the extent to which the test is incorporated in current 'clinical guidelines', as well as its impact on 'overall costs'.

DISCUSSION

The results of this study indicate that the main criterion 'clinical value' was found to be most important for comparing preferences for a POC test to its central laboratory equivalent (i.e. a relative weight of 51.8%). This main criterion includes the subcriteria 'clinical utility', 'technical performance', 'negative predictive value' (NPV), 'positive predictive value' (PPV), and 'risks'. Its subcriterion 'clinical utility' was assigned the highest relative weight (i.e. 42.8%), indicating that the expert panel considered "the extent to which a correct (treatment) decision, based on the test result, has added value in terms of clinical outcomes" as highly important. Overall, the current study found a relative preference of implementing and/or using the POC CRP

test instead of its central laboratory test in primary care, whereas the POC HbA_{1c} test was not preferred over its central laboratory equivalent. These results are in line with the (previously reported) broad diffusion of the POC CRP test of 80% in the Netherlands, and with the limited diffusion of the HbA_{1c} test of 19% [25].

Interestingly, results indicate that the expert panel considered the test's NPV to be far more important than its PPV (i.e. overall weight 7.1% vs. 1.5%), which is most likely attributable to the strong gatekeeping function of GPs. This gatekeeping function implies that GPs aim to prevent unnecessary referrals to secondary care, by ruling out serious conditions in the heterogeneous patient population that visit GPs' offices, thereby preventing (unnecessary) further testing and treatment with the associated patient burden and healthcare expenditures. Because of this focus on ruling out diseases or conditions, diagnostic tests for use in primary care are primarily aimed at having a high NPV, instead of a high PPV (as is required for ruling in). This was also suggested in the guideline from Hopstaken et al. (2015) [43].

Results indicate a strong preference for the POC CRP and POC HbA_{1c} test as compared to their laboratory equivalents with regard to the subcriterion 'turn-around-time' (i.e. a weight of 90.0% and 89.8%, respectively). The experts however assigned a strong preference for the POC CRP test with regard to the subcriterion 'clinical utility' (89.4%), as opposed to only a slight preference for POC HbA_{1c} (53.5%). This difference may be explained by considering the tests' intended use, because the POC CRP test is intended to immediately affect the decision to prescribe antibiotics, whereas the POC HbA_{1c} test is used to monitor a chronic condition. This finding is in line with previous research: Cals and Van Weert found that GPs especially preferred a POC test in case it can be used to rapidly exclude a serious condition [39], and Howick et al found that GPs desired POC tests to help them diagnose acute conditions [11]. In addition, another benefit of the POC CRP test is that its use is not limited to patients with acute cough, but it may also be used to diagnose or exclude other inflammatory processes [47]. In contrast, the use of the POC HbA_{1c} test is limited to the monitoring (and increasingly to the diagnosis) of diabetes [48, 49].

Strengths and limitations

The main strength of the current study is that the set of criteria that were selected from literature were validated in interviews with a multidisciplinary team of 12 experts. This has most likely contributed to preventing the AHP structure from becoming too complex by categorizing all aspects into a list of 20 subcriteria, while simultaneously ensuring that all relevant aspects remain incorporated.

As the GP is (inevitably) very important in the eventual decision on whether or not to implement and use a POC test, a strength of this study is that three GPs participated both in the expert panel and in the actual AHP session. These GPs were chosen in such a way that they

had a different number of years working experience and/or a different viewpoint regarding the use of POC tests in their general practice, which likely resulted in a more representative weighting of the criteria. On the other hand, it is arguable that this overrepresentation of GPs may have biased the results in favor of their preferences. More specifically, confirmation bias may have been present [50], as their current POC test use (i.e. one GP used POC HbA_{1c}, whereas all three GPs used POC CRP) may have strengthened their preference towards the POC CRP test. However, a separate analysis revealed that when only the GPs' opinions would have been incorporated in the AHP analysis, this would not have changed the overall conclusion (data not shown).

As the health insurer and the POC expert were not able to attend the actual AHP session, they completed the session individually afterwards. These two experts were provided with the scores given by the experts in the actual AHP session, as well as with the reasoning behind these scores. Consequently, only they had the opportunity to adjust their answers accordingly. It can therefore not be ruled-out that some changes in the scores given by the experts during the AHP group session would have occurred if those two experts would have attended the group session. However, subgroup analyses revealed that the overall conclusion was unaffected when excluding these two experts from the analyses (data not shown).

In the current study, the 20 subcriteria were clustered into four main criteria, each containing five subcriteria. Preference weights have been assigned to each of these subcriteria as well as to the four main criteria. The risk of such a clustered evaluation is that the weight of a subcriterion depends on the weight assigned to its main criterion. To illustrate this, the subcriterion 'clinical guidelines' was assigned a weight of 34.1%, but as the main criterion it belongs to (i.e. 'socio-political context') was only assigned a weight of 13.8%, the overall weight of 'clinical guidelines' is only 4.7%. Thereby, this clustered evaluation may result in a (slight) underestimation of the overall weight of some important criteria. The main benefit of a clustered evaluation, however, is that, by assigning weights to both subcriteria as well as main criteria, bias of overweighting main criteria is avoided, which would occur when weights of partially overlapping subcriteria were summed to calculate the weights of the main criteria.

Recommendations

It is not yet certain whether the results of the current study can be transferred to other POC tests and to other settings. Although the strong differences in the degree of implementation as well as in the application of the CRP and the HbA_{1c} test likely enhance the transferability of the study's results, it is not certain whether using one set of criterion weights (as in the current study) can reliably be applied to such a wide variety of tests. Therefore, future AHP sessions are recommended to find out whether criteria weights specific to the type of test are required.

Although the results of the current study indicate a preference for the POC CRP test (over its

laboratory equivalent), this cannot guarantee its (further) implementation and use in clinical practice. To make such a decision, additional (temporal) costs (and efforts) related to switching to a POC test have to be considered. These costs for example include costs related to the purchase and maintenance of the POC testing device, to obtaining the patient's (blood) sample, as well as to educating the test's users. Nevertheless, the insights provided into other factors that play a role in the decision to implement and use POC tests in Dutch general practices, may enhance the efficiency of both their development and implementation.

CONCLUSION

In conclusion, the list of criteria that was identified in the current study may facilitate efficient implementation and use of existing POC tests. In addition, it is likely valuable to predict, in an early stage of POC test development, their likelihood of implementation and use in general practices. The insights obtained in the barriers and facilitating factors of POC tests can be used for optimizing their development, as it allows test developers to focus on the criteria that were found to be most important, thereby increasing the probability of successful test implementation and use.

DECLARATIONS

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the faculty of Behavioural, Management and Social Sciences of the University of Twente, Enschede, the Netherlands. All participants provided verbal consent to participate.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was not funded.

Authors' contributions

MK, MH, EE, RK and MIJ were involved in the conception and design of the study. EE and MK performed the literature review. EE performed the expert interviews, and EE, MK and MH coordinated the AHP session. MK, MH, and EE analyzed the data, and all authors were involved in the interpretation of the results. MK drafted the manuscript, and MH, EE, HK, RH, MIJ and RK were major contributors in critically reviewing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank all experts who participated in the interviews and/or in the AHP session, as this study could not have been performed without their valuable input and expertise.

REFERENCES

- Wiener, J.M. and J. Tilly, Population ageing in the United States of America: implications for public programmes. *Int J Epidemiol*, 2002. 31(4): p. 776-81.
- Akhmetov, I. and R.V. Bubnov, Assessing value of innovative molecular diagnostic tests in the concept of predictive, preventive, and personalized medicine. *EPMA J*, 2015. 6: p. 19.
- Hickner, J., et al., Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. *J Am Board Fam Med*, 2014. 27(2): p. 268-74.
- Larsson, A., R. Greig-Pylypczuk, and A. Huisman, The state of point-of-care testing: a European perspective. *Ups J Med Sci*, 2015. 120(1): p. 1-10.
- Price, C.P., Point of care testing. *BMJ*, 2001. 322(7297): p. 1285-8.
- Laurence, C.O., et al., Patient satisfaction with point-of-care testing in general practice. *Br J Gen Pract*, 2010. 60(572): p. e98-104.
- Gialamas, A., et al., Does point-of-care testing lead to the same or better adherence to medication? A randomised controlled trial: the PoCT in General Practice Trial. *Med J Aust*, 2009. 191(9): p. 487-91.
- Bubner, T.K., et al., Effectiveness of point-of-care testing for therapeutic control of chronic conditions: results from the PoCT in General Practice Trial. *Med J Aust*, 2009. 190(11): p. 624-6.
- National Institute of Biomedical, I., et al., Improving healthcare accessibility through point-of-care technologies. *Clin Chem*, 2007. 53(9): p. 1665-75.
- Cals, J.W., et al., Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. *BMJ*, 2009. 338: p. b1374.
- Howick, J., et al., Current and future use of point-of-care tests in primary care: an international survey in Australia, Belgium, The Netherlands, the UK and the USA. *BMJ Open*, 2014. 4(8): p. e005611.
- Jones, C.H., et al., Primary care clinicians' attitudes towards point-of-care blood testing: a systematic review of qualitative studies. *BMC Fam Pract*, 2013. 14: p. 117.
- Lewandowski, K., K. Gregory, and D. Macmillan, Assuring quality in point-of-care testing: evolution of technologies, informatics, and program management. *Arch Pathol Lab Med*, 2011. 135(11): p. 1405-14.
- Greenhalgh, T., et al., Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q*, 2004. 82(4): p. 581-629.
- Grol, R., et al., Improving patient care: the implementation of change in health care, second edition. 2013: John Wiley & Sons Ltd. 392.
- Thokala, P., et al., Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*, 2016. 19(1): p. 1-13.
- Saaty, T.L., *The Analytic Hierarchy Process: planning, priority setting, resource allocation*. 1980, New York: McGraw-Hill.
- Hummel, J.M., J.F. Bridges, and I.J. MJ, Group decision making with the analytic hierarchy process in benefit-risk assessment: a tutorial. *Patient*, 2014. 7(2): p. 129-40.
- Verlee, L., et al., [Summary of NHG practice guideline 'Acute cough']. *Ned Tijdschr Geneesk*, 2012. 156(0): p. A4188.
- Minnaard, M.C., et al., C-reactive protein point-of-care testing and associated antibiotic prescribing. *Fam Pract*, 2016. 33(4): p. 408-13.
- Cals, J.W., et al., Point-of-care C-reactive protein testing and antibiotic prescribing for respiratory tract infections: a randomized controlled trial. *Ann Fam Med*, 2010. 8(2): p. 124-33.
- Huang, Y., et al., Association between point-of-care CRP testing and antibiotic prescribing in respiratory tract infections: a systematic review and meta-analysis of primary care studies. *Br J Gen Pract*, 2013. 63(616): p. e787-94.
- Aabenhus, R., et al., Biomarkers as point-of-care tests to guide prescription of antibiotics in patients with acute respiratory infections in primary care. *Cochrane Database Syst Rev*, 2014(11): p. CD010130.
- Tonkin-Crine, S.K., et al., Clinician-targeted interventions to influence antibiotic prescribing behaviour for acute respiratory infections in primary care: an overview of systematic reviews. *Cochrane Database Syst Rev*, 2017. 9: p. CD012252.
- Kip, M.M.A., et al., Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners. *Prim Health Care Res Dev*, 2017: p. 1-12.
- Minnaard, M.C., et al., The added diagnostic value of five different C-reactive protein point-of-care test devices in detecting pneumonia in primary care: A nested case-control study. *Scand J Clin Lab Invest*, 2015. 75(4): p. 291-5.
- Cals, J.W., et al., C-reactive protein point of care testing and physician communication skills training for lower respiratory tract infections in general practice: economic evaluation of a cluster randomized trial. *J Eval Clin Pract*, 2011. 17(6): p. 1059-69.
- Oppong, R., et al., Cost-effectiveness of point-of-care C-reactive protein testing to inform antibiotic prescribing decisions. *Br J Gen Pract*, 2013. 63(612): p. e465-71.
- Pluddemann, A., et al., Primary care diagnostic technology update: point-of-care testing for glycosylated haemoglobin. *Br J Gen Pract*, 2011. 61(583): p. 139-40.

30. Grieve, R., et al., Near patient testing in diabetes clinics: appraising the costs and outcomes. *Health Technol Assess*, 1999. 3(15): p. 1-74.
31. Lenters-Westra, E. and R.J. Slingerland, Six of eight hemoglobin A1c point-of-care instruments do not meet the general accepted analytical performance criteria. *Clin Chem*, 2010. 56(1): p. 44-52.
32. Lenters-Westra, E. and R.J. Slingerland, Three of 7 hemoglobin A1c point-of-care instruments do not meet generally accepted analytical performance criteria. *Clin Chem*, 2014. 60(8): p. 1062-72.
33. Khunti, K., et al., Randomised controlled trial of near-patient testing for glycated haemoglobin in people with type 2 diabetes mellitus. *Br J Gen Pract*, 2006. 56(528): p. 511-7.
34. Marsh, K., et al., Multiple Criteria Decision Analysis for Health Care Decision Making--Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*, 2016. 19(2): p. 125-37.
35. Dolan, J.G., Shared decision-making--transferring research into practice: the Analytic Hierarchy Process (AHP). *Patient Educ Couns*, 2008. 73(3): p. 418-25.
36. Kleinveld, H.A., et al., Visiedocument NVKC Point-of-care testen in de eerste lijn. 2012: Heerlen. p. 8.
37. Briggs, C., S. Kimber, and L. Green, Where are we at with point-of-care testing in haematology? *Br J Haematol*, 2012. 158(6): p. 679-90.
38. Hofland, H.J., Point of care testing and selftest related consultations in general practices in the Netherlands: an exploratory study on general practitioners' experiences. 2010, University of Twente: Enschede. p. 30.
39. Cals, J. and H. van Weert, Point-of-care tests in general practice: hope or hype? *Eur J Gen Pract*, 2013. 19(4): p. 251-6.
40. Jackman, J., et al., Minding the Gap: An approach to determine critical drivers in the development of Point of Care diagnostics. *Point Care*, 2012. 11(2): p. 130-139.
41. de Vries, C., et al., Results of a survey among GP practices on how they manage patient safety aspects related to point-of-care testing in every day practice. *BMC Fam Pract*, 2015. 16: p. 9.
42. Heneghan, C., et al., Diagnostics Forum 2013 report: fast-tracking the evidence for implementing diagnostic tests. 2013, University of Oxford: Oxford. p. 16.
43. Hopstaken, R.M., et al., Richtlijn: point of care testing (POCT) in de huisartsenzorg. 2015.
44. TNO innovation for life, MIDI: MeetInstrument Determinanten van Innovaties. 2012: Leiden. p. 4.
45. Hummel, J.M., et al., The effects of team expert choice on group-decision making in collaborative new product development: a pilot study. *J Multi-Criteria Decis Anal*, 2000. 9(1-3): p. 90-8.
46. Hilgerink, M.P., et al., Assessment of the added value of the Twente Photoacoustic Mammoscope in breast cancer diagnosis. *Med Devices (Auckl)*, 2011. 4: p. 107-15.
47. Verbakel, J.Y., et al., Should all acutely ill children in primary care be tested with point-of-care CRP: a cluster randomised trial. *BMC Med*, 2016. 14(1): p. 131.
48. Jia, W., Standardising HbA1c-based diabetes diagnosis: opportunities and challenges. *Expert Rev Mol Diagn*, 2016. 16(3): p. 343-55.
49. World Health Organization, Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. Abbreviated report of a WHO consultation. 2011: Geneva. p. 25.
50. Glick, M., Believing is seeing: Confirmation bias. *J Am Dent Assoc*, 2017. 148(3): p. 131-132.

APPENDIX 1: LITERATURE SEARCH

A literature search was performed in Scopus and PubMed, using combinations of the following search terms: point-of-care testing OR point of care testing OR point-of-care test OR point of care test OR POC OR POC test OR POCT, general practitioner OR GP OR general practice, innovation, healthcare OR health care, implementation, measurement instrument OR factors OR determinants, and innovation. Potentially relevant articles were selected based on title and abstract, and the full text of the remaining articles was assessed to identify relevant articles. Only articles published in English or Dutch, and available from the university library or open access databases, were included. In addition, a search on the internet, using similar search terms, was conducted to obtain general information (like guidelines and reports), which could not be obtained via PubMed or Scopus. The literature search resulted in the inclusion of seven journal articles, one guideline, and three reports [1-11]. The literature search was performed in September 2015.

REFERENCES

1. Kleinveld, H.A., et al., Visiedocument NVKC Point-of-care testen in de eerste lijn. 2012: Heerlen. p. 8.
2. Briggs, C., S. Kimber, and L. Green, Where are we at with point-of-care testing in haematology? *Br J Haematol*, 2012. 158(6): p. 679-90.
3. Hofland, H.J., Point of care testing and selftest related consultations in general practices in the Netherlands: an exploratory study on general practitioners' experiences. 2010, University of Twente: Enschede. p. 30.
4. Jones, C.H., et al., Primary care clinicians' attitudes towards point-of-care blood testing: a systematic review of qualitative studies. *BMC Fam Pract*, 2013. 14: p. 117.
5. Howick, J., et al., Current and future use of point-of-care tests in primary care: an international survey in Australia, Belgium, The Netherlands, the UK and the USA. *BMJ Open*, 2014. 4(8): p. e005611.
6. National Institute of Biomedical, I., et al., Improving healthcare accessibility through point-of-care technologies. *Clin Chem*, 2007. 53(9): p. 1665-75.
7. Cals, J. and H. van Weert, Point-of-care tests in general practice: hope or hype? *Eur J Gen Pract*, 2013. 19(4): p. 251-6.
8. Jackman, J., et al., Minding the Gap: An approach to determine critical drivers in the development of Point of Care diagnostics. *Point Care*, 2012. 11(2): p. 130-139.
9. de Vries, C., et al., Results of a survey among GP practices on how they manage patient safety aspects related to point-of-care testing in every day practice. *BMC Fam Pract*, 2015. 16: p. 9.
10. Heneghan, C., et al., Diagnostics Forum 2013 report: fast-tracking the evidence for implementing diagnostic tests. 2013, University of Oxford: Oxford. p. 16.
11. Hopstaken, R.M., et al., Richtlijn: point of care testing (POCT) in de huisartsenzorg. 2015.

APPENDIX 2

Questions used for semi-structured interviews, and design of analytical hierarchy process session

This supplementary file contains an overview of the questions asked in the semi-structured interviews, as well as of the design of the analytical hierarchy process session. The interviews were conducted in Dutch. The interview questions (as shown below) were translated to English for the purpose of this publication.

Semi-structured interviews (approximate duration: 1 hour)

In a semi-structured interview (face-to-face or via telephone), the questions as shown below were asked to the participants. In agreement with the participants, the interviews were audio-recorded. The participants were informed that the results will be owned by the University of Twente, will only be used for scientific purposes, and will only be published or reported anonymously.

Questions with regard to the participant:

1. What is your gender?
2. What is your age?
3. What position do you currently hold?
4. For how long have you already been working in this position?
5. In what type of organization are you employed, and where is it located?
6. For general practitioners only: what is the distance to the nearest central laboratory?
7. For general practitioners only: which point-of-care tests do you have access to, and do you use these tests frequently?
8. For other professionals: with which point-of-care tests are you familiar?

Agree on the definition of point-of-care testing:

In the current Dutch guidelines, point-of-care testing within general practices is defined as [1]: “point of care testing (POCT) is the process of initiating, conducting, processing, interpreting, communicating, capturing, and follow-up of a laboratory test, by a healthcare employee, during the process of delivering healthcare to, and in near proximity of the patient.

Urine test strips, haemoglobin, and glucose measurement are classical examples of tests that can be performed point-of-care. D-dimer and C-reactive protein testing are examples of new point-of-care tests who are (gradually) finding their way to primary care.

Our current study only focuses on point-of-care tests performed within healthcare settings (i.e. within the general practice), and therefore does not include self-testing or point-of-care testing at home.

Question for clarification: Do you think that a blood pressure measurement is also a point-of-care test?

Explanation: no, in the current study we only focus on diagnostic tests that use a sample obtained from the human body (for example, blood urine, saliva).

Questions with regard to the list of criteria, their definitions, and the range of outcome measures:

1. Did you already study the list of criteria and their accompanying definitions? (if not, give the participants time to study this list)
2. Are the names of the criteria and their accompanying definitions clear, specific, and unambiguous?
3. Do you miss any criteria? If yes, please specify which, and why.
4. Do you feel that there is overlap between criteria? If yes, which criteria are this, and why? Do you feel that we could merge these criteria into one criterion?
5. Do you feel that there are irrelevant/redundant criteria in this list? If yes, please specify which criteria, and why.
6. During previous interviews, the following changes to this list were proposed (the interviewer specifies the most important changes that were proposed during previous interviews). How do you feel about these changes?
7. Do you feel that the difference between the criteria 'Technical performance', 'Positive Predictive Value', and 'Negative Predictive Value' is clear? Please explain.
8. For the Analytical Hierarchy Process session, it is important to determine a range for every criterion from the list. This may concern absolute numbers (e.g.: Time to perform the test: 4 minutes), or qualitative using a categorical scale (for example, for the criterion 'Connectivity': 'test result is automatically registered in the hospital information system', 'registration of the result requires an additional step (by the general practitioner)', or 'the test result is written down, and manually entered into the hospital information system'. From literature, an initial list of ranges was obtained (which is shown to the participant). Do you have any suggestions for improvement?
9. Do you have any other remarks or questions following this interview?

REFERENCES

1. Hopstaken, R.M., et al., Richtlijn: Point of care testing (POCT) in de huisartsenzorg. 2015.

Analytical Hierarchy Process (AHP) session
Approximate duration: 3 hours

Part I. The tables below show the design of the first part of the AHP session, with the 46 pairwise comparisons in each of the four subcriteria (user, organisation, clinical value, and policy making), as well as the comparisons of these four main criteria amongst each other.

Compare the relative importance of criterion A with criterion B, regarding the implementation of point-of-care (POC) tests in general practices																		
USER																		
extremely - very strong - strong - moderate - equal - moderate -strong - very strong - extremely																		
Satisfaction patient	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Clarity of procedure
Clarity of procedure	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	User-friendliness
User-Friendliness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Test interpretation
Test interpretation	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Turn-around-time
Satisfaction patient	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	User-friendliness
Clarity of procedure	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Test interpretation
User-friendliness	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Turn-around-time
Satisfaction patient	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Test interpretation
Clarity of procedure	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Turn-around-time
Satisfaction patient	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Turn-around-time

ORGANISATION																		
extremely - very strong - strong - moderate - equal - moderate-strong - very strong - extremely																		
Frequency of use	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Room for innovation
Room for innovation	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Workload
Workload	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Support, training and quality control

Support, training and quality control	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Connectivity
Frequency of use	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Workload
Room for innovation	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Support, training and quality control
Workload	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Connectivity
Frequency of use	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Support, training and quality control
Room for innovation	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Connectivity
Frequency of use	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Connectivity

CLINICAL VALUE																		
extremely - very strong - strong - moderate - equal - moderate -strong - very strong - extremely																		
Clinical utility	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Technical performance
Technical performance	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Negative predictive value
Negative predictive value	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Positive predictive value
Positive predictive value	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Risks
Clinical utility	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Negative predictive value
Technical performance	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Positive predictive value
Negative predictive value	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Risks
Clinical utility	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Positive predictive value
Technical performance	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Risks
Clinical utility	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Risks

POLICY MAKING																		
extremely - very strong - strong - moderate - equal - moderate -strong - very strong - extremely																		
Clinical guidelines	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Scientific evidence
Scientific evidence	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Reimbursement
Reimbursement	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Overall costs
Overall costs	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Legislations
Clinical guidelines	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Reimbursement
Scientific evidence	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Overall costs
Reimbursement	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Legislations
Clinical guidelines	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Overall costs
Scientific evidence	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Legislations
Clinical guidelines	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Legislations

CATEGORIES																		
extremely - very strong - strong - moderate - equal - moderate -strong - very strong - extremely																		
User	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Organisation
Organisation	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Clinical value
Clinical value	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Policy making
User	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Clinical value
Organisation	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Policy making
User	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Policy making

Part II. The tables on the next two pages show the design of the second part of the AHP session, in which the criteria are scored when the current situation (perform the test in a central laboratory) is compared to the situation in which a point-of-care (POC) analyser is available within the general practice. POC = point-of-care, QC = quality control.

In this part, the experts are asked to score criterion 'x' when the current situation (central laboratory test) is compared to the new situation (POC test within general practice). Three examples:

1. When considering the criterion 'Satisfaction Patient', which situation do you then prefer, and to what extent?
2. Does the current 'Room for innovation' within the general practice stimulate the use of a central laboratory test or a POC test?
3. Does the criterion 'Clarity of procedure' stimulate the use of a central laboratory test or a POC test? (compare the (entire) procedure performed in case of a POC test, with the situation in which a test is requested at a central laboratory, and the processing of the results once they are received).

C-reactive protein (CRP) test																			
extremely - very strong - strong - moderate - equal - moderate -strong - very strong - extremely																			
USER																			
Satisfaction patient	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Clarity of procedure	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
User-friendliness	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Test interpretation	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Turn-around-time	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
ORGANISATION																			
Frequency of use	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Room for innovation	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Workload	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Support, training, QC	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Connectivity	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
CLINICAL VALUE																			
Clinical utility	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Technical performance	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Negative predictive value	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Positive predictive value	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Risks	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
POLICY MAKING																			
Clinical guidelines	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Scientific evidence	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Reimbursement	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Overall costs	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice
Legislations	CRP reference test at lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	CRP POC test at general practice

Glycated haemoglobin (HbA _{1c})																			
extremely - very strong - strong - moderate - equal - moderate - strong - very strong - extremely																			
		USER																	
Satisfaction patient	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Clarity of procedure	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
User-friendliness	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Test interpretation	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Turn-around-time	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
ORGANISATION																			
Frequency of use	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Room for innovation	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Workload	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Support, training, qc	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Connectivity	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
CLINICAL VALUE																			
Clinical utility	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Technical performance	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Negative predictive value	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Positive predictive value	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Risks	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
POLICY MAKING																			
Clinical guidelines	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Scientific evidence	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Reimbursement	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Overall costs	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice
Legislation	HbA _{1c} reference test lab	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	HbA _{1c} POC test at general practice

APPENDIX 3

Results of interviews and analytical hierarchy process

This supplementary file contains an overview of the detailed results of the analysis of the analytical hierarchy process (AHP), including the weighting of the subcriteria, the group inconsistency and missing judgements during the analytical hierarchy process session, as well as the preferences regarding the subcriteria for the two alternatives (i.e. point-of-care (POC) C-reactive protein (CRP) vs. CRP in a central laboratory, as well as POC glycated haemoglobin (HbA_{1c}) vs. HbA_{1c} in a central laboratory).

Table 1. Performance matrix, showing the criteria included in the AHP, their description, their range, and their relative weight.

Determinant		Description	Range
Determinants in relation to the user			
1	Satisfaction patient	Extent to which the use of the POC test is expected to improve service for the patient.	High - low satisfaction patient
2	Clarity of procedure	Extent to which the procedures for using the POC test are clearly described in protocols and/or manuals.	Clear - unclear
3	User-friendliness	Extent to which the POC test is easy to perform by the layman.	Easy to perform - difficult to perform
4	Test interpretation	Extent to which the POC test result is easy to read and the various test results are easy to interpret.	Easy to interpret - difficult to interpret.
5	Turn-around-time (TAT)	Extent to which the POC test results are instantly available.	1 - 20 minutes (preferably <4 min.)
Determinants in relation to the organisation			
6	Frequency of use	Extent to which the test is used sufficiently with respect to the indication and the size of the general practice.	Often- less
7	Room for innovation	The extent to which the pressure of daily practice leaves room for innovation and a mind-set for change.	None – present
8	Workload	Extent to which the POC test can be implemented without dramatic changes in the current way of working and user's workload.	Increase, equal, decrease
9	Support, training and quality control	Extent to which the introduction, maintenance, and quality control of the POC test, as well as the training of personnel is sufficient and supported by a coordinator from the laboratory, manufacturer and general practice.	No support - coordinators from lab, manufacturer and general practice in lead for contracts and regular up-dates to provide collaboration.

10	Connectivity	Extent to which the POC test results and errors are registered in an information system (HIS)	Hand written - fully automatically registration of results/errors in HIS.
Determinants in relation to the clinical value			
11	Clinical utility	Extent to which a correct (treatment) decision, as based on the point-of-care (POC) test result, has added value in clinical outcomes.	Increase, equal, decrease of clinical outcome for the patient.
12	Technical performance	The extent to which the POC test is exact, precise, reliable and robust in the hands of the user.	Increase, same, decrease.
13	Negative Predictive Value	Negative predictive value: Proportion of negative results that are true negative, which enhances the user's ability to reliably rule out a condition.	Increase, equal, decrease.
14	Positive Predictive Value	Positive predictive value: Proportion of positive results that are true positive, which enhances the user's ability to reliably diagnose a condition.	Increase, equal, decrease.
15	Risks	The impact of a (wrong) treatment/advice based on a (wrong) test result.	Increase, equal, decrease of clinical outcome for the patient.
Determinants in relation to the socio-political context			
16	Clinical guidelines	Extent to which the POC test is implemented in national guidelines.	None – present
17	Scientific evidence	Extent to which the added value of the POC test is demonstrated (as compared with current practice) in scientific literature, in the right patient population and for a specific clinical pathway.	Not available - sufficient and high quality
18	Reimbursement	Is the POC test reimbursed for the general practitioner?	Not / by manufacturer / double consult (time consult >20 min.) / MI-S3 (Innovation incentive by insurer) / standard NZA rate
19	Overall costs	Extent to which the POC test is expected to decrease costs of the health care system.	Increase, equal, decrease.
20	Legislations	Extent to which the innovation fits into existing legislations.	None – present

Table 2. *The inconsistency during the AHP session:*

Group	Group inconsistency	Missing judgments
User	0.07	0/10
Organisation	0.04	2/10
Clinical value	0.01	4/10
Socio-political context	0.09	2/10
Categories	0.05	0/6
Overall	0.05	1.6/9.2

Figures 1a-d. The outcome of the AHP session when comparing the preference of the POC CRP test with the CRP test performed at a central laboratory, for each of the subcriteria. The grey bars represent the relative weight of the subcriteria. The squares represent the preference with regard to the POC CRP test, whereas the triangles represent the preference with regard to the CRP test performed at a central laboratory. The overall weight of the main criterion is shown in the title.

Figure 1a. Result of the AHP analysis on the alternatives POC CRP test and the CRP test performed at a central laboratory for the main criterion ‘user’.

Figure 1b. Result of the AHP analysis on the alternatives POC CRP test and the CRP test performed at a central laboratory for the main criterion ‘organisation’.

Figure 1c. Result of the AHP analysis on the alternatives POC CRP test and the CRP test performed at a central laboratory for the main criterion 'clinical value'.

Figure 1d. Result of the AHP analysis on the alternatives POC CRP test and the CRP test performed at a central laboratory for the main criterion 'socio-political context'.

Figures 2a-d. The outcome of the AHP session when comparing the preference of the POC HbA_{1c} test with the HbA_{1c} test performed at a central laboratory, for each of the subcriteria. The grey bars represent the relative weight of the subcriteria. The squares represent the preference with regard to the POC HbA_{1c} test, whereas the triangles represent the preference with regard to the HbA_{1c} test performed at a central laboratory. The overall weight of the main criterion is shown in the title.

Figure 2a. Result of the AHP analysis on the alternatives POC HbA_{1c} test and the HbA_{1c} test performed at a central laboratory for the main criterion ‘user’.

Figure 2b. Result of the AHP analysis on the alternatives POC HbA_{1c} test and the HbA_{1c} test performed at a central laboratory for the main criterion ‘organisation’.

Figure 2c. Result of the AHP analysis on the alternatives POC HbA_{1c} test and the HbA_{1c} test performed at a central laboratory for the main criterion 'clinical value'.

Figure 2d. Result of the AHP analysis on the alternatives POC HbA_{1c} test and the HbA_{1c} test performed at a central laboratory for the main criterion 'socio-political context'.

12

Towards Alignment in the Reporting of Economic Evaluations of Diagnostic Tests and biomarkers: the AGREEDT checklist

Michelle M.A. Kip
Maarten J. IJzerman
Martin Henriksson
Tracy Merlin
Milton C. Weinstein
Charles E. Phelps
Ron Kusters
Hendrik Koffijberg

Submitted

ABSTRACT

Objectives

General frameworks for conducting and reporting health economic evaluations are available but not specific enough to cover the intricacies of the evaluation of diagnostic tests and biomarkers. Such evaluations are typically complex and model-based because tests primarily affect health outcomes indirectly and real-world data on health outcomes are often lacking. Moreover, not all aspects relevant to the evaluation of a diagnostic test may be known, and explicitly considered for inclusion in the evaluation, leading to a loss of transparency and replicability. To address this challenge, this study aims to develop a comprehensive reporting checklist.

Methods

This study consisted of three main steps: 1) the development of an initial checklist based on a scoping review; 2) review and critical appraisal of the initial checklist by four independent experts; 3) development of a final checklist. Each item from the checklist is illustrated using an example from previous research.

Results

The scoping review followed by critical review by the four experts resulted in a checklist containing 43 items which ideally should be considered for inclusion in a model-based health economic evaluation. The extent to which these items were included, or discussed, in the studies identified in the scoping review varied substantially, with 13 items not being mentioned in ≥ 47 (75%) of the included studies.

Conclusions

The reporting checklist developed in this study may contribute to improved transparency and completeness of model-based health economic evaluations of diagnostic tests and biomarkers. Use of this checklist is therefore encouraged to enhance the interpretation, comparability, and – indirectly – the validity of the results of such evaluations.

INTRODUCTION

Detailed evaluation of the clinical utility and also health economic impact of new diagnostic tests prior to their implementation in clinical practice is important to limit overuse of tests, ensure benefits to patients, and support efficient use of health care resources [1]. Different frameworks have been developed for the phased evaluation of diagnostic tests [2-6]. All these frameworks recognize that after evaluating the safety, efficacy and accuracy of a diagnostic test, the impact of this test on health outcomes and costs should be determined. Evaluating tests in randomized controlled trials (RCTs), however, is often not feasible for ethical, financial, or other reasons, particularly in early test development stages [7-10]. Indeed, RCTs evaluating the impact of diagnostic tests on patient outcomes are rare [11]. As an alternative, method to develop decision-analytic models for the health economic evaluation of diagnostic tests, synthesizing all available evidence from different sources, have long been available [6, 12-16]. It is widely recognized that such models are a useful and valid alternative to evaluate the impact of new health technologies in general [17, 18], and diagnostic methods in particular [12, 14].

However, the comprehensive evaluation of the impact of new tests is typically much more complex than, for example, evaluation of the impact of new drugs. Amongst others, this is due to the indirect impact of tests on health outcomes by improved patient management (also referred to as “clinical utility” [19]), the use of combinations and sequences of tests in clinical practice (depending on previous test results), and the often complex interpretation of test outcomes. In practice, model-based impact evaluations of tests therefore actually involve the evaluation of diagnostic testing strategies (i.e. *test-treatment combinations*).

Owing to the complexity of these diagnostic testing strategies, many model-based impact evaluations of tests make use of simplified models which do not incorporate all aspects of clinical practice. Simplified models are used because 1) evidence regarding all aspects involved in health economic test evaluations might be lacking, 2) inclusion of all aspects likely increases model complexity, or 3) researchers may not be aware of all aspects of test evaluation. For example, it is often not reported how the incremental effect of a new test when used in combination with other tests is determined, and how the correlation between the outcomes of these different tests (applied solo or in sequence) is handled [20-23]. Similarly, the selection of patients in whom the test is performed, the consequences of chance findings, as well as the occurrence of test failures or indeterminate test results are often not reported [24-26]. Although simplifications of the decision-analytic models used for such evaluations may sometimes be necessary and can be adequately justified, implicit simplification due to unawareness of all relevant evaluation aspects, or without proper justification may lead to non-transparent and incorrect evaluation results.

General frameworks and guidelines regarding which aspects to include in decision-analytic modelling and how to report modelling outcomes are available [27-29], but not specific enough to cover the complexities of diagnostic test evaluation. Furthermore, previous research into (aspects of) diagnostic test evaluation mostly focused on specific diseases or on specific types or combinations of diagnostic tests [23, 30-35]. A generic and comprehensive overview of all potentially relevant aspects in health economic evaluation of diagnostic tests and biomarkers that may be used to guide such evaluations, is currently lacking.

The purpose of this paper is, therefore, to provide such an overview as a generic checklist, intended to be applicable to all types of diagnostic tests, and not specific to a single disease or condition, or subgroup of individuals. Thereby, this checklist aims to allow researchers to explicitly consider all aspects potentially relevant to the health economic evaluation of a specific test, from a societal perspective. Use of the checklist does not need to complicate such evaluations, as some aspects described may not be relevant to particular evaluations, but rather suggests that choices to exclude certain aspects are adequately justified.

METHODS

This study consisted of three main steps: 1) the development of an initial checklist based on a scoping review; 2) review and critical appraisal of the initial checklist by four experts (CEP, MCW, MH, and TM) not involved in the scoping review; 3) development of a final checklist based on the review by experts. Finally, each item from the checklist is illustrated using an example from previous research.

Scoping review

In the last decades, hundreds of model-based health economic evaluations of diagnostic tests have been published, across a wide range of medical contexts. A still narrow literature search in PubMed in January 2017, using the following combinations of search terms in title and abstract: (health economic OR cost-effectiveness) AND diagn* AND (model OR Markov OR tree OR modeling OR modelling) already resulted in a total of 1,844 articles. Besides the very large number of studies that have been published in this field, systematic identification of health economic evaluations is found to be challenging [36]. This is partly caused by the multitude of MeSH terms in PubMed related to diagnostic strategies (over 48 MeSH terms exist which include the words 'diagnostic' or 'diagnosis'). Because of these challenges, and the fact that different evaluations are very likely to in- and exclude the same aspects, a scoping review was performed instead of a systematic literature review, followed by critical appraisal by four independent experts. A key strength of a scoping review is that it can provide a rigorous and transparent method for mapping areas of research [37], in particular when an area is complex or has not been reviewed comprehensively before [38].

This scoping review was performed in PubMed in January 2017, searching for the following combination of search terms in the title of the article: (health economic OR cost-effectiveness) AND diagn* AND (model OR Markov OR tree OR modeling OR modelling) NOT diagnosed. The term 'NOT diagnosed' was added to prevent retrieving many articles including patients who are already diagnosed with a certain condition, instead of focusing on the diagnostic process itself. The search was limited to articles published in the last five years, in English or Dutch. Studies were excluded, based on title and abstract, if they did not concern original research, or did not evaluate the cost-effectiveness of the use of one or more tests (regardless of the effectiveness measure, for example, additional cost per additional correct diagnosis, or per additional quality-adjusted life year). One author screened studies for exclusion (MMAK) and consulted with a second author (HK) if necessary.

Design of the reporting checklist

All articles resulting from the scoping review were searched for items related to model-based health economic test evaluation of diagnostic tests that were either included explicitly in the evaluation, or that were only mentioned but not included (mostly in the introduction or discussion sections). Generic items, not specific to diagnostic test evaluation were not included in the new checklist as these are already covered in existing checklists. Examples of such generic items include choosing the time horizon and perspective of the evaluation [27-29]. However, some overlap remains as the checklist does include items which are considered applicable to diagnostic test evaluation that are only covered partially or at a high level in existing guidelines.

A thorough screening of all articles was performed by MMAK resulting in an initial list of items considered to be relevant. The decision to include an aspect in the checklist, and the definition of this aspect, was based upon agreement between MMAK and HK.

Critical appraisal and validation of the reporting checklist

As diagnostic tests and imaging are used for a large variety of (suspected) medical conditions, an expert panel with a broad field of experience was required for critical appraisal of the checklist. Therefore, the expert panel was composed in such a way that at least one expert was experienced in each of the different areas of interest (i.e. biomarkers or imaging), and in each of the different purposes of diagnostic testing (i.e. diagnosis, screening, monitoring and prognosis). In addition, to maximize the likelihood that the final checklist is generalizable to different countries and settings, the experts chosen lived in three different continents. Four experts were invited (CEP, MCW, MH, and TM) to participate via email, and none of them declined.

The initial checklist was critically appraised and validated independently by all four experts, who received the checklist via email. They were asked to provide individual, qualitative

judgments on whether all items in this list were clear and unambiguous, to indicate any missing or redundant items in this list, and to provide suggestions for further improvement.

Finalization of the reporting checklist

Based on the experts' suggestions, several changes were made to the reporting checklist. Those changes involved the rewording of items, removal of redundant items, as well as the addition of missing items to the checklist. As this checklist is intended to provide an exhaustive list of all aspects relevant to the health economic evaluations of diagnostic tests and biomarkers, all suggestions for the addition of missing items were adopted. All changes made to the checklist were decided upon agreement between MMAK and HK (for a full description see Appendix 1). The revised checklist was again critically appraised by all authors, and agreed upon. Finally, all articles included in the scoping review were reread by MMAK to assess whether the final checklist items were included or mentioned.

RESULTS

Results of the scoping review

The literature search resulted in 77 articles that were screened for inclusion in the scoping review, of which 14 articles were excluded. Of these, four articles did not specifically evaluate the cost-effectiveness of a (combination of) diagnostic test(s), two concerned a letter to the editor, and seven articles focused on methodological aspects of the evaluation of diagnostic strategies, for example, in the context of single disease, or on specific types or combinations of diagnostic tests (as mentioned in the introduction). In addition, one article was excluded because the full text could not be obtained or purchased by the university library, from online databases, the website of the publisher, or by contacting the authors. This resulted in a total of 63 studies that were identified from the scoping review. An overview of this selection process is provided in Figure 1.

A critical evaluation of the 63 articles resulted in an initial list of 29 items. These items were divided in six main topics: (1) time to presentation of the individual to the health professional (i.e. the clinical starting point), (2) use of diagnostic tests, (3) test performance and characteristics, (4) patient management decisions, (5) impact on health outcomes and costs, and (6) wider societal impact, which may accrue to patients, their families, and/or health care professionals. This societal impact may for example concern the impact on caregivers (in terms of time spent on hospital visits and caregiving, and the accompanying impact on productivity), on the health system or health professional (for example in terms of reduced patient visits), or on society (for example measures that aim to prevent widespread antibiotic resistance). Quantifying these aspects may provide a broader view on the potential impact of diagnostic testing.

Figure 1. Result of scoping review and checklist design process. This figure first gives an overview of the selection process of articles in the scoping review, the number of checklist items this resulted in, and subsequently shows the results of the expert appraisal on the items included in the final checklist.

Critical appraisal and validation of the reporting checklist

Following the critical appraisal of experts the list was updated, with one item being removed, and 15 items being added, resulting in a final reporting checklist consisting of 43 items, as shown in Table 1. Of these 15 added items, seven involved a further specification of the tests’ diagnostic performance, as included below item 3.2 in the checklist. The item that was removed considered the generalizability of the results, which was considered not to be specific to diagnostic test evaluations. The full reporting checklist, including an overview in which of the studies from the scoping review each of the items was included or considered, as well as an example for each of the items, is provided in Appendix 2. An overview of this process, including the scoping review and the critical appraisal by the experts, is shown in Figure 1. The final list of items in this reporting checklist, in chronological order from the start of the diagnostic trajectory and onwards, is illustrated in Figure 2.

Table 1: Reporting checklist to indicate which items were included in the health economic evaluation of diagnostic tests and biomarkers. *If an item is included in the quantitative analysis, indicate the corresponding model parameter(s) and evidence source(s). **If an item is excluded from the quantitative analysis, please explain why the exclusion was necessary.

ITEMS OF THE EVALUATION OF DIAGNOSTIC TESTS AND BIOMARKERS		INCLUDED IN EVALUATION	
TIME TO PRESENTATION			
Onset of complaints or onset of suspicion by physician and start of diagnostic trajectory		YES*	NO**
1.1	The study should contain a description of the individuals that enter the diagnostic pathway (i.e. a patient develops a (new) condition, or disease, which may or may not result in symptoms or complaints, or undergoes diagnostic testing as part of screening or genetic testing).		
1.2	Consider the time to start of the diagnostic trajectory, or the time until a monitoring or screening test is (repeatedly) performed (initiated by symptoms/complaints or initiated as part of regular screening or monitoring). <i>(The time between 1.1 and 1.2 is the time during which individuals are at risk of complications from disease and progression, in the absence of a diagnosis and thus also in the absence of treatment).</i>		
USE OF DIAGNOSTIC TESTS			
Decision regarding which diagnostic test(s) is/are performed, in which patients, and in what order		YES*	NO**
2	Specify for which purpose(s) the test is used (e.g. screening, diagnosing, monitoring, guide dosage, commencement or cessation of therapy, triaging, staging, prognostic) and define the entire diagnostic and clinical pathway.		
2.1	Consider whether more than two (possible) diagnostic strategies can be compared, each involving a single test or combination of tests.		
2.2	Consider whether the evaluated diagnostic strategies include multiple tests, which can be performed in parallel or in sequence).		
2.2.1	Consider whether some tests of the diagnostic work-up are performed conditional on previous test outcomes, leading to a selection of patients undergoing specific tests.		
2.3	Consider whether subgroups can be defined based on explicit criteria or patient characteristics, in which different tests would be performed (not solely dependent on previous test outcomes).		
2.4	Consider whether different tests are applied based on implicit (shared) decision-making (for example perceived condition or risk, or symptom presentation).		

TEST PERFORMANCE AND CHARACTERISTICS			
Diagnostic test performance and items related to the sampling and testing		YES*	NO**
3.1	Specify the costs of the diagnostic test.		
3.2	Specify test performance, in terms of sensitivity, specificity, negative predictive value (and its complement) and/or positive predictive value (and its complement), either or not combined with a decision rule or algorithm.		
3.2.1	Describe the evidence base for the estimated test performance.		
3.2.2	Describe the positivity criterion (i.e. cut-off value) applied to the test or testing strategy.		
3.2.3	Consider whether the estimated test performance may be biased, for example due to lack of evidence on conditional dependence or independence, lack of a (perfect) gold standard (i.e. classification bias), verification bias, analytic bias, spectrum bias, diagnostic review bias and incorporation bias.		
3.2.4	Describe how uncertainty in the test performance (ROC curve) was handled or explained, for example due to inter-rater and intra-rater reliability, or experience of the clinician.		
3.2.5	Describe the logic, or analysis, applied to choose the cut-off value (i.e. the point on the ROC curve) for the test		
3.2.6	Describe whether different test performances and cut-off values were considered for different subgroups of patients and/or environmental characteristics. For example: based on specific subgroup(s) of patients, timing of the test in the diagnostic trajectory, or selection of patients based on previous test outcomes (if any).		
3.2.7	Consider whether test performance is dependent on disease prevalence.		
3.2.8	Consider whether test performance is based on a combination of tests (and on a combination of areas under the Receiver Operating Characteristic (ROC) curves for each test).		
3.3	Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained.		
3.4	Consider the occurrence of test failures or indeterminate/not assessable results.		
3.5	Consider costs of retesting (after obtaining insufficient/unusable sample, or after test failure or indeterminate/not assessable result).		
3.6	Consider complications, risks or other negative/positive aspects directly related to obtaining the sample and/or performing the diagnostic test (either in the intervention or in the control strategy).		
3.7	Consider the time taken to: perform the test (including waiting time), until the test result is available, or until a management decision or treatment is initiated based on this test result (either in the intervention or in the control strategy).		

3.8	Consider the impact of additional knowledge gained by performing the diagnostic test (i.e., for a genetic test), or the occurrence and impact of chance findings.		
3.8.1	The impact of chance findings on performing additional tests is addressed.		
PATIENT MANAGEMENT DECISIONS			
Impact of a test on the diagnosis and/or patient management strategy (based on this diagnosis)		YES*	NO**
4.1	Clearly specify the impact of the test in selecting the patient management strategy.		
4.2	Consider whether other aspects besides test results themselves are part of the decision algorithm (and included in the evaluation). These may involve a shared decision-making process of the physician with patients/relatives, or aspects including coverage or physician adherence to treatment guidelines).		
4.3	Consider whether the impact of the test result on resulting/selected diagnosis or management strategy varies across subgroups (this difference should not only be caused by differences in diagnostic performance of the test, and does not need to include the impact on costs and/or health outcomes within this subgroup).‡		
4.4	Consider the consistency of test results over time (for example: genetic mutations may be affected by treatment prescribed after the initial diagnosis).		
4.5	Consider the impact of performing the test and providing and interpreting the result on the time spend/capacity of the health care professional(s) or the patient.		
IMPACT ON HEALTH OUTCOMES AND COSTS			
Impact of the patient management strategy on diseased and non-diseased individuals, in terms of health outcomes and costs		YES*	NO**
5	Evaluate the direct impact of the chosen patient management strategy on the number of (in)correctly diagnosed individuals, health outcomes, and/or costs.		
5.1	Consider the direct impact of the chosen patient management strategy, on health outcomes and/or costs. This concerns the entire period in which patient management may affect a patient's health and/or costs, and does not only involve the testing strategy itself.		
5.2	Consider whether the direct impact of the chosen patient management strategy on health outcomes and/or costs varies across subgroups. (This does not include only varying the incidence of a certain condition in a sensitivity analysis. The subgroups should be clearly defined, and preferably be identifiable based on patient characteristics).‡		
5.3	Consider patient's adherence to treatment (which includes aspects that may indicate (partial) non-adherence, for example, following only some of the treatment recommendations, as well as aspects that affect the degree of administration of treatment).		
5.4	Consider the occurrence (and consequences) of treatment-related adverse events.		

5.5	Describe the probability, or time it takes to observe that the patient management strategy proves to be effective over time, or that the patient cures spontaneously (regardless whether the patient received a correct or an incorrect diagnosis).		
5.6	Describe the probability of, or time it takes to repeat or extend the diagnostic work-up when the patient management strategy proves to be ineffective, either directly or over time (regardless whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which the patient receives no treatment, or unnecessary treatment.		
5.7	Describe the impact of ineffective or unnecessary treatment or management on health outcomes and/or costs (including both side effects and costs, and regardless whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which incorrectly no treatment is provided, or in which the treatment is delayed.		
5.7.1	Consider the impact of delay in treatment initiation on health outcomes and/or costs		
SOCIETAL IMPACT			
wider (societal) impact of the chosen diagnostics and management strategy		YES*	NO**
6.1	Consider the psychological impact of diagnostic outcome and management strategy on patients, value of knowing (in terms of reassurance or anxiety), patient preferences regarding undergoing diagnostic tests, and the (accompanying) impact on caregivers, or relatives, etc.		
6.1.1	Consider the impact of test outcomes on relatives themselves, regarding the value of knowing (spillover knowledge) and regarding subsequent testing and/or treatment in this group (in case of heritable genetic conditions, or contagious diseases).		
6.2	Consider the additional impact of diagnostic outcome and management strategy on the health system or healthcare professionals.		
6.3	Consider the additional impact of diagnostic outcome and management strategy for society.		

‡ Existing guidelines indicate that subgroup analyses are relevant when different strategies are likely to be (sub)optimal in different subgroups. Subgroup specific analyses can then be performed to address multiple decision problems. Here we consider scenarios where different tests may be used in different subgroups, depending on patient characteristics or previous test outcomes.

Figure 2. Overview of steps in diagnostic trajectory.

This figure gives a conceptual outline of the steps involved in the diagnostic trajectory, in chronological order from top to bottom. The numbers shown at the several steps correspond to the item numbers presented in Table 1.

The dashed lines represent steps of which the duration may vary substantially, for example the time between symptom onset and presentation to a clinician (which may vary from minutes in case of severe symptoms, to years for mild and gradually developing conditions). The arrows indicate situations in which either the diagnostic test (result) was not usable or indeterminate (item 3.3 - 3.5), or situations in which the treatment proves to be ineffective (items 5.5 - 5.7). As this may be caused by an incorrect diagnosis, the patient may undergo a subsequent round of diagnostic testing and (possibly) treatment. Alternatively, the diagnosis may be correct, but the treatment incorrect, in which an alternative treatment may be initiated.

*Although the (wider) societal impact of diagnostic testing often involves long-term effects, these effects may sometimes also become apparent in the short term.

Results indicate that health economic evaluations of diagnostic tests or biomarkers differ considerably in the items that have been explicitly included (or considered for inclusion) in the corresponding decision-analytic model (Table 1). Some of the items from the checklist were only included (or considered) in a few studies from the scoping review. For example, the impact of chance findings on performing additional tests, the consistency of test results over time, and the impact of test outcomes on relatives themselves were each only addressed in 3 of the 63 included studies. These items may not have been included in other studies because they were considered not relevant to the specific context, or because they were not considered due to unawareness of their relevance by the authors.

DISCUSSION

Strengths

A strength of this study is that it combines evidence from multiple sources, including a review of literature, as well as a validation by experts. As the items included in the checklist are defined in general terms, not limited to specific diseases, tests, care providers or patient management strategies, this reporting checklist can potentially be useful in performing and appraising health economic evaluations worldwide and across a broad spectrum of (novel) diagnostic technologies. In addition, as this checklist specifically focuses on health economic evaluations of diagnostic tests or biomarkers, an area for which no reporting checklists are yet available, it may be a useful extension to existing reporting checklists, such as the CHEERS checklist [27]. Finally, use of this checklist can also support *development* of health economic models through increased awareness of all potentially relevant evaluation aspects.

In addition, use of this checklist does not necessarily require more resources to be allocated to the evaluation, or increase the complexity of the resulting decision-analytic model. In general, deliberation on the relevance of all aspects is key, and aspects may be excluded from the evaluation whenever this can be adequately justified. For example, when evaluating the cost-effectiveness of a new point-of-care (POC) troponin test used by the general practitioner as compared to an existing, older POC troponin test, in the context where the new test would replace the old test, aspects such as ‘time to start of the diagnostic trajectory’, and ‘purpose of the test’ will not differ between strategies. In addition, ‘complication risks’ associated with taking the blood sample (in both POC tests) are likely extremely small, which could justify excluding these aspects from the analysis.

Limitations

Performing a systematic literature review was considered not possible given the very large number of published economic evaluations of diagnostic tests. Therefore, a scoping review was performed instead, by one reviewer. As the judgement regarding whether an aspect was incorporated in a health economic evaluation was sometimes found to be difficult, it cannot be excluded that these judgements may have differed slightly when performed by a different

reviewer. However, as four independent experts from different countries have critically reviewed the checklist, it is unlikely that important aspects have been missed. The scoping review may however have been subject to publication bias, as it may have omitted potentially relevant aspects from unpublished studies, as well as from method manuals (including those focusing on economic evaluations of other interventions or technologies in healthcare).

In addition, the expert appraisal resulted in the addition of 15 items to the checklist. Although this may seem to be a large extension to the items already identified in the scoping review, seven of these added items actually involved a further specification (i.e. a sub-item) of the test's diagnostic performance. It was found useful to further specify 'test performance' (i.e. item 3.2, which initially integrated several performance measures) into seven sub-items to further increase the transparency and comparability of health economic test evaluations.

Implications for practice

This study was intended to design a reporting checklist without formulating a quality judgement of the studies included in the scoping review, based on which items of the checklist they did or did not incorporate. Furthermore, some items may have been included implicitly in the health economic evaluations identified in the scoping review, which could thus not be identified by the reviewer. As scientific articles are often restricted in their length, there may often be insufficient space to mention the inclusion (or justified exclusion) of each of the items from this checklist. In these situations, authors are recommended to describe their use of this checklist in an appendix. More specifically, authors are recommended to describe which items from the checklist they included in their evaluation, and what evidence was used to inform them. Furthermore, they are recommended to explicitly state the reason(s) for excluding checklist items from their evaluation. Although it may be considered time-consuming to consider all 43 items of this checklist, it should be noted that the majority of these items are actually sub-items, which do not need to be considered if the overarching (higher-level) item is (justifiably) excluded from the evaluation.

In addition, it should be noted that not all items in this checklist can be considered of equal importance. For example, diagnostic performance will typically have a larger impact on health outcomes and costs when compared to considering the occurrence of test failures, or the consistency of test results over time. However, this checklist is designed to provide an exhaustive overview of all potentially relevant items, regardless of importance. Therefore, use of this checklist will likely increase the chance that all relevant aspects will be included in health economic evaluations of diagnostic tests and biomarkers. Ultimately it is up to the researchers to make a justifiable decision on which items to incorporate and which to exclude.

Finally, experiences regarding the use of this reporting checklist in practice may be valuable to further enhance its completeness and usability. Furthermore, given the rapid methodological

developments in the field of health economic evaluation of diagnostic tests regular updating of this checklist may be warranted.

Conclusion

Given the complexity and dependencies related to the use of diagnostic tests or biomarkers, researchers may not always be fully aware of all the different aspects potentially influencing the result of a model-based health economic evaluation. The use of the reporting checklist developed in this study may remedy this by increasing awareness of all potentially relevant aspects involved in such model-based health economic evaluations of diagnostic tests and biomarkers, and thereby also increase the transparency, comparability, and – indirectly – the validity of the results of such evaluations.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflicts of interest

Prof. Merlin reports that she was previously commissioned by the Australian Government to develop version 5.0 of the 'Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee'. Some of the content concerning 'Product Type 4 - Codependent Technologies' influenced the guidance suggested in the current paper. Prof. Weinstein reports that he is a consultant to OptumInsight on unrelated topics. All other authors declare that there is no conflict of interest.

REFERENCES

1. Price, C.P., Evidence-based laboratory medicine: is it working in practice? *Clin Biochem Rev*, 2012. 33(1): p. 13-9.
2. Ferrante di Ruffano, L., et al., Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*, 2012. 344: p. e686.
3. Guyatt, G.H., et al., A framework for clinical evaluation of diagnostic technologies. *CMAJ*, 1986. 134(6): p. 587-94.
4. Silverstein, M.D. and B.J. Boland, Conceptual framework for evaluating laboratory tests: case-finding in ambulatory patients. *Clin Chem*, 1994. 40(8): p. 1621-7.
5. Anonychuk, A., et al., A framework for assessing the value of laboratory diagnostics, in *Healthcare Management Forum - Fall 2012*. 2012.
6. Phelps, C.E. and A.I. Mushlin, Focusing technology assessment using medical decision theory. *Med Decis Making*, 1988. 8(4): p. 279-89.
7. Schunemann, H.J., et al., Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*, 2008. 336(7653): p. 1106-10.
8. Moons, K.G., Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem*, 2010. 56(4): p. 537-41.
9. Bossuyt, P.M., J.G. Lijmer, and B.W. Mol, Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*, 2000. 356(9244): p. 1844-7.
10. Biesheuvel, C.J., D.E. Grobbee, and K.G. Moons, Distraction from randomization in diagnostic research. *Ann Epidemiol*, 2006. 16(7): p. 540-4.
11. Ferrante di Ruffano, L., et al., A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J Clin Epidemiol*, 2012. 65(3): p. 282-7.
12. Schaafsma, J.D., et al., Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-eff-

- fectiveness. *J Clin Epidemiol*, 2009. 62(12): p. 1248-52.
13. Bossuyt, P.M.M. and K. McCaffery, Additional Patient Outcomes and Pathways in Evaluations of Testing, in *Medical Tests-White Paper Series*. 2009: Rockville (MD).
 14. Koffijberg, H., B. van Zaane, and K.G. Moons, From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol*, 2013. 13: p. 12.
 15. Merlin, T., The use of the 'linked evidence approach' to guide policy on the reimbursement of personalized medicines. *Personalized Medicine*, 2014. 11(4): p. 435-448.
 16. Merlin, T., et al., The "linked evidence approach" to assess medical tests: a critical analysis. *Int J Technol Assess Health Care*, 2013. 29(3): p. 343-50.
 17. Briggs, A. and K. Claxton, *Decision Modelling for Health Economic Evaluation*, in NY. 2006, Oxford University Press: New York.
 18. Sculpher, M.J., et al., Whither trial-based economic evaluation for health care decision making? *Health Econ*, 2006. 15(7): p. 677-87.
 19. Bossuyt, P.M., et al., Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem*, 2012. 58(12): p. 1636-43.
 20. Jethwa, P.R., et al., Cost-effectiveness of digital subtraction angiography in the setting of computed tomographic angiography negative subarachnoid hemorrhage. *Neurosurgery*, 2013. 72(4): p. 511-9; discussion 519.
 21. Koffijberg, H., Cost-effectiveness analysis of diagnostic tests. *Neurosurgery*, 2013. 73(3): p. E558-60.
 22. Novielli, N., A.J. Sutton, and N.J. Cooper, Meta-analysis of the accuracy of two diagnostic tests used in combination: application to the ddimer test and the wells score for the diagnosis of deep vein thrombosis. *Value Health*, 2013. 16(4): p. 619-28.
 23. Novielli, N., N.J. Cooper, and A.J. Sutton, Evaluating the cost-effectiveness of diagnostic tests in combination: is it important to allow for performance dependency? *Value Health*, 2013. 16(4): p. 536-41.
 24. Otero, H.J., et al., Accuracy, risk and the intrinsic value of diagnostic imaging: a review of the cost-utility literature. *Acad Radiol*, 2012. 19(5): p. 599-606.
 25. Ding, A., J.D. Eisenberg, and P.V. Pandharipande, The economic burden of incidentally detected findings. *Radiol Clin North Am*, 2011. 49(2): p. 257-65.
 26. Mueller, M., et al., Subgroup Discovery for Test Selection: A Novel Approach and Its Application to Breast Cancer Diagnosis, in *Advances in Intelligent Data Analysis VIII: 8th International Symposium on Intelligent Data Analysis, IDA 2009, Lyon, France, August 31 - September 2, 2009. Proceedings*, N.M. Adams, et al., Editors. 2009, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 119-130.
 27. Husereau, D., et al., Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health*, 2013. 16(2): p. e1-5.
 28. Philips, Z., et al., Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*, 2006. 24(4): p. 355-71.
 29. Sanders, G.D., et al., Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA*, 2016. 316(10): p. 1093-103.
 30. Dowdy, D.W., et al., Impact and cost-effectiveness of current and future tuberculosis diagnostics: the contribution of modelling. *Int J Tuberc Lung Dis*, 2014. 18(9): p. 1012-8.
 31. Wimo, A., et al., Health economic evaluation of treatments for Alzheimer's disease: impact of new diagnostic criteria. *J Intern Med*, 2014. 275(3): p. 304-16.
 32. Raymakers, A.J., et al., Diagnostic strategies incorporating computed tomography angiography for pulmonary embolism: a systematic review of cost-effectiveness analyses. *J Thorac Imaging*, 2014. 29(4): p. 209-16.
 33. Oosterhoff, M., M.E. van der Maas, and L.M. Steuten, A Systematic Review of Health Economic Evaluations of Diagnostic Biomarkers. *Appl Health Econ Health Policy*, 2016. 14(1): p. 51-65.
 34. Sailer, A.M., et al., Cost-effectiveness modelling in diagnostic imaging: a stepwise approach. *Eur Radiol*, 2015. 25(12): p. 3629-37.
 35. Lee, J., et al., A generalized health economic and outcomes research model for the evaluation of companion diagnostics and targeted therapies. *Expert Rev Pharmacoecon Outcomes Res*, 2013. 13(3): p. 361-70.
 36. Glanville, J. and S. Paisley, Identifying economic evaluations for health technology assessment. *Int J Technol Assess Health Care*, 2010. 26(4): p. 436-40.
 37. Arksey, H. and L. O'malley, Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 2005. 8(1): p. 19-32.
 38. Mays, N., E. Roberts, and J. Popay, Synthesising research evidence, in *Methods for studying the delivery and organisation of health services*, N.F.a.P.A.a.A.C.a.N. Black, Editor. 2001, Routledge: London.

APPENDIX 1

Description of items added to or excluded from the checklist, based on experts' critical revisions

Items added to the checklist

Category	Item
Test performance and characteristics	Specify the costs of the diagnostic test.
	Describe the evidence base for the estimated test performance.
	Describe the positivity criterion (i.e. cut-off value) applied to the test or testing strategy.
	Consider whether the estimated test performance may be biased, for example due to lack of evidence on conditional dependence or independence, lack of a (perfect) gold standard (i.e. classification bias), verification bias, analytic bias, spectrum bias, diagnostic review bias and incorporation bias.
	Describe how uncertainty in the test performance (ROC curve) was handled or explained, for example due to inter-rater and intra-rater reliability, or experience of the clinician.
	Describe the logic, or analysis, applied to choose the cut-off value (i.e. the point on the ROC curve) for the test.
	Consider whether test performance is dependent on disease prevalence.
	Consider whether test performance is based on a combination of tests (and on a combination of areas under the Receiver Operating Characteristic (ROC) curves for each test).
	Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained.
	Consider costs of retesting (after obtaining insufficient/unusable sample, or after test failure or indeterminate/not assessable result).
Patient management decisions	The impact of chance findings on performing additional tests is addressed.
	Clearly specify the impact of the test in selecting the patient management strategy.
	Consider the consistency of test results over time (for example: genetic mutations may be affected by treatment prescribed after the initial diagnosis).
Societal impact	Consider the impact of performing the test and providing and interpreting the result on the time spend/capacity of the health care professional(s) or the patient.
	Consider the impact of test outcomes on relatives themselves, regarding the value of knowing (spillover knowledge) and regarding subsequent testing and/or treatment in this group (in case of heritable genetic conditions, or contagious diseases).

Items removed from the checklist

Category	Item
Societal impact	Likelihood that the results can be generalized to other countries and/or settings (for example different patient groups based on age, severity or other characteristics).

Description of reasons for changes made to the checklist

Items added to the checklist

Eleven items were added to the category ‘test performance and characteristics’. The item ‘specify the costs of the diagnostic test’ was included as this is specific for and very important in the health economic evaluation of diagnostic tests. The following seven items (as shown on the previous page) were added with the goal to more accurately specify how the diagnostic performance was either estimated or derived from literature, and to incorporate uncertainty or bias of this diagnostic performance in the health economic evaluation.

The item ‘Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained’ was added to the checklist, whereas the previous version of the checklist incorporated this issue within the item ‘usefulness or usability of the results, or usability of the sample that is obtained’. Therefore, this item in the checklist was changed to ‘consider costs of retesting (after test failure or indeterminate/not assessable result)’, in order to assess the usability of the sample and the usability of the test result in separate items. In addition, the costs of retesting after an indeterminate/not assessable result, or after obtaining an unusable sample, was added.

In addition, although the impact of diagnostic tests on chance findings was already included in the checklist, it was decided to explicitly include an item regarding addressing the impact of chance findings on performing additional tests.

In the category ‘patient management decisions’, three items were added which aim to 1) explicitly specify the aim of the diagnostic test (e.g. is the test used for diagnosis, screening, disease staging), 2) to consider the consistency of test results over time (as outcomes of diagnostic tests may change over time, for example due to spontaneous cure, or by the occurrence of genetic mutations due to treatment), and 3) to not only consider the time the test itself takes, but also the time it takes to perform the test (by the health care professional) and the time it takes to interpret the results.

Finally, in the category ‘societal impact’, the previous version of the checklist already incorporated the impact of a diagnostic test on relatives in terms of reassurance or caregiving, but this item did not specifically address the impact of a test on relatives themselves. Therefore,

an item was added to specify the impact of test outcomes on subsequent testing and/or treatment of relatives.

Items removed from the checklist

The item regarding the generalizability of the study's findings was excluded as it was agreed upon that this item should be included/considered in any health economic evaluation, and is not specific to diagnostic test evaluation.

APPENDIX 2

Table 1. Overview and definitions of the final reporting checklist. This table shows the checklist of items that should be considered for inclusion in health economic evaluations of diagnostics tests and biomarkers. Each item is illustrated using an example from a selection of previously published papers.

TIME TO PRESENTATION			Item included in study	Item mentioned (but not included)
Onset of complaints or onset of suspicion by physician and start of diagnostic trajectory				
1.1	The study should contain a description of the individuals that enter the diagnostic pathway (i.e. a patient develops a (new) condition, or disease, which may or may not result in symptoms or complaints, or undergoes diagnostic testing as part of screening or genetic testing).	Example from study: [1] Patients present with chest pain at the general practitioner.	[2-64]	
1.2	Consider the time to start of the diagnostic trajectory, or the time until a monitoring or screening test is (repeatedly) performed (initiated by symptoms/complaints or initiated as part of regular screening or monitoring).(The time between 1.1 and 1.2 is the time during which individuals are at risk of complications from disease and progression, in the absence of a diagnosis and thus also in the absence of treatment).	Example from study: [65] In patients with non-traumatic intracerebral hemorrhage, the time since onset of symptoms and presentation at the emergency department influences the clinical condition at presentation.	[2, 4, 13, 19, 20, 33, 34, 37, 43, 57]	[25, 53, 56]
USE OF DIAGNOSTIC TESTS			Item included in study	Item mentioned (but not included)
Decision regarding which diagnostic test(s) is/are performed, in which patients, and in what order				
2.	Specify for which purpose(s) the test is used (e.g. screening, diagnosing, monitoring, guide dosage, commencement or cessation of therapy, triaging, staging, prognostic) and define the entire diagnostic and clinical pathway.	Example from study: [65] Computed tomography angiography can be used to detect macrovascular causes in patients with non-traumatic intracerebral hemorrhage.	[2-64]	
2.1	Consider whether more than two (possible) diagnostic strategies can be compared, each involving a single test or combination of tests.	Example from study: [13] The main economic analysis considered five strategies for comparing the incremental cost-effectiveness of high-sensitive troponin testing to standard troponin testing (reference).	[2-5, 7-9, 11-13, 16-19, 21, 24, 26-28, 30, 31, 34, 37-39, 41, 42, 46, 47, 49, 53, 55-64]	[15, 20, 22, 29, 33]

2.2	Consider whether the evaluated diagnostic strategies include multiple tests, which can be performed in parallel or in sequence).	Example from study: [65] Patients with non-traumatic intracerebral hemorrhage first undergo computed tomography angiography. Patients with a negative result undergo additional testing.	[3-6, 8-14, 16-19, 21, 24-49, 51-54, 56-64]	[2]
2.2.1	Consider whether some tests of the diagnostic work-up are performed conditional on previous test outcomes, leading to a selection of patients undergoing specific tests.	Example from study: [65] In patients with non-traumatic intracerebral hemorrhage, tests are performed in sequence conditional on previous test outcomes (see 2.2).	[3-6, 8-10, 14, 16-18, 20, 24-28, 30-40, 42-48, 51, 53, 54, 56-62, 64]	[13, 49]
2.3	Consider whether subgroups can be defined based on explicit criteria or patient characteristics, in which different tests would be performed (not solely dependent on previous test outcomes).	Example from study: [1] The point-of-care troponin test is only performed in patients with at least 4 hours chest pain.	[3, 27, 34, 40, 44, 48, 54]	[2, 9, 13, 26, 30, 32, 36, 46, 51]
2.4	Consider whether different tests are applied based on implicit (shared) decision-making (for example perceived condition or risk, or symptom presentation).	Example from study: [1] In chest pain patients, the general practitioner decides, based on the perceived risk of the patient, whether to use a point-of-care troponin test and/or electrocardiogram, or to immediately refer the patient to the hospital.	[9, 10, 14, 17, 34, 35, 37, 43, 44, 51, 54, 58, 63]	[13, 20, 26, 27, 30-32, 40, 46, 55, 57, 61]
TEST PERFORMANCE AND CHARACTERISTICS			Item included in study	Item mentioned (but not included)
Diagnostic test performance and items related to the sampling and testing				
3.1	Specify the costs of the diagnostic test.	Example from study: [1] The costs of the point-of-care troponin test were estimated at €15.	[2-22, 24-64]	
3.2	Specify test performance, in terms of sensitivity, specificity, negative predictive value (and its complement) and/or positive predictive value (and its complement), either or not combined with a decision rule or algorithm.	Example from study: [1] The accuracy of the point-of-care troponin test is based on previously published literature.	[2-22, 24-64]	

3.2.1	Describe the evidence base for the estimated test performance.	Example from study: [1] The sensitivity and specificity of the point-of-care troponin test were derived from Diercks et al.	[2-11, 13-22, 24-63]	
3.2.2	Describe the positivity criterion (i.e. cut-off value) applied to the test or testing strategy.	Example from study: [66] Anemia was defined as hemoglobin below 13.7 g/dL (8.5 mmol/L) for males and below 12.1 g/dL (7.5 mmol/L) for females.	[2-4, 10, 19, 30, 31, 40, 42, 44, 56, 59]	[9, 13]
3.2.3	Consider whether the estimated test performance may be biased, for example due to lack of evidence on conditional dependence or independence, lack of a (perfect) gold standard (i.e. classification bias), verification bias, analytic bias, spectrum bias, diagnostic review bias and incorporation bias.	Example from study: [51] Without a true reference standard diagnostic test for latent tuberculosis infection, it is impossible to ascertain the true proportion of individuals with latent tuberculosis infections among those referred with a positive tuberculin skin test.	[4, 9, 16, 22, 30, 42, 45, 51, 62-64]	[2, 3, 7, 15, 21, 29, 31, 34, 44, 47-50, 53, 55, 59]
3.2.4	Describe how uncertainty in the test performance (ROC curve) was handled or explained, for example due to inter-rater and intra-rater reliability, or experience of the clinician.	Example from study: [51] Inter-reader variability of tuberculin skin test may lead to misclassifications and FPs, particularly in low-prevalence populations.	[2, 4, 7, 17, 29, 30, 42, 44, 45]	[5, 19, 51, 55]
3.2.5	Describe the logic, or analysis, applied to choose the cut-off value (i.e. the point on the ROC curve) for the test	Example from study: [4] A cataract test categorized a patient as needing evaluation by a clinician if any elements of a composite triage test was 'abnormal' in either eye: imaging classification 'outside normal limits' or intraocular pressure > 21 mmHg or visual activity of 6/12 or poorer.	[3, 4, 31, 40, 42]	[9, 10, 13, 36, 44]
3.2.6	Describe whether different test performances and cut-off values were considered for different subgroups of patients and/or environmental characteristics. For example: based on specific subgroup(s) of patients, timing of the test in the diagnostic trajectory, or selection of patients based on previous test outcomes (if any).	Example from study: [4] At the start of the study, principal investigators reviewed and agreed on the reference standard (definitions of glaucoma, ocular hypertension, glaucoma suspect and normal), and how to define the spectrum of the disease (mild, moderate, and severe).	[4, 21, 32, 37-40, 42, 44, 46, 48, 52, 53, 56, 58, 62]	[2, 9, 11, 13, 22, 24, 34, 36, 45, 54, 55]

3.2.7	Consider whether test performance is dependent on disease prevalence.	Example from study: [51] Inter-reader variability of tuberculin skin test may lead to misclassifications and FPs, particularly in low-prevalence populations.	[3, 4, 8, 20, 21, 26, 29, 30, 34, 62]	[2, 9, 11, 16, 19, 31, 42, 47, 51, 57]
3.2.8	Consider whether test performance is based on a combination of tests (and on a combination of areas under the Receiver Operating Characteristic (ROC) curves for each test).	Example from study: [4] For the triage analysis of cataract, a composite test (including three components) was compared with a reference standard of clinical decision 'do not discharge'.	[4, 5, 16, 25, 29, 30, 34, 42, 46, 52, 53, 56, 58-61]	[2, 9, 13, 49]
3.3	Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained.	Example from study: [51] For the base-case analysis, it was assumed that 3% of tests required to be repeated due to processing failure (of the sample) or an indeterminate result.	[3, 10, 29, 30, 51, 60]	
3.4	Consider the occurrence of test failures or indeterminate/not assessable results.	Example from study: [51] For the base-case analysis, it was assumed that 3% of tests required to be repeated due to processing failure (of the sample) or an indeterminate result.	[5, 9, 15, 24, 26, 28, 30, 42, 44-47, 51, 52, 59, 61, 62]	[4, 17, 22, 35, 43, 49, 57]
3.5	Consider costs of retesting (after obtaining insufficient/unusable sample, or after test failure or indeterminate/not assessable result).	Example from study: [51] For the base-case analysis, it was assumed that 3% of tests required to be repeated due to processing failure (of the sample) or an indeterminate result.	[9, 26, 28, 30, 44, 51]	[15, 43]
3.6	Consider complications, risks or other negative/positive aspects directly related to obtaining the sample and/or performing the diagnostic test (either in the intervention or in the control strategy).	Example from study: [65] Digital subtraction angiography involves a catheterization, and the most common complications involved a groin hematoma, and (possible) thromboembolic complications (both 2%).	[4, 5, 8, 9, 26, 28, 30, 39, 41-44, 46, 47, 57, 59, 64]	[6, 10, 17, 25, 31, 32, 35, 40, 50, 52, 60]
3.7	Consider the time taken to: perform the test (including waiting time), until the test result is available, or until a management decision or treatment is initiated based on this test result (either in the intervention or in the control strategy).	Example from study: [1] The duration of a point-of-care troponin test is estimated at 10 minutes.	[11, 13, 19, 20, 22, 38, 41, 42, 45, 48, 51, 53, 54, 57, 61, 62]	[4, 18, 24, 25, 27, 30, 36, 44, 46, 50, 55]

3.8	Consider the impact of additional knowledge gained by performing the diagnostic test (i.e., for a genetic test), or the occurrence and impact of chance findings.	Example from study: [28] The study includes the indirect costs associated with incidental findings during calcium scoring and cardiac computed tomography angiography.	[28, 30]	[4, 11, 13, 14, 16, 21, 39, 43, 57, 62, 64]
3.8.1	The impact of chance findings on performing additional tests is addressed.	Example from study: [28] Incidental cases (as mentioned in 3.8) would require non-contrast chest computed tomography a few months later at an additional costs of €80.00.	[28, 30]	[40]
PATIENT MANAGEMENT DECISIONS				
Impact of a test on the diagnosis and/or patient management strategy (based on this diagnosis)			Item included in study	Item mentioned (but not included)
4.1	Clearly specify the impact of the test in selecting the patient management strategy.	Example from study: [4] The diagnostic procedures either ends when a test is negative, or invasive coronary angiography is performed when the last test is positive or inconclusive.	[2-22, 24-26, 28, 30-36, 38, 39, 41-53, 56-59, 61-64, 67]	[27, 29, 37, 40, 55, 60]
4.2	Consider whether other aspects besides test results themselves are part of the decision algorithm (and included in the evaluation). These may involve a shared decision-making process of the physician with patients/relatives, or aspects including coverage or physician adherence to treatment guidelines).	Example from study: [65] Among patients with non-traumatic intracerebral hemorrhage, not all patients with (initial) negative results underwent digital subtraction angiography, owing to reluctance of patients and their treating clinicians because of a (small) risk of complications.	[2, 4, 5, 9, 11, 17, 20, 30, 32-34, 36-38, 43-45, 48, 51, 54, 56, 57, 62, 64]	[3, 6, 7, 10, 13, 21, 24, 31, 42, 52, 53, 58, 61, 63]
4.3	Consider whether the impact of the test result on resulting/selected diagnosis or management strategy varies across subgroups (this difference should not only be caused by differences in diagnostic performance of the test, and does not need to include the impact on costs and/or health outcomes within this subgroup)	Example from study: [68] When an aneurysm is diagnosed, the strategy of whether or not to (surgically) treat a patient, may be influenced by the patient's age.	[2, 4, 12, 13, 15, 21, 24, 30, 31, 33-35, 39, 40, 42, 48, 50, 52, 54, 58]	[6, 19, 37, 41, 60]
4.4	Consider the consistency of test results over time (for example: genetic mutations may be affected by treatment prescribed after the initial diagnosis).	Example from study: [54] Among patients with tuberculosis, smear-negative cases may progress to smear-positive, and all individuals with active disease may spontaneously self-cure.	[9, 53, 54]	

4.5	Consider the impact of performing the test and providing and interpreting the result on the time spend/capacity of the health care professional(s) or the patient.	Example from study: [50] The modified transesophageal echocardiography imaging procedure takes, on average, only 4.5 minutes to perform.	[15, 19, 20, 24, 31, 41, 42, 51, 52]	[29, 38, 50]
IMPACT ON HEALTH OUTCOMES AND COSTS			Item included in study	Item mentioned (but not included)
Impact of the patient management strategy on diseased and non-diseased individuals, in terms of health outcomes and costs				
5	Evaluate the direct impact of the chosen patient management strategy on the number of (in)correctly diagnosed individuals, health outcomes, and/or costs.	Example from study: [69] The overall impact of an extensive laboratory work-up in anemia patients on the (correct) diagnosis, on treatment decisions and on accompanying costs was quantified.	[2-64]	
5.1	Consider the direct impact of the chosen patient management strategy, on health outcomes and/or costs. This concerns the entire period in which patient management may affect a patient's health and/or costs, and does not only involve the testing strategy itself.	Example from study: [1] As acute coronary syndrome may impact the patient's well-being over his or her entire lifetime, the impact on quality of life, as well as the (accompanying) impact on societal costs, was estimated over a patient's lifetime.	[4-7, 11-15, 20, 24, 25, 35, 41-44, 46, 48, 49, 58]	[2, 8-10, 16-19, 21-23, 26-34, 36-39, 45, 47, 50-55, 57, 60-64]
5.2	Consider whether the direct impact of the chosen patient management strategy on health outcomes and/or costs varies across subgroups. (This does not include only varying the incidence of a certain condition in a sensitivity analysis. The subgroups should be clearly defined, and preferably be identifiable based on patient characteristics).	Example from study: [69] Among patients presenting with anemia in primary care, the results in terms of costs and health outcomes (i.e. correct underlying causes) were presented separately for the different underlying causes.	[2, 8, 10, 12-15, 21-24, 28, 30, 33, 38-40, 42, 44, 50, 54, 64]	[6, 35, 48, 53, 58]
5.3	Consider patient's adherence to treatment (which includes aspects that may indicate (partial) non-adherence, for example, following only some of the treatment recommendations, as well as aspects that affect the degree of administration of treatment).	Example from study: [54] Individuals on TB treatment may successfully complete treatment, fail, default (become lost to follow-up) or die.	[8, 11, 23, 30, 33, 38, 42, 51, 53, 54, 56, 58, 62, 63]	[61]

5.4	Consider the occurrence (and consequences) of treatment-related adverse events.	Example from study: [54] Individuals who fail or default from treatment may acquire resistance to the drugs they have received.	[3, 5, 8, 12, 15, 26, 30, 33, 41, 42, 44, 49-52, 54, 56]	[6, 9, 21, 31, 34, 35, 58, 63]
5.5	Describe the probability, or time it takes to observe that the patient management strategy proves to be effective over time, or that the patient cures spontaneously (regardless whether the patient received a correct or an incorrect diagnosis).	Example from study: [69] Although a patient may be diagnosed with the incorrect underlying cause of anemia, this same patient may (by coincidence) receive a management or treatment strategy that is considered effective.	[2, 9-11, 20-24, 26, 31-34, 42, 44, 45, 48, 52-54, 56, 58, 62-64]	[61]
5.6	Describe the probability of, or time it takes to repeat or extend the diagnostic work-up when the patient management strategy proves to be ineffective, either directly or over time (regardless whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which the patient receives no treatment, or unnecessary treatment.	Example from study: [69] Although a patient may be diagnosed with the correct underlying cause of anemia, the management or treatment strategy (as chosen by the general practitioner) may be considered ineffective.	[2, 4, 6-12, 16-18, 21-24, 26, 28, 30-36, 40, 42-46, 48, 52-54, 58, 62-64]	[57, 61]
5.7	Describe the impact of ineffective or unnecessary treatment or management on health outcomes and/or costs (including both side effects and costs, and regardless whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which incorrectly no treatment is provided, or in which the treatment is delayed.	Example from study: [69] The impact of ineffective treatment was quantified (in terms of repeat consultations and additional healthcare costs).	[2-5, 7, 8, 11, 13, 16, 17, 20-22, 24-26, 28, 30-36, 38, 39, 41-43, 45-48, 50-54, 56, 59, 61, 62, 64]	[6, 9, 12, 14, 15, 18, 19, 23, 27, 37, 40, 44, 57, 60, 63]
5.7.1	Consider the impact of delay in treatment initiation on health outcomes and/or costs	Example from study: [1] In patients with acute coronary syndrome, a delayed referral to the emergency care (and consequently a delay in treatment initiation), is associated with worse health outcomes.	[4, 5, 7, 11, 13, 21, 24-26, 30, 33, 35, 38, 39, 42, 45, 46, 48, 52-54, 59, 61, 64]	[6, 8, 9, 12, 14, 15, 17-19, 23, 27, 34, 37, 40, 44, 51, 57, 60, 62]

SOCIETAL IMPACT Wider (societal) impact of the chosen diagnostics and management strategy			Item included in study	Item mentioned (but not included)
6.1	Consider the psychological impact of diagnostic outcome and management strategy on patients, value of knowing (in terms of reassurance or anxiety), patient preferences regarding undergoing diagnostic tests, and the (accompanying) impact on caregivers, or relatives, etc.	Example from study: [14] Among patients with hereditary colorectal cancer and polyposis syndromes, it was (in the base-case analysis) assumed that first-, second-, and third-degree relatives would be contacted.	[4, 14, 42, 48, 56, 58]	[2, 6, 8, 9, 15, 21, 25-27, 30-32, 34, 35, 43, 57]
6.1.1	Consider the impact of test outcomes on relatives themselves, regarding the value of knowing (spillover knowledge) and regarding subsequent testing and/or treatment in this group (in case of heritable genetic conditions, or contagious diseases).	Example from study: [14] If the relative (of the patient with hereditary colorectal cancer and polyposis syndromes) was found to carry the pathogenic variant, he or she was presumably offered intensive colorectal surveillance with colonoscopy at recommended intervals.	[14, 56]	[27]
6.2	Consider the additional impact of diagnostic outcome and management strategy on the health system or healthcare professionals.	Example from study: [69] If an anemia patient receives a correct treatment immediately, this lowers the number of follow-up appointments at the general practitioner.	[37, 52, 54]	[4, 5, 9, 15, 17, 19-21, 25, 27, 29, 30, 32, 35, 40, 42, 45, 48, 51, 60, 61]
6.3	Consider the additional impact of diagnostic outcome and management strategy for society.	Example from study: [70] Use of the procalcitonin test in intensive care unit patients with sepsis limits the duration of antibiotic therapy, which in turn may reduce antibiotic resistance.	[16, 29, 54]	[2, 3, 6, 8, 9, 11, 19, 20, 24, 33, 34, 36, 37, 42, 53, 55, 61-63]

REFERENCES

1. Kip, M.M.A., et al., The cost-utility of point-of-care troponin testing to diagnose acute coronary syndrome in primary care. *BMC Cardiovasc Disord*, 2017. 17(1): p. 213.
2. Lubell, Y., et al., Modelling the Impact and Cost-Effectiveness of Biomarker Tests as Compared with Pathogen-Specific Diagnostics in the Management of Undifferentiated Fever in Remote Tropical Settings. *PLoS One*, 2016. 11(3): p. e0152420.
3. Peixoto, H.M., et al., Cost-effectiveness analysis of rapid diagnostic tests for G6PD deficiency in patients with *Plasmodium vivax* malaria in the Brazilian Amazon. *Malar J*, 2016. 15: p. 82.
4. Azuara-Blanco, A., et al., Automated imaging technologies for the diagnosis of glaucoma: a comparative diagnostic study for the evaluation of the diagnostic accuracy, performance as triage tests and cost-effectiveness (GATE study). *Health Technol Assess*, 2016. 20(8): p. 1-168.
5. Gregory, S., et al., Cost-Effectiveness of Integrating a Clinical Decision Rule and Staged Imaging Protocol for Diagnosis of Appendicitis. *Value Health*, 2016. 19(1): p. 28-35.
6. Cerantola, Y., et al., Cost-effectiveness of multiparametric magnetic resonance imaging and targeted biopsy in diagnosing prostate cancer. *Urol Oncol*, 2016. 34(3): p. 119 e1-9.
7. Karl, J.W., E. Swart, and R.J. Strauch, Diagnosis of Occult Scaphoid Fractures: A Cost-Effectiveness analysis. *J Bone Joint Surg Am*, 2015. 97(22): p. 1860-8.
8. Lubitz, C.C., et al., Cost-Effectiveness of Screening for Primary Aldosteronism and Subtype Diagnosis in the Resistant Hypertensive Patients. *Circ Cardiovasc Qual Outcomes*, 2015. 8(6): p. 621-30.
9. Nicholson, A., et al., The clinical effectiveness and cost-effectiveness of the PROGENSA(R) prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation. *Health Technol Assess*, 2015. 19(87): p. i-xxxi, 1-191.
10. Sato, M., et al., Cost-Effectiveness Analysis of the Diagnosis and Treatment of Primary Aldosteronism in Japan. *Horm Metab Res*, 2015. 47(11): p. 826-32.
11. Suen, S.C., E. Bendavid, and J.D. Goldhaber-Fiebert, Cost-effectiveness of improvements in diagnosis and treatment accessibility for tuberculosis control in India. *Int J Tuberc Lung Dis*, 2015. 19(9): p. 1115-24, i-xv.
12. Whiting, P., et al., Viscoelastic point-of-care testing to assist with the diagnosis, management and monitoring of haemostasis: a systematic review and cost-effectiveness analysis. *Health Technol Assess*, 2015. 19(58): p. 1-228, v-vi.
13. Westwood, M., et al., High-sensitivity troponin assays for the early rule-out or diagnosis of acute myocardial infarction in people with acute chest pain: a systematic review and cost-effectiveness analysis. *Health Technol Assess*, 2015. 19(44): p. 1-234.
14. Gallego, C.J., et al., Next-Generation Sequencing Panels for the Diagnosis of Colorectal Cancer and Polyposis Syndromes: A Cost-Effectiveness Analysis. *J Clin Oncol*, 2015. 33(18): p. 2084-91.
15. Morris, S., et al., Cost-effectiveness of diagnostic laparoscopy for assessing resectability in pancreatic and periampullary cancer. *BMC Gastroenterol*, 2015. 15: p. 44.
16. Yin, Z.G., J.B. Zhang, and K.T. Gong, Cost-Effectiveness of Diagnostic Strategies for Suspected Scaphoid Fractures. *J Orthop Trauma*, 2015. 29(8): p. e245-52.
17. Pershad, J., et al., Cost-effectiveness of diagnostic approaches to suspected appendicitis in children. *J Am Coll Surg*, 2015. 220(4): p. 738-46.
18. You, J.H., et al., Cost-effectiveness analysis of the Xpert MTB/RIF assay for rapid diagnosis of suspected tuberculosis in an intermediate burden area. *J Infect*, 2015. 70(4): p. 409-14.
19. Schmid, K.B., et al., Smear plus Detect-TB for a sensitive diagnosis of pulmonary tuberculosis: a cost-effectiveness analysis in an incarcerated population. *BMC Infect Dis*, 2014. 14: p. 678.
20. Larson, B.A., et al., Finding a needle in the haystack: the costs and cost-effectiveness of syphilis diagnosis and treatment during pregnancy to prevent congenital syphilis in Kalamo District of Zambia. *PLoS One*, 2014. 9(12): p. e113868.
21. Mather, R.C., 3rd, et al., Cost-effectiveness analysis of the diagnosis of meniscus tears. *Am J Sports Med*, 2015. 43(1): p. 128-37.
22. Cheng, H.M., et al., Cost-effectiveness of non-invasive central blood pressure monitoring in the diagnosis of hypertension. *Am J Hypertens*, 2015. 28(5): p. 604-14.
23. Trakada, G., et al., A health-economic analysis of diagnosis and treatment of obstructive sleep apnea with continuous positive airway pressure in relation to cardiovascular disease. The Greek experience. *Sleep Breath*, 2015. 19(2): p. 467-72.
24. Shah, M., et al., Cost-effectiveness of novel algorithms for rapid diagnosis of tuberculosis in HIV-infected individuals in Uganda. *AIDS*, 2013. 27(18): p. 2883-92.
25. van Karnebeek, C.D., et al., Health economic evaluation of plasma oxysterol screening in the diagnosis of Niemann-Pick Type C disease among intellectually disabled using discrete event simulation. *Mol Genet Metab*, 2015. 114(2): p. 226-32.

26. Deppen, S.A., et al., Cost-effectiveness of initial diagnostic strategies for pulmonary nodules presenting to thoracic surgeons. *Ann Thorac Surg*, 2014. 98(4): p. 1214-22.
27. Gould, M., et al., Cost-effectiveness and diagnostic effectiveness analyses of multiple algorithms for the diagnosis of Lynch syndrome. *Dig Dis Sci*, 2014. 59(12): p. 2913-26.
28. Ferreira, A.M., et al., Cost-effectiveness of different diagnostic strategies in suspected stable coronary artery disease in Portugal. *Arq Bras Cardiol*, 2014. 102(4): p. 391-402.
29. Assefa, L.M., et al., Diagnostic accuracy and cost-effectiveness of alternative methods for detection of soil-transmitted helminths in a post-treatment setting in western Kenya. *PLoS Negl Trop Dis*, 2014. 8(5): p. e2843.
30. Cooper, N.A., et al., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis. *Health Technol Assess*, 2014. 18(24): p. 1-201, v-vi.
31. Phelps, C.E., et al., Cost effectiveness of a gene expression score and myocardial perfusion imaging for diagnosis of coronary artery disease. *Am Heart J*, 2014. 167(5): p. 697-706 e2.
32. Bentley, T.G., et al., Cost effectiveness of a 92-gene assay for the diagnosis of metastatic cancer. *J Med Econ*, 2014. 17(8): p. 527-37.
33. Collins, I.J., et al., Cost-effectiveness of early infant HIV diagnosis of HIV-exposed infants and immediate antiretroviral therapy in HIV-infected children under 24 months in Thailand. *PLoS One*, 2014. 9(3): p. e91004.
34. Bermingham, S.L., The appropriate use of neuroimaging in the diagnostic work-up of dementia: an economic literature review and cost-effectiveness analysis. *Ont Health Technol Assess Ser*, 2014. 14(2): p. 1-67.
35. de Rooij, M., et al., Cost-effectiveness of magnetic resonance (MR) imaging and MR-guided targeted biopsy versus systematic transrectal ultrasound-guided biopsy in diagnosing prostate cancer: a modelling study from a health care perspective. *Eur Urol*, 2014. 66(3): p. 430-6.
36. van't Hoog, A.H., et al., Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: a decision analysis. *PLoS One*, 2013. 8(12): p. e82786.
37. Guerra, R.L., et al., Cost-effectiveness of routine diagnostic evaluation of pulmonary tuberculosis in a primary care unit in Brazil. *Int J Tuberc Lung Dis*, 2013. 17(10): p. 1336-40.
38. Choi, H.W., et al., Cost-effectiveness of Xpert(R) MTB/RIF for diagnosing pulmonary tuberculosis in the United States. *Int J Tuberc Lung Dis*, 2013. 17(10): p. 1328-35.
39. Holmes, M.W., et al., The cost-effectiveness of diagnostic management strategies for children with minor head injury. *Arch Dis Child*, 2013. 98(12): p. 939-44.
40. Yang, Z., N. Clark, and K.T. Park, Effectiveness and cost-effectiveness of measuring fecal calprotectin in diagnosis of inflammatory bowel disease in adults and children. *Clin Gastroenterol Hepatol*, 2014. 12(2): p. 253-62 e2.
41. Zanocco, K., et al., Cost effectiveness of intra-operative pathology examination during diagnostic hemithyroidectomy for unilateral follicular thyroid neoplasms. *J Am Coll Surg*, 2013. 217(4): p. 702-10.
42. Mowatt, G., et al., The diagnostic accuracy and cost-effectiveness of magnetic resonance spectroscopy and enhanced magnetic resonance imaging techniques in aiding the localisation of prostate abnormalities for biopsy: a systematic review and economic evaluation. *Health Technol Assess*, 2013. 17(20): p. vii-xix, 1-281.
43. Ohno, M. and A. Caughey, The role of noninvasive prenatal testing as a diagnostic versus a screening tool--a cost-effectiveness analysis. *Prenat Diagn*, 2013. 33(7): p. 630-5.
44. Beynon, R., et al., The diagnostic utility and cost-effectiveness of selective nerve root blocks in patients considered for lumbar decompression surgery: a systematic review and economic model. *Health Technol Assess*, 2013. 17(19): p. 1-88, v-vi.
45. Kansal, A.R., et al., Cost-effectiveness of a FISH assay for the diagnosis of melanoma in the USA. *Expert Rev Pharmacoecon Outcomes Res*, 2013. 13(3): p. 371-80.
46. Walker, S., et al., Cost-effectiveness of cardiovascular magnetic resonance in the diagnosis of coronary heart disease: an economic evaluation using data from the CE-MARC study. *Heart*, 2013. 99(12): p. 873-81.
47. Boldt, J., et al., Cost-effectiveness of cardiovascular magnetic resonance and single-photon emission computed tomography for diagnosis of coronary artery disease in Germany. *J Cardiovasc Magn Reson*, 2013. 15: p. 30.
48. Wilson, E.C., et al., The cost-effectiveness of a novel SIAscopic diagnostic aid for the management of pigmented skin lesions in primary care: a decision-analytic model. *Value Health*, 2013. 16(2): p. 356-66.
49. Widjaja, E., B. Li, and L.S. Medina, Diagnostic evaluation in patients with intractable epilepsy and normal findings on MRI: a decision analysis and cost-effectiveness study. *AJNR Am J Neuroradiol*, 2013. 34(5): p. 1004-9, S1-2.
50. Koffijberg, H., B. van Zaane, and K.G. Moons, From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol*, 2013. 13: p. 12.
51. Shah, M., et al., QuantiFERON-TB gold in-tube implementation for latent tuberculosis diagnosis in a public health clinic: a cost-effective-

- ness analysis. *BMC Infect Dis*, 2012. 12: p. 360.
52. Najafzadeh, M., et al., Cost-effectiveness of using a molecular diagnostic test to improve preoperative diagnosis of thyroid cancer. *Value Health*, 2012. 15(8): p. 1005-13.
 53. Winetsky, D.E., et al., Screening and rapid molecular diagnosis of tuberculosis in prisons in Russia and Eastern Europe: a cost-effectiveness analysis. *PLoS Med*, 2012. 9(11): p. e1001348.
 54. Menzies, N.A., et al., Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. *PLoS Med*, 2012. 9(11): p. e1001347.
 55. de Oliveira, M.R., et al., Cost-effectiveness of diagnostic for malaria in Extra-Amazon Region, Brazil. *Malar J*, 2012. 11: p. 390.
 56. del Campo, M.T., et al., Cost-effectiveness of different screening strategies (single or dual) for the diagnosis of tuberculosis infection in healthcare workers. *Infect Control Hosp Epidemiol*, 2012. 33(12): p. 1226-34.
 57. Ward, M.J., et al., Cost-effectiveness of diagnostic strategies for evaluation of suspected subarachnoid hemorrhage in the emergency department. *Acad Emerg Med*, 2012. 19(10): p. 1134-44.
 58. Biasutti, M., et al., Cost-effectiveness of magnetic resonance imaging with a new contrast agent for the early diagnosis of Alzheimer's disease. *PLoS One*, 2012. 7(4): p. e35559.
 59. Breijer, M.C., et al., Diagnostic strategies for endometrial cancer in women with postmenopausal bleeding: cost-effectiveness of individualized strategies. *Eur J Obstet Gynecol Reprod Biol*, 2012. 163(1): p. 91-6.
 60. Heinzel, A., et al., Cost-effectiveness analysis of FET PET-guided target selection for the diagnosis of gliomas. *Eur J Nucl Med Mol Imaging*, 2012. 39(7): p. 1089-96.
 61. Abimbola, T.O., et al., Cost-effectiveness of tuberculosis diagnostic strategies to reduce early mortality among persons with advanced HIV infection initiating antiretroviral therapy. *J Acquir Immune Defic Syndr*, 2012. 60(1): p. e1-7.
 62. Hughes, R., et al., The cost effectiveness of Nucleic Acid Amplification Techniques for the diagnosis of tuberculosis. *Respir Med*, 2012. 106(2): p. 300-7.
 63. Bajaj, J.S., et al., Diagnosis and treatment of minimal hepatic encephalopathy to prevent motor vehicle accidents: a cost-effectiveness analysis. *Hepatology*, 2012. 55(4): p. 1164-71.
 64. Holmes, M.W., et al., The cost-effectiveness of diagnostic management strategies for adults with minor head injury. *Injury*, 2012. 43(9): p. 1423-31.
 65. van Asch, C.J., et al., Diagnostic yield and accuracy of CT angiography, MR angiography, and digital subtraction angiography for detection of macrovascular causes of intracerebral haemorrhage: prospective, multicentre cohort study. *BMJ*, 2015. 351: p. h5762.
 66. Schop, A., et al., The effectiveness of a routine versus an extensive laboratory analysis in the diagnosis of anemia in general practice. Manuscript submitted, 2017.
 67. Dowdy, D.W., et al., Impact and cost-effectiveness of current and future tuberculosis diagnostics: the contribution of modelling. *Int J Tuberc Lung Dis*, 2014. 18(9): p. 1012-8.
 68. Schermerhorn, M.L., et al., Cost-effectiveness of surgery for small abdominal aortic aneurysms on the basis of data from the United Kingdom small aneurysm trial. *J Vasc Surg*, 2000. 31(2): p. 217-26.
 69. Kip, M.M.A., et al., ANNALS EXPRESS: Assessing the cost-effectiveness of a routine versus an extensive laboratory work-up in the diagnosis of anaemia in Dutch general practice. *Ann Clin Biochem*, 2017: p. 4563217748984
 70. Kip, M.M., et al., A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis. *J Med Econ*, 2015. 18(11): p. 944-53.

13

General discussion

GENERAL DISCUSSION

Despite the rapid increase in the number of available diagnostic tests, many of these tests are not actually used in current clinical practice. As the demand for healthcare is increasing, which is attributable to the aging population, the increasing incidence of chronic diseases, the rise of personalized medicine as well as technological improvements, the number of available tests is however expected to rise even further [1, 2]. Besides an increase in healthcare costs, this also raises concerns about test overuse, and about the potential negative impact of over testing on patients' health [3, 4]. Therefore, a thorough evaluation of the expected impact of new diagnostic tests, in terms of costs and health outcomes, is essential to improve the efficiency of diagnostic test implementation and use, while simultaneously providing affordable and good quality healthcare.

The aim of this thesis was three-fold: (1) investigate the consequences of existing as well as new diagnostic tests or diagnostic testing strategies in terms of health outcomes and costs, based on case studies in three major disease areas, (2) investigate which aspects affect the implementation and use of (new) diagnostic tests, with a focus on laboratory biomarkers, and (3) and to provide recommendations and guidance on investigating the health economic impact of such diagnostic tests.

Investigating the impact of new healthcare technologies on costs and health outcomes, compared with the standard-of-care, is referred to as Health Technology Assessment (HTA) [5, 6]. HTA can inform decisions regarding the implementation and use of (new) diagnostic tests, or testing strategies, in clinical practice. Though, owing to the rapid development of diagnostic tests, and because of the (many) different roles diagnostic tests can potentially fulfill in a diagnostic pathway [7], assessment of new medical tests should preferably be performed in early stages of product development, which is referred to as 'early HTA' [8]. More specifically, early HTA has been defined as "all methods used to inform industry and other stakeholders about the potential value of new medical products in development, including methods to quantify and manage uncertainty" [8]. Following this definition, the insights obtained through such an early HTA can be used to guide decisions during diagnostic test development, define minimum performance thresholds, support pricing and reimbursement strategies, and (thereby) improve the likelihood of successful test implementation and use [8-12].

This chapter will evaluate the main findings of this thesis into the perspective of current knowledge, and further elaborate on conceptual and methodological issues related to (early) HTA of diagnostic tests. In addition, the implications for practice will be described and recommendations for further research will be formulated for each of the three main aims of this thesis. Finally, the main conclusions will be drawn.

Implications for practice

Challenges in quantifying the impact of diagnostic testing strategies

When evaluating the consequences of diagnostic tests or diagnostic testing strategies in terms of health outcomes and costs, such an evaluation should ideally assess the utility of diagnostic testing. This utility involves the degree to which test use is associated with a change in health outcomes, such as preventing death, restoring or maintaining health, or resolving uncertainty [13]. However, in practice, identifying a test's clinical utility may be complicated because the diagnostic accuracy of a test (i.e. sensitivity and specificity), cannot be translated in a 1:1 ratio to the management decisions made by the clinician. To illustrate this, chapter 3 has shown that a non-elevated point-of-care (POC) troponin level does not guarantee that the patient is not (unnecessarily) referred to the hospital. Also, published literature [14] and chapter 7 have shown that a procalcitonin (PCT) level below its recommended threshold for antibiotic discontinuation only leads to stopping antibiotic therapy (within 24 hours) in 44% of sepsis patients. In addition, results from chapter 10 indicated that when the diagnostic performance of a combination of cardiac markers (used in a coronary care unit) is assumed to increase, the accompanying increase in patients' discharge rates is much smaller. Similar examples have been described in literature [15-19]. In other words, the eventual management decision is (besides diagnostic test outcomes) affected by several factors, including the judgement of the treating clinician, patient's preferences, and costs of the chosen management or treatment decision. Consequently, the potential impact of a (single) diagnostic test is limited.

Although the abovementioned issues have indicated that evaluating the impact of diagnostic tests is complex, the results of this thesis have also shown that surveys and expert elicitations are likely of added value to estimate the impact of diagnostic tests on the resulting patient management decisions. It should however be acknowledged that this intended behavior (as expressed by clinicians), does not always seem to reflect actual behavior in clinical practice [20]. Thus, in order to assess the real-life utility of a diagnostic test, it is required to quantify the impact of diagnostic test results on actual treatment and management decisions, as well as the accompanying impact on health outcomes and/or patient's well-being.

Aspects that affect test implementation and use in clinical practice

As illustrated in chapters 3 and 4, the decision to implement and use a test in clinical practice is not only affected by its expected cost-effectiveness, but also depends on the type of test that is evaluated. A previous study indicated that 63% of Dutch general practitioners (GPs) would like a POC test for the diagnosis of acute cardiac disease [21]. This corresponds to the findings from a survey (chapter 3), indicating that the majority of GPs consider a POC troponin test to be of added value for this purpose. Furthermore, this survey showed that GPs' gut feeling plays a (very) important role in the decision to immediately refer a patient to the hospital. Despite these abovementioned preferences of GPs with regard to the use of a POC troponin test, the

results from this survey also indicated that only 5 out of 126 GPs actually use a POC troponin test within their general practice. In addition, the majority of these GPs also agreed that this test had to fulfil several requirements: the test should be performed with a finger prick blood sample (instead of a venous blood sample), the POC analyzer should be reimbursed, and test results should be available within ten minutes. As many of the currently available POC troponin tests do not fulfill these requirements, this may (partly) explain the mismatch between the desire of the GPs with regard to this test and its actual implementation in clinical practice. Therefore, it is crucial to get insights in barriers and facilitators of test use during early development stages. This may allow incorporating users' requirements during this development process, which likely enhances the likelihood of POC troponin implementation and use.

Furthermore, although chapter 4 indicated that the use of a POC troponin test is likely safe and cost-saving, the diagnostic accuracy of many of the currently available POC troponin tests may not yet allow to rule out acute coronary syndrome (ACS) with sufficient certainty. The remaining uncertainty is likely another reason why the implementation of this test is still very limited in clinical practice. The issue of diagnostic uncertainty may be (partly) overcome by the use of a clinical decision rule (CDR) [22-24], either or not combined with a POC cardiac marker (chapter 2). Such a CDR may facilitate decision making in chest pain patients, and is likely cost-effective. In turn, this CDR may increase the likelihood of POC test implementation.

However, the decision to implement a (new) diagnostic test in clinical practice is not solely based on its diagnostic performance and its expected impact on decision making. To illustrate this, consider the use of PCT-guidance in intensive care patients with sepsis. Although previous research has shown that the use of PCT-guidance is expected to reduce antibiotic exposure [14, 25-32] without compromising health outcomes [25, 26, 28-30, 32], and possibly even leading to lower mortality rates [14], the use of this test is still limited in clinical practice. This low uptake may be explained by the relatively high cost of the PCT test when compared to other routinely performed laboratory tests. However, this thesis indicated that, when considering costs during the initial hospitalization episode, the use of PCT-guidance is expected to be cost saving, as the higher initial costs of this test are offset by cost-savings further downstream the patient's management pathway. Also, when applying a willingness-to-pay threshold of €60,000/QALY, PCT-guidance can be considered cost-effective from a societal perspective within a one year time horizon. Despite this, those relatively high initial costs were found to hamper implementation and use of the PCT test. One explanation for this is because many hospital laboratories receive an annual budget for performing their laboratory analyses [33], which implies that the costs of the laboratory tests performed are directly paid from the laboratory's budget. Consequently, these hospital laboratories are likely reluctant or unable to offer such (relatively) expensive tests as the PCT test. However, hospital laboratories increasingly charge the costs of diagnostic tests to the requesting hospital department, and these laboratories, in general, are expected to be more willing to offer such relatively expensive tests.

(Early) HTA in diagnostic test evaluation – reporting of these assessments

As illustrated in this thesis, diagnostic test evaluations are complex, especially because these tests (in general) do not affect health outcomes directly, but instead have an indirect impact by affecting patient management decisions [34]. Because of this complexity, many economic evaluations of diagnostic tests were found not to consider all aspects which are relevant to the appropriate estimation of the cost-effectiveness of these tests, or the consideration of these aspects was not explicitly mentioned in these evaluations (chapter 12). Consequently, many of the published health economic diagnostic test evaluations are not sufficiently transparent in describing the methods used [35, 36]. Therefore, the goal of chapter 12 was to develop a checklist of items that ideally should be incorporated (or at least considered for inclusion) when conducting health economic evaluations of diagnostic tests. The use of this checklist enables authors to explicitly state which aspects they included in their evaluation, and which aspects they did not include (and for what reason). As reported previously, transparent and complete reporting of methods and findings facilitates interpretation and comparison of such evaluations [37]. Hopefully, the use of this checklist, in combination with existing general checklists or guidelines on economic evaluations of healthcare technologies, will contribute to the transparent and complete reporting of health economic evaluations of diagnostic tests and biomarkers. As suggested by Husereau et al, adequate reporting of such evaluations may facilitate their interpretation, provide a means for comparing these evaluations, and ultimately lead to better health decisions [37].

Recommendations for further research

Challenges in quantifying the impact of diagnostic testing strategies

An issue that complicates the evaluation of diagnostic tests in experimental settings, is the fact that in clinical practice, a diagnosis is usually based on the results from multiple tests [39]. More specifically, diagnostic tests are often combined with a clinical decision rule or with other diagnostic (laboratory) tests in parallel or sequence. However, when the diagnostic performance of such combinations of tests is estimated, independency between tests is commonly assumed, which is (thus) often invalid [38, 39]. Therefore, future research should aim to (further) improve methods available for evaluating the overall performance of a combination of tests or a sequence of tests.

Although RCTs are still considered the experimental design which provides the highest quality of evidence [40] for assessing the consequences of diagnostic tests, RCTs are often too expensive, too time consuming, or too complicated to perform [41-45]. As the number of diagnostic tests available is tremendous (and because this number is still increasing), the use of combinations of diagnostic tests will likely increase even faster. In turn, this will lead to numerous potential diagnostic testing strategies. In addition, the long time-frame required to assess long-term outcomes, or risk related to testing, make it practically infeasible to evaluate

the impact of all potential and relevant test combinations in clinical practice. Consequently, in situations where RCTs cannot be easily carried out, or where the dynamics of the health system are too difficult to account for in (other) experimental study settings, or where too many diagnostic testing strategies have to be compared, the use of (early) health economic models can be considered a valid alternative to quantify the impact of diagnostic testing [46]. Therefore, improvement of guidance regarding when an RCT (or another type of experimental study) is required or whether a health economic model can suffice, is recommended.

Aspects that affect test implementation and use in clinical practice

In this thesis, almost all studies that investigated the cost-effectiveness of a new laboratory test (or a new combination of tests), found only a minimal impact of the test on clinical effectiveness or quality-adjusted life years (QALYs). This limited impact of (new) diagnostic tests on health outcomes can be explained by the fact that diagnostic test results are not considered in isolation (as mentioned previously), but always along with other diagnostic strategies, patient's clinical signs, symptoms and medical history, and with the judgement of the treating clinician [14-19]. In turn, this further limits the potential benefit of experimental study designs of (new) diagnostic tests, and increases the potential added value of early HTA. Therefore, we postulate that, although the field of early HTA is emerging, its added value is still underestimated. This is (at least partly) explained by a lack of knowledge among stakeholders involved in the process of healthcare technology development and implementation. To illustrate this, a survey regarding the health economic knowledge among company directors or managers of medical device manufactures revealed that 60% rated themselves as having low or no knowledge of health economics prior to attending a workshop on this topic [47]. Consequently, effort should be spend on increasing knowledge about (the importance of) early HTA among all stakeholders involved in the process of technology development and implementation.

(Early) HTA in diagnostic test evaluation and the role of value-based healthcare

A recently published scoping review into the historical context of early HTA indicated that the reference to 'early' in this setting reflects the uncertainty in clinical evidence as well as the mechanisms to reduce or mitigate this uncertainty in evidence development [8]. This lack of clinical evidence is thus inevitable in the majority of early HTAs. As an unfortunate consequence, peer review of these kind of studies (when submitted to scientific journals) often indicates that the lack of patient outcome data is considered a major limitation. Moreover, this likely hampers (rapid) publication, or even prevents such studies from being published, indicating a risk of publication bias. Although this (again) illustrates that experimental studies continue to be preferred among reviewers, data obtained through such studies are commonly of limited value to evaluate the dynamics in the health system. More specifically, in experimental study settings it is (for example) not always possible to account for specific features of the health care system, including difficulties in defining the target population, or compliance to guidelines [8]. One such example in which investigating the real-life impact of a diagnostic test may be

practically unfeasible, is illustrated in chapter 4. In this chapter, the use of the POC troponin test in primary care depended on the patient's electrocardiogram (if performed), the severity of symptoms (as judged by the GP), as well as the duration of these symptoms. In such situations, early HTA offers a valuable alternative to experimental study designs to evaluate the dynamics of the health care system.

As reported previously, most recently published studies of (early) health economic models use the QALY as the single outcome measure. However, value in healthcare comprises various elements, ranging from elements related to patient health improvement, to wider elements of value, including other benefits to the patient (e.g. reduced out-of-pocket payments, return to work, reassurance), to benefits for caregivers, the health and societal care system, and to society as a whole [48]. These wider elements of value are however not captured in the QALY. Yet, several case studies in this thesis have shown that the decision on whether or not to implement and use a test in clinical practice is affected by many more factors besides the test's impact on cost-effectiveness and subsequent management or treatment. Therefore, when assessing the impact of (new) diagnostic tests, it is recommended to not only focus on its impact on costs and QALYs, but to also consider wider elements of value brought by the use of these tests. In turn, this will likely facilitate, on the one hand, the probability of successful test implementation and use, and on the other hand, prevent ongoing investments in diagnostic tests that are unlikely to be implemented in clinical practice.

General conclusion

From this thesis it can be concluded that HTA, and especially early HTA, is highly important to evaluate the expected impact of a diagnostic test (or combination of tests) on health outcomes and costs, prior to their implementation in clinical practice. Besides early-stage economic models, the use of multi-criteria decision analysis is valuable to estimate the factors that hamper or facilitate the implementation and use of newly developed diagnostic tests.

In order to fully quantify the value of diagnostic testing, it is important to also consider wider elements of value besides the impact on patients' health, as these will affect test implementation and use. In addition, the use of stakeholder elicitation is considered of added value to handle evidence gaps in early stage economic models. The value of (early) HTA however still tends to be underestimated.

The use of the developed checklist for reporting health economic evaluations of diagnostic tests and biomarkers, will hopefully contribute to the transparency and comparability of such evaluations. In turn, this may facilitate the interpretation of such studies for stakeholders involved in the decision-making process, and thereby ultimately lead to optimization of health decisions.

REFERENCES

- Wiener, J.M. and J. Tilly, Population ageing in the United States of America: implications for public programmes. *Int J Epidemiol*, 2002. 31(4): p. 776-81.
- Akhmetov, I. and R.V. Bubnov, Assessing value of innovative molecular diagnostic tests in the concept of predictive, preventive, and personalized medicine. *EPMA J*, 2015. 6: p. 19.
- U. S. Preventive Services Task Force, et al., Screening for Thyroid Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*, 2017. 317(18): p. 1882-1887.
- Treadwell, J. and M. McCartney, Overdiagnosis and overtreatment: generalists--it's time for a grassroots revolution. *Br J Gen Pract*, 2016. 66(644): p. 116-7.
- Devlin, N.J. and J. Sussex, Incorporating multiple criteria in HTA: methods and processes. 2011, Office of Health Economics: London. p. 60.
- Wahlster, P., et al., Balancing costs and benefits at different stages of medical innovation: a systematic review of Multi-criteria decision analysis (MCDA). *BMC Health Serv Res*, 2015. 15: p. 262.
- Bossuyt, P.M., et al., Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*, 2006. 332(7549): p. 1089-92.
- Ijzerman, M.J., et al., Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *Pharmacoeconomics*, 2017. 35(7): p. 727-740.
- Vallejo-Torres, L., et al., Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach. *Int J Technol Assess Health Care*, 2008. 24(4): p. 459-64.
- Ijzerman, M.J. and L.M. Steuten, Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Health Econ Health Policy*, 2011. 9(5): p. 331-47.
- Markiewicz, K., J.A. van Til, and I.J. MJ, Medical devices early assessment methods: systematic literature review. *Int J Technol Assess Health Care*, 2014. 30(2): p. 137-46.
- Ijzerman, M.J., Webinar: Early Health Technology Assessment: potential to inform pricing and reimbursement during R and D. 2016, ISPOR New Zealand.
- Bossuyt, P.M., et al., Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem*, 2012. 58(12): p. 1636-43.
- de Jong, E., et al., Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: a randomised, controlled, open-label trial. *Lancet Infect Dis*, 2016.
- Azuara-Blanco, A., et al., Automated imaging technologies for the diagnosis of glaucoma: a comparative diagnostic study for the evaluation of the diagnostic accuracy, performance as triage tests and cost-effectiveness (GATE study). *Health Technol Assess*, 2016. 20(8): p. 1-168.
- Karl, J.W., E. Swart, and R.J. Strauch, Diagnosis of Occult Scaphoid Fractures: A Cost-Effectiveness Analysis. *J Bone Joint Surg Am*, 2015. 97(22): p. 1860-8.
- Nicholson, A., et al., The clinical effectiveness and cost-effectiveness of the PROGENSA(R) prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation. *Health Technol Assess*, 2015. 19(87): p. i-xxxi, 1-191.
- Mather, R.C., 3rd, et al., Cost-effectiveness analysis of the diagnosis of meniscus tears. *Am J Sports Med*, 2015. 43(1): p. 128-37.
- Shah, M., et al., Cost-effectiveness of novel algorithms for rapid diagnosis of tuberculosis in HIV-infected individuals in Uganda. *AIDS*, 2013. 27(18): p. 2883-92.
- Lijmer, J.G., M. Leeflang, and P.M. Bossuyt, Proposals for a phased evaluation of medical tests. *Med Decis Making*, 2009. 29(5): p. E13-21.
- Howick, J., et al., Current and future use of point-of-care tests in primary care: an international survey in Australia, Belgium, The Netherlands, the UK and the USA. *BMJ Open*, 2014. 4(8): p. e005611.
- Backus, B.E., et al., A prospective validation of the HEART score for chest pain patients at the emergency department. *Int J Cardiol*, 2013. 168(3): p. 2153-8.
- Body, R., et al., The Manchester Acute Coronary Syndromes (MACS) decision rule for suspected cardiac chest pain: derivation and external validation. *Heart*, 2014. 100(18): p. 1462-8.
- Haasenritter, J., et al., Ruling out coronary heart disease in primary care: external validation of a clinical prediction rule. *Br J Gen Pract*, 2012. 62(599): p. e415-21.
- Nobre, V., et al., Use of procalcitonin to shorten antibiotic treatment duration in septic patients: a randomized trial. *Am J Respir Crit Care Med*, 2008. 177(5): p. 498-505.
- Bouadma, L., et al., Use of procalcitonin to reduce patients' exposure to antibiotics in intensive care units (PRORATA trial): a multicentre randomised controlled trial. *Lancet*, 2010. 375(9713): p. 463-74.
- Annane, D., et al., Procalcitonin levels to guide antibiotic therapy in adults with non-microbiologically proven apparent severe sepsis: a

- randomised controlled trial. *BMJ Open*, 2013. 3(2).
28. Deliberato, R.O., et al., Clinical and economic impact of procalcitonin to shorten antimicrobial therapy in septic patients with proven bacterial infection in an intensive care setting. *Diagn Microbiol Infect Dis*, 2013. 76(3): p. 266-71.
 29. Hochreiter, M., et al., Procalcitonin to guide duration of antibiotic therapy in intensive care patients: a randomized prospective controlled trial. *Crit Care*, 2009. 13(3): p. R83.
 30. Schroeder, S., et al., Procalcitonin (PCT)-guided algorithm reduces length of antibiotic treatment in surgical intensive care patients with severe sepsis: results of a prospective randomized study. *Langenbecks Arch Surg*, 2009. 394(2): p. 221-6.
 31. Shehabi, Y., et al., Procalcitonin algorithm in critically ill adults with undifferentiated infection or suspected sepsis. A randomized controlled trial. *Am J Respir Crit Care Med*, 2014. 190(10): p. 1102-10.
 32. Xu, X.L., et al., [Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment of sepsis patients]. *Zhonghua Yi Xue Za Zhi*, 2017. 97(5): p. 343-346.
 33. Medisch Contact. Win-win in het lab. 2005; Available from: <https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/win-win-in-het-lab.htm>.
 34. Bossuyt, P.M.M. and K. McCaffery, Additional Patient Outcomes and Pathways in Evaluations of Testing, in *Medical Tests-White Paper Series*. 2009: Rockville (MD).
 35. Oliveira, M.R.F., et al., Systematic Review of Health Economic Evaluations of Diagnostic Tests in Brazil: How accurate are the results? *Clinics (Sao Paulo)*, 2017. 72(8): p. 499-509.
 36. Oosterhoff, M., M.E. van der Maas, and L.M. Steuten, A Systematic Review of Health Economic Evaluations of Diagnostic Biomarkers. *Appl Health Econ Health Policy*, 2016. 14(1): p. 51-65.
 37. Husereau, D., et al., Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health*, 2013. 16(2): p. e1-5.
 38. Novielli, N., et al., How is evidence on test performance synthesized for economic decision models of diagnostic tests? A systematic appraisal of Health Technology Assessments in the UK since 1997. *Value Health*, 2010. 13(8): p. 952-7.
 39. Novielli, N., N.J. Cooper, and A.J. Sutton, Evaluating the cost-effectiveness of diagnostic tests in combination: is it important to allow for performance dependency? *Value Health*, 2013. 16(4): p. 536-41.
 40. Schunemann, H.J., et al., Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*, 2008. 336(7653): p. 1106-10.
 41. Bossuyt, P.M., J.G. Lijmer, and B.W. Mol, Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*, 2000. 356(9244): p. 1844-7.
 42. Moons, K.G., Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem*, 2010. 56(4): p. 537-41.
 43. Biesheuvel, C.J., D.E. Grobbee, and K.G. Moons, Distraction from randomization in diagnostic research. *Ann Epidemiol*, 2006. 16(7): p. 540-4.
 44. Hlatky, M.A., et al., Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*, 2009. 119(17): p. 2408-16.
 45. Ferrante di Ruffano, L., et al., A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J Clin Epidemiol*, 2012. 65(3): p. 282-7.
 46. Koffijberg, H., B. van Zaane, and K.G. Moons, From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol*, 2013. 13: p. 12.
 47. Craven, M.P., et al., Engaging with economic evaluation methods: insights from small and medium enterprises in the UK medical devices industry after training workshops. *Health Res Policy Syst*, 2012. 10: p. 29.
 48. Henshall, C., T. Schuller, and H.T.P. Forum, Health technology assessment, value-based decision making, and innovation. *Int J Technol Assess Health Care*, 2013. 29(4): p. 353-9.

The background is a dark blue gradient with numerous overlapping circles of varying sizes and opacities, creating a bokeh effect. The circles range from small, faint dots to large, bright, semi-transparent spheres. The overall aesthetic is modern and tech-oriented.

Summary

SUMMARY

The demand for healthcare has strongly increased in the last decades, due to the aging population, the increase in the number of people with chronic diseases, the rise of personalized medicine, and technological advances. Although this was accompanied by a strong increase in the number of available diagnostic tests, only a limited number of these tests are actually implemented and used in clinical practice. However, this rising number of diagnostic tests available raises concerns about potential overuse of tests that are already implemented in clinical practice, the accompanying (potential) negative health impact of diagnostic testing, for example caused by (unnecessary) distress about test results, as well as rising healthcare costs. Therefore, evaluating the expected health economic impact of new diagnostic tests or diagnostic strategies is highly important to enhance efficient test implementation and use, and to ensure the provision of affordable and good quality healthcare.

The aim of this thesis was therefore to investigate the health economic impact of (new) diagnostic tests or diagnostic testing strategies within three disease areas (i.e. acute coronary syndrome, sepsis and anaemia). In addition, this thesis investigated which aspects affect the implementation and use of (new) diagnostic tests (with a focus on laboratory biomarkers), and to provide recommendations and guidance on investigating the health economic impact of such diagnostic tests.

Investigating the potential added value and health economic impact of new healthcare technologies compared with the standard-of-care is referred to as Health Technology Assessment (HTA) and it can be used to support decisions regarding test implementation and use. However, owing to the rapid development of new diagnostic tests, assessment of these new tests should preferably be performed in early test development stages, which is referred to as 'early HTA'. More specifically, although regular HTA is mainly used to assess the safety, effectiveness and cost-effectiveness of a technology, early HTA considers aspects that are relevant to inform the test's development process, including implementation barriers, the predicted uptake of a new test, requirements of a new medical test, as well as patient needs.

With regard to the three main disease areas, focuses the first part of this thesis (**chapters 2, 3 and 4**) on early HTA of point-of-care (POC) tests to rule out acute coronary syndrome (ACS) in primary care:

Chapter 2 concerns an early HTA of a future clinical decision rule (CDR) incorporating a POC test, for diagnosing ACS in primary care as compared to current practice. The annual incidence and referral rates of chest pain patients, as well as the incidence of ACS, were estimated based on a literature review and on a Dutch and Belgian registration study. The potential impact of a future CDR on costs and effects (i.e., correct referral decisions) was estimated for several

scenarios. Results indicate that a reduction of at least 29% of unnecessary referrals (i.e. referrals of patients without ACS) is required to make the addition of a POC test as predictor in a CDR cost-saving.

Chapter 3 describes the results of a survey distributed among general practitioners (GPs), investigating the perceived added value with regard to a POC troponin test, as well as its effect on referral decisions and test requirements. A total of 126 GPs (15.1%) completed at least 75% of the questions. The results indicated that 67.1% of GPs believed that a POC troponin test has moderate to high added value, and that this test may decrease (immediate) hospital referral rates. Although GPs consider the test of potential added value to exclude ACS early on, actual test implementation will depend on test characteristics, including test duration, type of blood sample required, and reimbursement of the analyzer.

Chapter 4 estimates the expected cost-effectiveness of using a POC troponin test to rule out ACS in a hypothetical cohort of Dutch chest pain patients (>35 years) consulting a GP, using a patient-level health economic model. The model incorporated both symptom duration, selection of patients in whom the POC test would be performed, as well as test performance at different time points. The POC troponin strategy was found to decrease the referral rate among non-ACS patients from 38.46% to 31.85%. Despite a small increase in non-referral among ACS patients (i.e. from 0.22% to 0.27%), the overall health effect is negligible, while substantial cost savings (i.e. €77 per patient) can be achieved. Besides, current developments in POC troponin tests will likely further improve their diagnostic performance. Therefore, future prospective (randomized) studies into the safety and diagnostic accuracy of POC troponin are warranted to investigate whether those developments make these tests to a safe and cost-effective diagnostic tool for diagnosing ACS in primary care.

The second part of this thesis (**chapters 5, 6 and 7**) concerns the cost-effectiveness of a procalcitonin (PCT)-guided algorithm for antibiotic discontinuation in intensive care patients with sepsis:

Chapter 5 illustrates the cost-effectiveness of such a PCT-guided antibiotic treatment algorithm in intensive care patients with sepsis. The results of a systematic literature review regarding the effectiveness of such an algorithm served as input for a health economic model. Results indicate that use of a PCT algorithm is expected to reduce the duration of antibiotic prescription with 1.7 days, as well as decrease the duration of hospital stay with 3.4 days and a decrease in direct hospital costs with -€3,503 per patient. However, as both the duration of hospital stay and antibiotic treatment are relatively short in the Netherlands compared to other countries, this model should ideally be populated with country-specific data to make valid per country estimations.

Chapter 6 shows the response to a published randomized controlled trial (RCT), regarding the use of a PCT-guided treatment algorithm in intensive care patients with sepsis in the Netherlands. When using the results of this RCT to calculate the costs of this algorithm, results indicate that it could not be conclusively demonstrated that the use of a PCT algorithm is cost saving. As the duration of intensive care stay was found to have the largest impact on the overall costs, future studies should aim to estimate the impact of such a PCT algorithm on the duration of intensive care stay.

Chapter 7 concerns a trial-based cost-effectiveness analysis in which the resource use data from this recently published RCT (as described in chapter 6) were used as input parameters. The results indicate that the impact of this PCT algorithm on costs is minimal (i.e. €-65), although both the duration of antibiotic use and in-hospital mortality are lower compared with standard-of-care (i.e. -1.2 days and -7.9%, respectively). When applying a one-year time horizon, the high healthcare-related costs occurring in sepsis survivors are however expected to lead to a (non-significant) increase in costs of the PCT group compared to standard-of-care (i.e. +€2,704). This results in an incremental cost-effectiveness ratio (ICER) of ~60,000/quality-adjusted life year (QALY) gained, which cannot be considered cost-effective when applying a WTP threshold of €20,000/QALY. Long-term follow-up studies are however required to fully quantify the costs and health impact of such a PCT algorithm from the societal perspective.

The third part of this thesis (**chapters 8 and 9**) concerns an evaluation of the costs and effects of using an extensive laboratory analysis for diagnosis and treatment of anaemia in primary care:

Chapter 8 describes a study on the effectiveness of two different laboratory work-ups to detect the underlying cause of anaemia by GPs. In this study, effectiveness is defined as the percentage of anaemia patients in whom the correct underlying cause is diagnosed by the GP, when using an extensive versus a routine laboratory work-up. To investigate this, an online survey was distributed among 836 GPs. This survey consisted of six cases from an existing cohort of anaemia patients. In the first three cases, GPs could choose themselves which tests to perform (from a previously defined set of 14 relevant laboratory tests), i.e. the routine work-up. In the second three cases, the results of all 14 tests, i.e. the extensive work-up, were directly presented to the GPs. Results indicate that, compared to an expert panel, the use of this extensive work-up, improves GPs' ability to diagnose patients with the correct underlying cause of anaemia.

Subsequently, **chapter 9** describes the impact of both laboratory work-ups, in terms of the number of correctly diagnosed underlying causes, associated treatment decisions and accompanying costs. Besides a 6.4% improvement in the number of patients in whom the correct underlying cause of anaemia is diagnosed, costs are expected to increase slightly from

€842 per patient in the routine work-up to €845 per patient in the extensive work-up. However, as the additional costs of the extensive work-up are found to be minimal, while the chance of early diagnosis of the correct underlying cause increased, use of the extensive laboratory work-up was recommended.

The final part of this thesis (**chapters 10, 11 and 12**) discusses methodological issues related to performing an (early) HTA of a diagnostic test or testing strategy:

Chapter 10 illustrates how expert elicitation can be used to provide estimations for model input parameters in situations where scientific evidence is lacking. This is illustrated using a case study on the cost-effectiveness of the combined use of heart-type fatty acid binding protein (H-FABP), high-sensitive troponin, and copeptin, as compared to conventional serial high-sensitive troponin testing, to rapidly rule out non-ST elevation myocardial infarction in patients presenting with chest pain in the emergency department. The results indicate that the use of this triple biomarker in combination with high-sensitive troponin measurement after two and six hours, is expected to be the most cost-effective strategy compared to current practice. Furthermore, this case study showed that expert elicitation can be a valuable tool for early HTA to provide an initial estimate of the cost-effectiveness of new tests prior to their further development and (finally) their implementation in clinical practice.

Chapter 11 served to get insight in which features affect the implementation and use of diagnostic tests in clinical practice. This was illustrated using a case study on two POC tests in primary care (i.e. the POC glycated haemoglobin [HbA_{1c}] test), and the POC C-reactive protein [CRP] test). After a literature review and validation by 12 experts, a list of 20 criteria was obtained. Results of an analytical hierarchy process (AHP) session with 10 of these experts revealed that the test's clinical relevance, its technical performance, and risks (associated with making a treatment or management decision based on the test result) were considered most important with relative weights of 22.2%, 12.6% and 8.5%, respectively. The POC CRP test was preferred over its central laboratory equivalent, whereas the POC HbA_{1c} test was not. The insights obtained from this study may both steer test development and improve the implementation use of existing POC tests in general practices.

Chapter 12 describes the development of a comprehensive reporting checklist, which is intended to increase awareness of all aspects potentially relevant in investigating the impact of diagnostic tests and biomarkers in a model-based health economic evaluation. This checklist was designed based on a scoping review which was critically assessed by four independent experts who judged the clarity, redundancy and completeness of the checklist. This resulted in a list of 43 aspects that ideally should be considered for explicit inclusion in, or exclusion from a health economic evaluation. Thereby, this reporting checklist may contribute to improving the completeness and transparency of such model-based health economic evaluations.

The results of this thesis have indicated that it is highly important to evaluate the expected impact of a diagnostic test (or combination of tests) on health outcomes and costs, prior to their implementation in clinical practice. In this regard, HTA, and especially early HTA, represent a highly valuable set of techniques to evaluate this impact. Besides health economic models, other techniques that are part of early HTA, for example multi-criteria decision analysis, are valuable to obtain insight in factors that hamper or facilitate the implementation and use of new diagnostic tests. In order to fully quantify the value of diagnostic testing, wider elements of value besides the impact on costs and patients' health should be considered. Such wider elements of value for example include being reassured of not having a serious illness, or the knowledge of having a genetic disorder, which may affect family planning. Besides a test's cost-effectiveness, these broader elements of value will likely also affect its implementation and use.

Samenvatting

SAMENVATTING

De zorgvraag is in de laatste decennia sterk toegenomen. Oorzaken van deze toename zijn de vergrijzing, de groei in het aantal mensen met een chronische ziekte, de opkomst van personalized medicine, en technologische vooruitgang. Ondanks dat het aantal beschikbare diagnostische testen tegelijkertijd eveneens sterk is toegenomen, wordt slechts een klein deel van deze testen daadwerkelijk geïmplementeerd en gebruikt in de dagelijkse praktijk. Echter, deze toename in beschikbare diagnostische testen geeft ook aanleiding tot bezorgdheid omtrent mogelijk overmatig gebruik van reeds geïmplementeerde diagnostiek, de bijbehorende (potentiële) negatieve gezondheidssimpact door bijvoorbeeld (onnodige) ongerustheid over testuitslagen, alsmede een mogelijke toename in kosten van de gezondheidszorg. Het is daarom van groot belang om de verwachte gezondheidseconomische impact van nieuwe diagnostische testen of teststrategieën te evalueren om zowel efficiënte implementatie als efficiënt gebruik van deze testen te bevorderen, en om kwalitatief goede en betaalbare zorg te bewerkstelligen.

Het doel van dit proefschrift was daarom om de gezondheidseconomische impact van (nieuwe) diagnostische testen of diagnostische teststrategieën te onderzoeken binnen drie ziektegebieden (namelijk acuut coronair syndroom, sepsis, en anemie). Daarnaast is in dit proefschrift onderzoek gedaan naar welke aspecten de implementatie en het gebruik van (nieuwe) diagnostische testen beïnvloeden. De focus ligt hierbij op laboratoriumdiagnostiek. Eveneens was het doel van dit proefschrift om aanbevelingen te formuleren en sturing te bieden op het gebied van onderzoek naar de gezondheidseconomische impact van deze diagnostische testen.

Het onderzoeken van de potentiële toegevoegde waarde en gezondheidseconomische impact van nieuwe technologieën in de gezondheidszorg ten opzichte van de standaardzorg wordt 'Health Technology Assessment' genoemd. Deze evaluatie van medische technologieën kan bijdragen aan het maken van een weloverwogen beslissingen omtrent de implementatie en het gebruik ervan. Echter, omdat nieuwe diagnostische testen in een hoog tempo ontwikkeld worden, wordt evaluatie van nieuwe testen bij voorkeur al in vroegtijdige stadia van de technologieontwikkeling uitgevoerd. Terwijl (reguliere) evaluatie van gezondheidstechnologie vooral gebruikt wordt om de veiligheid, effectiviteit en kosteneffectiviteit van een technologie te bepalen, is vroegtijdige evaluatie vooral gericht op aspecten die relevant zijn voor het ondersteunen van beslissingen *tijdens het ontwikkelproces*. Dergelijke vroegtijdige evaluaties kunnen inzicht geven in welke aspecten implementatie van een test in de weg staan, in het verwachte gebruik van een nieuwe test, in eisen met betrekking tot een nieuwe test, alsmede in de wensen van de patiënt.

Met betrekking tot de drie ziektegebieden, is het eerste deel van dit proefschrift (**hoofdstuk 2, 3 en 4**) gericht op vroegtijdige technologie evaluatie van point-of-care (POC) testen ten behoeve van het uitsluiten van acuut coronair syndroom (ACS) in de eerste lijn:

Hoofdstuk 2 beschrijft een vroegtijdige technologie evaluatie van een toekomstige klinische beslisregel, inclusief POC test, ten behoeve van het diagnosticeren van ACS in de eerste lijn, vergeleken met de standaardzorg. De jaarlijkse incidentie van zowel ACS als van patiënten met pijn op de borst, en het bijbehorende percentage doorverwijzingen, zijn geschat op basis van een literatuur review en op basis van een registratiestudie in Nederland en België. De potentiële impact van een toekomstige klinische beslisregel op kosten en effecten (namelijk, het aantal juiste verwijsbeslissingen) is uitgerekend voor diverse scenario's. De resultaten tonen aan dat de afname in onnodige doorverwijzingen (verwijzingen van patiënten zonder ACS) tenminste 29% moet bedragen om de toevoeging van een POC test, als onderdeel van een klinische beslisregel, kosteneffectief te maken.

Hoofdstuk 3 beschrijft de resultaten van een vragenlijst verspreid onder huisartsen, betreffende de verwachte toegevoegde waarde van het gebruik van een POC troponine test, het effect op verwijsbeslissingen, en eisen die ze aan de test stellen. De enquête werd ingevuld door 126 huisartsen (minstens 75% van de vragen), wat overeenkomt met een respons van 15.1%. De resultaten tonen aan dat 67.1% van de huisartsen verwacht dat een POC troponine test van gemiddelde tot grote toegevoegde waarde is, en dat deze test kan leiden tot een afname in het aantal patiënten dat (direct) naar het ziekenhuis doorverwezen wordt. Ondanks dat huisartsen verwachten dat de test van toegevoegde waarde is om ACS vroegtijdig uit te kunnen sluiten, hangt de uiteindelijke implementatie af van testeigenschappen, waaronder de duur van de test, het type bloedmonster dat benodigd is, alsmede van vergoeding van het POC apparaat.

In **hoofdstuk 4** is de verwachte kosteneffectiviteit van het gebruik van een POC troponine test om ACS uit te sluiten in kaart gebracht voor een hypothetisch cohort van Nederlandse patiënten (>35 jaar) dat zich met pijn op de borst bij de huisarts meldt. Hiervoor is een gezondheidseconomisch model ontwikkeld waarin individuele patiënten zijn gesimuleerd. In dit model zijn zowel de duur van symptomen, de selectie van patiënten waarin de POC test wordt uitgevoerd, als de diagnostische performance van de test op verschillende tijdstippen meegenomen. De resultaten tonen aan dat de POC troponine strategie leidt tot een afname in doorverwijzingen van patiënten zonder ACS van 38.46% naar 31.85%. Ondanks een kleine toename in patiënten met ACS die niet doorverwezen worden (van 0.22% naar 0.27%), is het totale effect op gezondheidsuitkomsten verwaarloosbaar, terwijl er substantiële kostenbesparingen bereikt kunnen worden (€77 per patiënt). Het is daarnaast aannemelijk dat verdere ontwikkelingen van deze POC troponine testen leiden tot een verdere verbetering in hun diagnostische performance. Daarom is het aan te bevelen om

prospectieve (gerandomiseerde) studies uit te voeren naar de veiligheid en diagnostische performance van POC troponine testen, om te onderzoeken of deze testen een veilig en kosteneffectief diagnostisch hulpmiddel kunnen zijn ten behoeve van het uitsluiten van ACS in de huisartsenpraktijk.

Het tweede deel van dit proefschrift (**hoofdstuk 5, 6 en 7**) gaat over de kosteneffectiviteit van een algoritme voor het stoppen met antibiotica op geleide van de procalcitonine (PCT) test, in patiënten met sepsis op de intensive care (IC):

Hoofdstuk 5 illustreert de kosteneffectiviteit van het gebruik van de PCT test om de duur van de antibioticabehandeling te bepalen in IC patiënten met sepsis. De resultaten van een systematische literatuur review over de effectiviteit van een dergelijk algoritme zijn gebruikt als input voor een gezondheidseconomisch model. De resultaten laten zien dat het gebruik van dit PCT algoritme naar verwachting leidt tot een afname in de duur van de antibioticabehandeling met 1.7 dagen, een afname in de duur van de ziekenhuisopname met 3.4 dagen, en een afname in directe ziekenhuiskosten met €3,503 per patiënt. Echter, aangezien zowel de duur van ziekenhuisverblijf als van antibioticabehandelingen relatief kort is in Nederland vergeleken met andere landen, is het aan te bevelen om het model te heranalyseren met Nederlandse data.

Hoofdstuk 6 toont de correspondentie naar aanleiding van publicatie van de resultaten van een gerandomiseerde vergelijkende studie, betreffende het gebruik van een behandelalgoritme op geleide van de PCT test, in patiënten met sepsis op de IC in Nederland. In dit hoofdstuk zijn de resultaten van deze gerandomiseerde studie gebruikt om de kosten van een dergelijk PCT algoritme in kaart te brengen. De resultaten tonen aan dat het onzeker is of het gebruik van een dergelijk algoritme ook kosteneffectief is in Nederland. Aangezien de duur van verblijf op de IC de grootste impact heeft op de totale kosten, is het voor toekomstige studies van belang om de impact van een dergelijk PCT algoritme op deze verblijfsduur nauwkeurig in kaart te brengen.

Hoofdstuk 7 betreft een kosteneffectiviteitsanalyse gebaseerd op het zorggebruik in de gerandomiseerde studie die beschreven is in hoofdstuk 6. De resultaten tonen aan dat de impact van het gebruik van dit PCT algoritme op kosten minimaal is (€-65), terwijl zowel de duur van antibioticagebruik als de sterfte tijdens de ziekenhuisopname afnemen in vergelijking met de standaardzorg (respectievelijk -1.2 dagen en -7.9%). Wanneer een tijdshorizon van één jaar wordt gehanteerd is de verwachting echter dat de hoge zorgkosten onder mensen die sepsis overleefd hebben leiden tot een (niet-significante) toename in kosten in de PCT groep vergeleken met standaardzorg (+€2,704). Dit resulteert in een incrementele kosteneffectiviteitsratio van ~€60,000 per gewonnen levensjaar in volledige gezondheid (QALY). Gebruik van dit PCT algoritme kan daarom niet kosteneffectief beschouwd worden bij

een 'willingness-to-pay' drempelwaarde van €20,000/QALY. Vervolgonderzoek is nodig om de gezondheidseconomische impact van een dergelijk PCT algoritme over een langere termijn, vanuit het maatschappelijk perspectief, in kaart te brengen.

Het derde deel van dit proefschrift (**hoofdstuk 8 en 9**) betreft de evaluatie van de kosten en effecten van het gebruik van uitgebreide laboratoriumdiagnostiek voor de diagnostiek en behandeling van anemie in de eerste lijn:

Hoofdstuk 8 beschrijft een studie over de effectiviteit van huisartsen met betrekking tot het vaststellen van de onderliggende oorzaak van anemie aan de hand van twee verschillende laboratoriumaanvragen. In deze studie is de effectiviteit gedefinieerd als het percentage patiënten met anemie waarin de juiste onderliggende oorzaak is vastgesteld door de huisarts, wanneer het gebruik van een uitgebreide laboratoriumaanvraag wordt vergeleken met een routine laboratoriumaanvraag. Hiervoor is een online vragenlijst verspreid onder 836 huisartsen. De vragenlijst bestond uit zes casussen van patiënten met anemie. In de eerste drie casussen konden huisartsen zelf kiezen welke testen ze uit zouden laten voeren (op basis van een vooraf gedefinieerde set van 14 laboratoriumtesten), dit werd de 'routine werkwijze' genoemd. In de volgende drie casussen kregen huisartsen direct het resultaat van alle 14 laboratoriumtesten te zien, en dit werd de 'uitgebreide werkwijze' genoemd. De resultaten tonen aan dat, vergeleken met een panel van experts, het gebruik van de uitgebreide werkwijze leidt tot een toename in het aantal patiënten waarin huisartsen in staat zijn om de juiste onderliggende oorzaak van anemie te diagnosticeren.

In **hoofdstuk 9** is vervolgens de impact van beide laboratoriumwerkwijzen geschat op zowel het aantal juiste diagnoses, de behandelbeslissingen als de bijbehorende kosten. Naast een 6.4% toename in het aantal patiënten waarin de correcte onderliggende oorzaak van anemie wordt gediagnosticeerd, wordt een kleine toename in kosten verwacht van €842 per patiënt in de routine werkwijze naar €845 per patiënt in de uitgebreide werkwijze. Vanwege deze minimale toename in kosten, en een toename in het aantal patiënten waarin de juiste onderliggende oorzaak kan worden vastgesteld, wordt gebruik van de uitgebreide werkwijze aanbevolen.

Het laatste deel van dit proefschrift (**hoofdstuk 10, 11 en 12**) beschrijft methodologische aspecten gerelateerd aan het uitvoeren van een (vroegtijdige) evaluatie van een diagnostische test of teststrategie:

Hoofdstuk 10 illustreert hoe schattingen van experts gebruikt kunnen worden in gezondheidseconomische modellen wanneer wetenschappelijk bewijs ontbreekt. Dit wordt geïllustreerd aan de hand van een case studie naar het gebruik van een combinatie van biomarkers om een non-ST elevatie myocard infarct snel uit te kunnen sluiten in patiënten

die zich met pijn op de borst op de spoedeisende hulp melden. Deze case studie vergelijkt de kosteneffectiviteit van gecombineerd gebruik van de heart-type fatty acid-binding protein (H-FABP) test, hoog-sensitieve troponine, en copeptine, met conventionele diagnostiek waarbij hoog-sensitieve troponine in serie gemeten wordt. De resultaten laten zien dat gebruik van deze combinatie van testen, aangevuld met troponine testen twee en zes uur na ziekenhuisopname, naar verwachting de meest kosteneffectieve strategie is ten opzichte van de standaardzorg. Daarnaast toont deze case studie aan dat schattingen van experts waardevol kunnen zijn voor het informeren van een vroegtijdige evaluatie van een medische technologie, om een initiële schatting te maken van de kosteneffectiviteit van nieuwe testen voordat ze verder worden ontwikkeld en (uiteindelijk) in de dagelijkse praktijk worden geïmplementeerd.

Hoofdstuk 11 heeft als doel om inzicht te verkrijgen in welke aspecten invloed hebben op de implementatie en het gebruik van diagnostische testen in de dagelijkse praktijk. Dit is geïllustreerd aan de hand van een case studie over twee POC testen in de eerste lijn, namelijk de POC geglyceerd hemoglobine (HbA_{1c}) test, en de POC C-reactief proteïne (CRP) test. Aan de hand van een literatuur review en validatie door 12 experts is een lijst van 20 criteria verkregen. Uit een Analytisch Hiërarchisch Proces (AHP) sessie met 10 van deze experts kwam naar voren dat de klinische relevantie van de test, de technische performance, en risico's (die gepaard gaan met de behandel- of managementbeslissing op basis van het testresultaat) als meest belangrijke criteria werden gezien, met relatieve gewichten van respectievelijk 22.2%, 12.6% en 8.5%. Wat de CRP test betreft hadden de experts de voorkeur voor een POC analyzer ten opzichte van dezelfde test in een centraal laboratorium, terwijl dezelfde experts geen voorkeur aangaven voor de HbA_{1c} test op een POC analyzer. De inzichten verkregen in deze studie kunnen zowel bijdragen aan het sturen van testontwikkeling als aan het verbeteren van de implementatie en gebruik van bestaande POC testen in de huisartsenpraktijk.

Hoofdstuk 12 beschrijft de ontwikkeling van een uitgebreide checklist, welke als doel heeft om het bewustzijn te vergroten omtrent het opnemen van aspecten die de impact van diagnostische tests en biomarkers beïnvloeden in gezondheidseconomische modellen. Deze checklist is ontworpen op basis van een scoping review, welke kritisch is beoordeeld door vier experts. Deze experts beoordeelden de duidelijkheid, aanwezigheid van overbodige aspecten, en compleetheid van de lijst. Dit resulteerde in een lijst van 43 aspecten waarvan in een gezondheidseconomische evaluatie idealiter expliciet de overweging gemaakt moet worden om deze aspecten te includeren, dan wel onderbouwd te excluderen. Op deze manier kan deze checklist kan bijdragen aan het verbeteren van de compleetheid en transparantie van dergelijke gezondheidseconomische evaluaties.

De conclusie van dit proefschrift is dat het van groot belang is om de verwachte impact van een diagnostische test (of combinatie van testen) op kosten en effecten in kaart te brengen, voordat deze test wordt geïmplementeerd in de praktijk. Evaluatie van gezondheids-

technologie (oftewel 'Health Technology Assessment'), en vooral vroegtijdige evaluatie, omvatten een aantal erg waardevolle technieken om deze impact in kaart te brengen. Naast gezondheidseconomische modellen zijn andere technieken die deel uit maken van vroegtijdige evaluatie van gezondheidstechnologie, waaronder multi-criteria decision analysis (MCDA), waardevol om inzicht te krijgen in factoren die de implementatie en het gebruik van nieuwe testen bevorderen dan wel belemmeren. Om de waarde van diagnostiek volledig te kunnen kwantificeren, en om de implementatie en gebruik van (nieuwe) testen in te kunnen schatten, is het echter van belang om naar bredere aspecten te kijken dan enkel naar de impact op kosten en de impact op de gezondheidstoestand van de patiënt. Dergelijke aspecten omvatten bijvoorbeeld geruststelling omtrent het niet hebben van een ernstige aandoening, of bewust zijn van het hebben van een genetische aandoening, wat beslissingen omtrent het krijgen van kinderen kan beïnvloeden. Naast de kosteneffectiviteit van een test, zullen deze bredere waarde-aspecten naar alle waarschijnlijkheid ook invloed hebben op testimplementatie en gebruik.

The background of the image is a dark blue gradient. It is decorated with numerous semi-transparent circles of various sizes, creating a bokeh effect. Some circles are a lighter, brighter blue, while others are a darker, more muted blue. They are scattered across the frame, with some overlapping each other.

Dankwoord

Daar staan we dan! Toen ik zo'n acht jaar geleden als HBO'er voor het eerst een promotie bij mocht wonen was ik wel enigszins onder de indruk, en ik weet nog goed dat ik dacht: 'Dat ga ik nooit doen!' Maar een aantal jaren later mag ik toch met trots zeggen dat ik mijn proefschrift ga verdedigen. Ondanks dat je vaak hoort hoe zwaar promoveren kan zijn, kan ik gelukkig vooral zeggen: 'promoveren is leuk!' Ik wil in dit dankwoord dan ook terugblikken op deze plezierige en ontzettend leerzame periode.

Ron, het enthousiasme waarmee jij als begeleider van mijn masteropdracht vertelde over het belang van evaluatie van laboratoriumdiagnostiek, is ook op mij overgeslagen. Ik wil je hartelijk bedanken voor het vertrouwen dat je als promotor in mij hebt gesteld om mijn master thesis een vervolg te geven in de vorm van dit promotieonderzoek. Dankzij jou grote netwerk zijn er veel leuke en leerzame samenwerkingen tot stand gekomen, wat uiteindelijk heeft geresulteerd in dit proefschrift. Bedankt voor je kritische blik, je onmisbare klinische invalshoek, en niet te vergeten, voor het feit dat ik nu weet waar de beste Bossche bollen te koop zijn!

Maarten, met jou kennis van de methodologie en je 'helikopterview' vormde je als promotor samen met Ron een perfecte combinatie. Ik heb grote bewondering voor hoe snel je kunt schakelen tussen alle onderwerpen waarbij je betrokken bent. Ondanks de vele afstudeerders en PhDs die je begeleidt, weet je binnen een paar minuten precies de verbeterpunten in onze onderzoeken aan te wijzen. Bedankt voor je vertrouwen, je enthousiasme, en voor het feit dat je me tijdens mijn onderzoek al de kans hebt gegeven om aan vervolgprojecten te werken. Ik kijk uit naar onze verdere samenwerking, en ik wens je heel veel succes en plezier in Melbourne!

Voor één van de grootste uitdagingen tijdens mijn promotieonderzoek is mijn copromotor van erg grote waarde geweest. Erik, zonder al jou kennis over modelleren en programmeren waren deze kosteneffectiviteitsmodellen niet tot stand gekomen. Daarnaast wist jij altijd het geduld op te brengen om mij iets uit te leggen over verdelingen, multiële imputaties, of bootstraps, waarbij ik je regelmatig met glazige ogen aan heb gekeken omdat ik het (nog steeds) niet begreep. En niet te vergeten: bedankt voor je kritische blik, je humor, en de vele koppen koffie of thee. Daarnaast ben ik je nog steeds dank verschuldigd dat je een uur hebt omgereden om me bij een B&B in Brussel af te zetten (terwijl ik het verkeerde adres in de navigatie had ingevoerd), en jij vervolgens nog helemaal naar huis moest rijden.

Ook gaat mijn dank uit naar mijn paranimfen die op deze belangrijke dag naast mij staan. Marieke, ik vond het ontzettend leuk dat ik jou paranimf mocht zijn. Dat ik jou ook als mijn paranimf zou vragen stond al geruime tijd vast! Ook al vertoonden onze werkzaamheden de afgelopen zes jaar niet ontzettend veel overlap, we hadden altijd genoeg gespreksstof in de vorm van trouwen, verbouwen of verhuizen. Bedankt voor je luisterend oor en de gezelligheid, ik had me geen betere kamergenoot kunnen wensen!

Jorieke, mijn schoonzusje, dat we vorig jaar samen de Nijmeegse Vierdaagse hebben gelopen was in één woord onvergetelijk. Omdat dit vele uren training vergt, was het onvermijdelijk dat je veel over mijn onderzoek aan hebt moeten horen. Desondanks heb je altijd interesse getoond, en ik vind het daarom ontzettend leuk dat je vandaag mijn paranimf wilt zijn!

Daarnaast wil ik de leden van mijn promotiecommissie hartelijk bedanken voor het lezen en beoordelen van mijn proefschrift, en voor het aanwezig zijn bij mijn verdediging.

Ik wil de Raad van Bestuur van het Jeroen Bosch Ziekenhuis bedanken voor de financiële ondersteuning die het mogelijk heeft gemaakt om dit promotieonderzoek uit te voeren.

Tevens gaat mijn dank uit naar alle huisorganisaties voor het verspreiden van de enquêtes, en naar alle huisartsen en cardiologen die deel hebben genomen aan de interviews en enquêtes voor diverse artikelen in dit proefschrift. Zonder jullie waardevolle input waren deze hoofdstukken niet tot stand gekomen, hartelijk dank hiervoor!

In de afgelopen vier jaar heb ik met ontzettend leuke en gemotiveerde mensen mogen samenwerken. Helaas wordt dit dankwoord te lang om iedereen persoonlijk te bedanken, maar ik wil toch even bij een aantal van deze samenwerkingen stilstaan.

Om te beginnen de samenwerking met het team uit het Albert Schweitzer Ziekenhuis in Dordrecht. Beste dr. Levin, Mark-David, volgens mij zijn we er goed in geslaagd om de ‘communicatiekloof’ tussen een gezondheidswetenschapper en een internist-hematoloog te overbruggen. Bedankt voor je sterke betrokkenheid en je klinische blik! Karlijn, bedankt voor je input en je altijd snelle reacties op mijn vragen. Onze geheel toevallige gezamenlijke wandeltocht zal ik niet snel vergeten. Annemarie, ik heb bewondering voor hoe jij onderzoek combineert met het drukke leven van coschappen. Bedankt voor onze prettige samenwerking! Tevens wil ik alle coauteurs bedanken voor hun waardevolle input op onze twee manuscripten.

Daarnaast gaat mijn dank uit naar de vakgroep huisartsgeneeskunde van de Universiteit Maastricht. Beste Robert, ik heb bewondering voor jou enthousiasme en – afgaande op de tijdstippen van je e-mails – voor de diverse nachtelijke uren die je hebt gewerkt om je proefschrift af te ronden. Bedankt voor de fijne samenwerking!

Prof. dr. Dinant, beste Geert-Jan, naast onze samenwerking voor bovengenoemd onderzoek aan de Universiteit Maastricht, was je input ook van grote waarde voor het project over de kosteneffectiviteit van anemiediagnostiek, door middel van het kritisch reviewen van de enquête, alsmede door de feedback die je op beide artikelen hebt geleverd. Hartelijk dank voor je input en voor het plaatsnemen in mijn promotiecommissie!

I would like to thank the members from the SAPS steering committee for sharing the data from this highly valuable study, and for critically reviewing our manuscript. Dr. Maarten Nijsten and Arezoo Shajiei, thank you for your great help in obtaining the data from the database and with writing the manuscript. Also, I would like to thank dr. Jos van Oers for helping me with selecting the most relevant and up-to-date literature.

Also, I would like to express my gratitude to professor Tracy Merlin, professor Charles Phelps, professor Milton Weinstein and dr. Martin Henriksson for critically reviewing our checklist and manuscript. Your valuable inputs have greatly contributed to improving the completeness and quality of the checklist. Thank you for the very nice cooperation!

Sofie, het laatste artikel was niet op tijd afgerond zonder jou hulp met het schrijven van de R code. Hartelijk dank daarvoor! Karin, bedankt voor je ondersteuning hierbij, en voor je hulp bij diverse statistische vraagstukken, met name in de beginfase van mijn promotieonderzoek.

Daarnaast gaat mijn dank uit naar dr. Marjan Hummel. Beste Marjan, ondanks dat je helaas al enige tijd geen collega meer van ons bent, bleef je erg betrokken bij het onderzoek naar factoren die implementatie van point-of-care testen beïnvloeden. Jouw hulp was van grote waarde om de analyses af te ronden in een vakgebied waarin ik zelf nog iets minder bekend was.

Tevens gaat mijn dank uit naar dr. Rogier Hopstaken, en alle andere experts die deel hebben genomen aan de hiervoor genoemde studie. Hartelijk dank voor jullie deelname aan de interviews, en aan de AHP-sessie. Zonder jullie input hadden we deze studie niet kunnen doen. Rogier, bedankt voor je verdere input bij het afronden van dit manuscript, en voor het feit dat je deel uitmaakt van mijn promotiecommissie.

Also, I would like to thank everyone who was involved in the project I was already working on before my PhD started. Therefore, I would like to thank my former colleagues from Panaxea b.v., as well as Henk Nijzing, Stefan Leichtl, and all other members from Thermo Fisher Scientific who were involved in this project, for our pleasant and constructive collaboration.

Uiteraard zou ik niet terug kunnen kijken op vier leuke jaren zonder gezellige collega's. Daarom wil ik al mijn collega's van de vakgroep HTSR hartelijk bedanken! De dagjes Winterberg, de 'bakbokaal', onze deelname aan het beachvolleybaltoernooi, het 'heel HTSR bakt'-boek met bijbehorend keukenschort dat we voor ons huwelijk hebben mogen ontvangen, en de vakgroeputjes (waaronder de kanotocht die bijna uitmondde in een zwemles), hebben deze periode tot een onvergetelijke tijd gemaakt! Bovendien waren mijn collega's altijd erg welwillende proefpersonen als ik weer eens een nieuw recept voor taart, cake of koekjes had uitgeprobeerd.

Een belangrijke spil in deze vakgroep wordt ongetwijfeld gevormd door het secretariaat. Joke, Annelies, Maaïke, Tanja, en Ingrid, bedankt voor jullie ondersteuning. Annelies, ook al ben je helaas geen HTSR-collega meer, je luisterend oor en interesse is nog steeds van grote waarde. Dankjewel voor je gezelligheid, en voor de ondersteuning die je geboden hebt bij het regelen van mijn promotie! Ingrid, ook al was mijn promotie voor jou de eerste waarbij je echt betrokken was, je ziet dat het allemaal gelukt is! Hartelijk dank voor je ondersteuning hierbij!

Ik wil ook alle studenten bedanken die een bijdrage hebben geleverd aan mijn promotieonderzoek. Ik heb ervan genoten dat ik als begeleider ook van jullie kon leren, en heb zelfs een aantal VBA-codes aan jullie te danken! In het bijzonder wil ik de afstudeerders bedanken wiens opdrachten uiteindelijk tot een artikel in mijn proefschrift hebben geleid: Marco, Maran, Soraya, Elra, hartelijk dank voor jullie bijdrage!

Dit proefschrift was niet tot stand gekomen zonder ondersteuning van familie en vrienden. Daarom wil ik graag even stil staan bij iedereen die de afgelopen vier jaar al mijn verhalen, inclusief (kleine) frustraties, aan heeft moeten horen.

Als eerste wil ik mijn vrienden en vriendinnen bedanken. De continue interesse die jullie getoond hebben in de voortgang van mijn proefschrift heb ik ontzettend gewaardeerd. Daarnaast zorgden jullie er ook voor dat ik soms even niet met mijn onderzoek bezig was, zodat ik na het weekend weer fris aan de nieuwe werkweek kon beginnen! Bedankt voor alle gezelligheid in de vorm van de vele verjaardagen, barbecues, bowling-uitjes, spelletjes-avonden, high-tea middagen, en de (te) lange, maar erg gezellige wandelingen.

Daarnaast waren mijn badmintontrainingen, de wedstrijden, (de kantine na afloop), en de bijbehorende gezelligheid de perfecte manier om mijn hoofd leeg te maken, al moet ik zeggen dat een drukke werkdag niet altijd positieve invloed had op mijn kwaliteiten op het veld.

Ook mijn schoonouders verdienen een plekje in dit dankwoord. Dat jullie altijd interesse hebben getoond in mijn onderzoek blijkt wel uit de vele succeswensen die ik heb mogen ontvangen wanneer ik naar een congres ging of ergens een presentatie gaf. Bedankt daarvoor!

Pap, mam, jullie hebben Patrick en mij altijd gemotiveerd om te gaan studeren, maar het belangrijkste was nog wel dat jullie altijd trots op ons waren, onder het motto: zolang je maar je best doet, en doet wat je leuk vindt. Jullie nuchterheid en rust, gecombineerd met de altijd aanwezige interesse in mijn onderzoek, heb ik ontzettend gewaardeerd. Dank jullie wel!

Patrick, als ik vroeger een 9 voor een toets had gehaald, was jou commentaar dat ik 30% teveel had gedaan, want een 6 is immers ook voldoende. Natuurlijk moest ik dit met een korreltje

zout nemen, maar het hielp mij om te relativieren, en ook tijd voor leuke dingen (zoals samen met jou en Jorieke) niet te vergeten. Dankjewel daarvoor, en nu op naar het volgende feest(je)!

Lieve Tom, jij hebt me altijd dat kleine duwtje in de rug gegeven om mij te motiveren om geen uitdaging uit de weg te gaan. Vaak wordt gezegd dat verhuizen, verbouwen, of samen op vakantie gaan de ultieme relatietesten zijn. Voeg daar (gelijktijdig) promoveren aan toe, en dan nog hebben we ze allemaal met glans doorstaan. Tom, bedankt voor je geduld, je vertrouwen, en de rust die je mij brengt. Zonder jouw steun was dit zeker niet gelukt!

Curriculum Vitae & Publications

CURRICULUM VITAE

Michelle Kip was born on the 19th of January 1989 in Oldenzaal, the Netherlands. In 2007 she graduated from secondary school at the Twents Carmel College in Oldenzaal. From 2007 to 2011, she studied Biology and Medical Laboratory Research at Saxion University of Applied Sciences in Enschede. During this study, she did a five-month internship as a laboratory technician at the laboratory hospital of Medisch Spectrum Twente in Enschede. For her bachelor thesis, she investigated changes in gene expression in dedifferentiating human chondrocytes, at the Tissue Regeneration group of the University of Twente, Enschede. In 2010, she defended her bachelor thesis cum laude. For this thesis, she was awarded the 'Zilveren Vlam', which is an annual prize for the best graduate student in Biology and Medical Laboratory Research in the Netherlands.

Following this, she started with the (pre)master program of the master Health Sciences at the University of Twente. In 2012 she graduated cum laude, with her master thesis about the cost-effectiveness of a triple biomarker for early exclusion of acute coronary syndrome in chest pain patients. Following her graduation, she worked at Panaxea b.v. (a consultancy company), for two years. During this period, she worked on several projects in the field of (early) health technology assessment of diagnostic tests, pharmaceuticals, and medical devices. From 2009 to 2014, she also worked part-time as a laboratory technician at Medlon (at Medisch Spectrum Twente).

In 2014, she started as a PhD candidate at the department of Health Technology and Services Research (HTSR) at the University of Twente. Her PhD project concerned (early) health technology assessment of point-of-care and laboratory diagnostics. The results of this PhD project are described in this dissertation.

Since June 2017, she is involved in the UCAN CAN-DU project, a collaboration between the Netherlands and Canada, to set-up a large, multicenter, prospective trial, aiming to optimize treatment strategies in patients with juvenile idiopathic arthritis. Since December 2017, she works on a cost-effectiveness analysis of IMPACT, a technology aimed at stimulating repair of cartilage tissue by using stem cells, in collaboration with the University Medical Center Utrecht.

After her PhD trajectory, Michelle will start working as a postdoctoral researcher at the department HTSR on the two abovementioned projects. She hopes to continue and expand her collaboration with different (international) disciplines and institutions, in order to contribute to further improving the cost-effectiveness and quality of care.

LIST OF PUBLICATIONS

Journal papers

Annemarie Schop, Michelle M.A. Kip, Karlijn Stouten, Soraya Dekker, Jürgen Riedl, Ron J. van Houten, Joost van Rosmalen, Geert-Jan Dinant, Maarten J. IJzerman, Hendrik Koffijberg, Patrick J.E. Bindels, Ron Kusters, Mark-David Levin (2017). The effectiveness of a standardized versus a personalized laboratory analysis in the diagnosis of anaemia in general practice. *Ann Clin Biochem*, p. 4563217748680

Michelle M.A. Kip, Annemarie Schop, Karlijn Stouten, Soraya Dekker, Geert-Jan Dinant, Hendrik Koffijberg, Patrick J.E. Bindels, Maarten J. IJzerman, Mark-David Levin, Ron Kusters. (2017) Assessing the cost-effectiveness of a routine versus an extensive laboratory work-up in the diagnosis of anaemia in Dutch general practice. *Ann Clin Biochem*, p. 4563217748984.

Robert T.A. Willemsen, Michelle M.A. Kip, Hendrik Koffijberg, Ron Kusters, Frank Buntinx, Jan F.C. Glatz, Geert-Jan Dinant (2017). Early Health Technology Assessment of future clinical decision rule aided triage of patients presenting with chest pain in primary care. *Prim Health Care Res Dev*, (18): p.1 -13.

Michelle M.A. Kip, Hendrik Koffijberg, Marco J. Moesker, Maarten J. IJzerman, Ron Kusters (2017). The cost-utility of point-of-care troponin testing to diagnose acute coronary syndrome in primary care. *BMC Cardiovasc Disord*. 17(1): p. 213.

Michelle M.A. Kip, Amber M. Noltjes, Hendrik Koffijberg, Maarten J. IJzerman, Ron Kusters (2017). Improving early exclusion of acute coronary syndrome in primary care: the added value of point-of-care troponin as stated by general practitioners. *Prim Health Care Res Dev*, p. 1-12.

Tommy S. de Windt, Juliette C. Sorel, Lucienne A. Vonk, Michelle M.A. Kip, Maarten J. IJzerman, Daniel B.F. Saris (2017). Early health economic modelling of single-stage cartilage repair. Guiding implementation of technologies in regenerative medicine. *J Tissue Eng Regen Med*. 11(10): p. 2950-2959..

Michelle M.A. Kip, Hendrik Koffijberg, Maarten J. IJzerman, Ron Kusters (2016). Procalcitonin to guide antibiotic stewardship in intensive care. *Lancet Infect Dis*. 16(8). p. 888

Michelle M.A. Kip, Lotte M.G. Steuten, Hendrik Koffijberg, Maarten J. IJzerman, Ron Kusters (2016). Using expert elicitation to estimate the potential impact of improved diagnostic performance of laboratory tests: a case study on rapid discharge of suspected non-ST elevation myocardial infarction patients. *J Eval Clin Pract*.

Roosmarijn T.M. van Hooijdonk, Lotte M.G. Steuten, Michelle M.A. Kip, Helma Monteban, Marianne R. Mulder, Floris van Braam Houckgeest, Johannes P. van der Sluijs, Ameen Abu-Hanna, Peter E. Spronk, Marcus J. Schultz (2015). Health economic evaluation of a strict glucose control guideline implemented using point-of-care testing in three intensive care units in The Netherlands. *Appl Health Econ and Health Policy*. 13(4). p. 399-407.

Michelle M.A. Kip, Helma Monteban, Lotte M.G. Steuten (2015). Long-term cost-effectiveness of Oncotype DX versus current clinical practice from a Dutch cost perspective. *J Comp Eff Res*. 4(5). p. 433-445.

Michelle M.A. Kip, Ron Kusters, Maarten J. IJzerman, Lotte M.G. Steuten (2015). A PCT algorithm for discontinuation of antibiotic therapy is a cost-effective way to reduce antibiotic exposure in adult intensive care patients with sepsis. *J Med Econ*. 18(11). p. 944-953.

Bin Ma, Jeroen C.H. Leijten, Ling Wu, Michelle M.A. Kip, Clemens A. van Blitterswijk, Janine N. Post, H.B.J. (Marcel) Karperien (2013). Gene expression profiling of dedifferentiated human articular chondrocytes in monolayer culture. *Osteoarthritis cartilage*. 21(4). p. 599-603.

Conference contributions

Michelle M.A. Kip, Jos A. van Oers, Maarten W.N. Nijsten, Maarten J. IJzerman, Hendrik Koffijberg, Ron Kusters (2018). Procalcitonin testing to guide antibiotic treatment duration in intensive care patients with sepsis: a trial-based cost-effectiveness analysis relates an increased survival rate to a minimal impact on costs. In: *Voorjaarscongres NVKC*, April 11th - 13th, 2018, Papendal, The Netherlands.

Michelle M.A. Kip, J. Marjan Hummel, Elra B. Eppink, Hendrik Koffijberg, Rogier M. Hopstaken, Maarten J. IJzerman, Ron Kusters (2018). Understanding the adoption and use of point-of-care tests in Dutch general practices using Multi-Criteria Decision Analysis. In: *Voorjaarscongres NVKC*, April 11th - 13th, 2018, Papendal, The Netherlands.

Michelle M.A. Kip, Annemarie Schop, Karlijn Stouten, Soraya Dekker, Geert-Jan Dinant, Hendrik Koffijberg, Patrick J.E. Bindels, Maarten J. IJzerman, Mark-David Levin, Ron Kusters (2017). A standardized versus a personalized laboratory work-up in the diagnosis of anemia in Dutch general practices: a cost-effectiveness analysis. Poster and podium presentation. In: *Voorjaarscongres NVKC*, April 5th - 7th, 2017, Scheveningen, The Netherlands. (winner of poster award)

Michelle M.A. Kip, Ron Kusters (2016). Procalcitonin testing: added value and cost-effectiveness. Podium presentation. In: *Transformation forum*, December 8th, 2016, Mechelen, Belgium.

Michelle M.A. Kip, Ron Kusters (2016). Procalcitonin testing: added value and cost-effectiveness. Podium presentation. In: Transformation forum, December 1st, 2016, Rotterdam, The Netherlands.

Michelle M.A. Kip, Hendrik Koffijberg, Ron Kusters, Maarten J. IJzerman (2016). Value of diagnostic information: how to show the value of your innovation to clinicians and decision makers in healthcare. Podium presentation. In: MIRA day, November 3rd, 2016, Enschede.

Michelle M.A. Kip and Ron Kusters (2016). Point-of-care troponine testen in de huisartsenpraktijk: wat is nodig voor implementatie? Podium presentation. In: World of Technology & Science, October 8th, 2016, Utrecht, The Netherlands.

Michelle M.A. Kip, Hendrik Koffijberg, Marco J. Moesker, Maarten J. IJzerman, Ron Kusters (2016). The cost-effectiveness of point-of-care troponin testing to exclude acute coronary syndrome in primary care. Poster. In: SMDM, 16th Biennial European Conference, Special Focus on Improving Diagnostic Decision Making, June 6th – 12th, 2016, London, United Kingdom.

Michelle M.A. Kip, Amber M. Noltes, Hendrik Koffijberg, Ron Kusters (2016). Point-of-care troponin testing in Dutch primary care: preferences and referral decisions of general practitioners. Poster and podium presentation. In: Voorjaarscongres NVKC, April 13th – 15th, 2016, Papendal, The Netherlands.

Michelle M.A. Kip, Ron Kusters (2016). De kosteneffectiviteit van een triple hartmarker voor het snel uitsluiten van NSTEMI. Podium presentation. In: Wetenschapsmiddag Jeroen Bosch Ziekenhuis, November 11th, 2015, 's-Hertogenbosch, The Netherlands.

Michelle M.A. Kip, Marco J. Moesker, Lotte M.G. Steuten, Ron Kusters (2015). Assessing the cost-effectiveness of point-of-care testing for patients with symptoms suggestive of acute coronary syndrome. Poster and podium presentation. In: 68e Voorjaarscongres NVKC, April 15th – 17th, 2015, Veldhoven, The Netherlands. (winner of poster award)

Marloes van der Maas, Michelle M.A. Kip, Gertjan Mantjes, Lotte M.G. Steuten (2015). A procalcitonin algorithm used in adult ICU patients with sepsis saves costs by reducing antibiotic resistance and C. difficile infections. Poster. In: ISPOR 18th Annual European Congress, November 7th – 11th, 2015, Milan, Italy.

Michelle M.A. Kip, Marina van den Bosch, Kathelijn Fischer, Rienk Y.J. Tamminga, Isabelle Lepage-Nefkens (2014). Cost-effectiveness analysis evaluating Factor VIII as primary prophylaxis treatment for patients with severe haemophilia A in The Netherlands. Poster. In: ISPOR 17th Annual European Congress, November 8th – 12th, 2014, Amsterdam, The Netherlands.

Michelle M.A. Kip, Ron Kusters, Maarten J. IJzerman, Lotte M.G. Steuten (2014). Cost-effectiveness of a PCT algorithm for discontinuation of antibiotic therapy in adult intensive care patients with sepsis. Poster. In: MIRA day, November 27th, 2014, Enschede, The Netherlands.

Michelle M.A. Kip, Ron Kusters, Lotte M.G. Steuten (2014). Cost minimization in the intensive care unit: the added value of procalcitonin. Poster. In: NVKC voorjaarscongres 2014, April 9th – 11th, 2014, Veldhoven, The Netherlands.

Lotte M.G. Steuten, Michelle M.A. Kip, Roosmarijn van Hooijdonk, Helma Monteban, Peter Spronk, Marcus Schultz (2014). PDB73 – Economic evaluation of blood glucose point-of-care testing in the intensive care unit. Value in Health, 17 (2). A250. ISSN 1098-3015.

Lotte M.G. Steuten, Martijn Schutten, Michelle M.A. Kip, Maarten J. IJzerman (2014). PHP128 – Assessing the impact of a multi-million public-private partnership for translational research: the centre for translational molecular medicine (CTMM). Value in Health, 17 (3). A31. ISSN 1098-3015.

Michelle M.A. Kip, Ron Kusters, Lotte M.G. Steuten (2013). Cost-minimization in the intensive care unit: the added value of procalcitonin. Poster. In: Topics in Intensive Care, December 11th – 12th, 2013, Lunteren, The Netherlands. (winner of poster award)

Ruud de Jonge, Lotte M.G. Steuten, Martijn Schutten, Michelle M.A. Kip, Helma Monteban (2013). Cost-effectiveness of Oncotype DX test vs. current clinical practice: a Dutch cost perspective. Breast, 22 (. 1). S113. ISSN 0960-9776.

Ruud de Jonge, Martijn Schutten, Michelle M.A. Kip, Lotte M.G. Steuten (2013). Long term cost-effectiveness of Oncotype DX test vs. current clinical practice from a Dutch cost perspective. Poster. In: 13th St. Gallen international breast cancer conference, St. Gallen, Switzerland.

Michelle M.A. Kip, Lotte M.G. Steuten, Ron Kusters (2013). Decreasing time to diagnosis in patients with acute chest pain. Poster. In: NVKC congress, April 17th – 19th, 2013, Veldhoven, The Netherlands.

Students supervised

Eric ten Bosch (MSc Health Sciences, 2017). Evaluation of the potential health benefits and cost savings of a point-of-care analyser for individuals with prediabetes. Supervised with Hendrik Koffijberg and Paul Smit.

Aniek N. Vollenbroek (MSc Health Sciences, 2017). The health economic impact of (mis) classifying pregnant women with gestational diabetes mellitus. Supervised with Ron Kusters.

Twan J. Seydel (MSc Health Sciences, 2017). Evaluating the health economic impact of misclassification during the cardiovascular risk assessment due to HDL-c measurement bias. Supervised with Ron Kusters.

Soraya Dekker (MSc Health Sciences, 2016). Diagnosis and treatment of anaemia in Dutch general practices: the effect of an extensive laboratory work-up. Supervised with Ron Kusters and Hendrik Koffijberg.

Elra B. Eppink (MSc Health Sciences, 2016). The development of a measurement instrument with determinants affecting the implementation of point-of-care tests in Dutch general practices. Supervised with J.M. (Marjan) Hummel and Ron Kusters.

Jeriela M.C. Alphenaar (MSc Health Sciences, 2015). The incremental cost-effectiveness of introducing PCR for early detection of E.coli infections associated with acute diarrhea compared to conventional microbiological methods. Supervised with Ron Kusters.

Laura M.J. van Maanen (MSc Health Sciences, 2015). Evaluatie van medische technologie: ziekenhuisgroep Twente. Supervised with J.M. (Marjan) Hummel.

Amber M. Noltes (MSc Health Sciences, 2015). Point-of-care troponin testing in Dutch primary care: General practitioners' opinions. Supervised with Ron Kusters and Hendrik Koffijberg.

Daphne J.J. Duteweert (MSc Health Sciences, 2015). The impact of C-reactive protein point-of-care tests on diagnostic test use and antibiotic prescriptions for patients with lower respiratory tract infections in primary care. Supervised with Ron Kusters and Hendrik Koffijberg.

Shannon van Hoorn (BSc Health Sciences, 2015). De diagnostiek van vitamine B1, B6, B12, foliumzuur, en vitamine D25 deficiëntie in de huisartsenpraktijk. Supervised with Ron Kusters.

Marco J. Moesker (MSc Health Sciences, 2014). Assessing the cost-effectiveness of point-of-care testing for primary care patients with symptoms suggestive of acute coronary syndrome: a threshold analysis. Supervised with Ron Kusters, Maarten J. IJzerman and Lotte M.G. Steuten.

Erwin E.A. Veneman (MSc Health Sciences, 2014). Assessment of the effectiveness and cost-effectiveness of reflective testing compared to regular testing with regard to iron deficiency anaemia in primary care. Supervised with Ron Kusters and Lotte M.G. Steuten.

TUSSENSCHIET:

*DIT VEL WORDT TIJDENS DE
AFWERKING VERWIJDERD*

SEPARATION:

*THIS SHEET WILL BE REMOVED
DURING THE BINDING PROCESS*

TUSSENSCHIET:

*DIT VEL WORDT TIJDENS DE
AFWERKING VERWIJDERD*

SEPARATION:

*THIS SHEET WILL BE REMOVED
DURING THE BINDING PROCESS*