Health technology assessments alongside the translational pathway of innovations in oncology

Providing guidance towards implementation

Melanie Lindenberg

Health technology assessments alongside the translational pathway of innovations in oncology

Providing guidance towards implementation

Melanie Lindenberg

Dit proefschrift is goedgekeurd door:

Promotor
Prof. dr. W.H. van Harten

Copromotor
Dr. V.P. Retèl

Cover and chapter intros: Rachel van Esschoten, DivingDuckDesign (www.divingduckdesign.nl)
Layout: Angela Lindenberg and Melanie Lindenberg
Printing: Ipskamp

The cover visualizes the path from a conceptual idea of an innovation to its use in the clinic. This yellow road refers to the "yellow brick road" which is a fictional element in, amongst others, the movie the Wizard of Oz and the musical the Wiz; where it symbolizes the road to success. The purple icons on the road represent the innovations that are part of this dissertation. The white icons represent the methods used to guide the innovations on the road to success.

The work described in this dissertation was performed at the Netherlands Cancer Institute – Antoni van Leeuwenhoek, Amsterdam, the Netherlands.
The studies presented in this dissertation were financially supported by the Netherlands Cancer Institute and research grants received from ZonMw (TIL study) and Intuitive Surgical (evaluation of the Da Vinci robot).

ISBN: 978-90-365-5068-0
DOI: 10.3990/1.9789036550680
This thesis is part of the Health Science Series, HSS 20-35, department Health Technology and Services Research, University of Twente, Enschede, the Netherlands. ISSN: 1878-4968.
The printing of this thesis was financially supported by the Netherlands Cancer Institute, University of Twente and Pfizer.

© 2020 Maria Anna Lindenberg, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

HEALTH TECHNOLOGY ASSESSMENTS ALONGSIDE THE TRANSLATIONAL PATHWAY OF INNOVATIONS IN ONCOLOGY
PROVIDING GUIDANCE TOWARDS IMPLEMENTATION

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. ir. A. Veldkamp,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
vrijdag 11 december 2020 om 10.45 uur

door

Maria Anna Lindenberg
Geboren op 23 maart 1992,
te Goes

Promotiecommissie

Voorzitter
Prof. dr. T.A.J. Toonen

Promotor
Prof. dr. W.H. van Harten

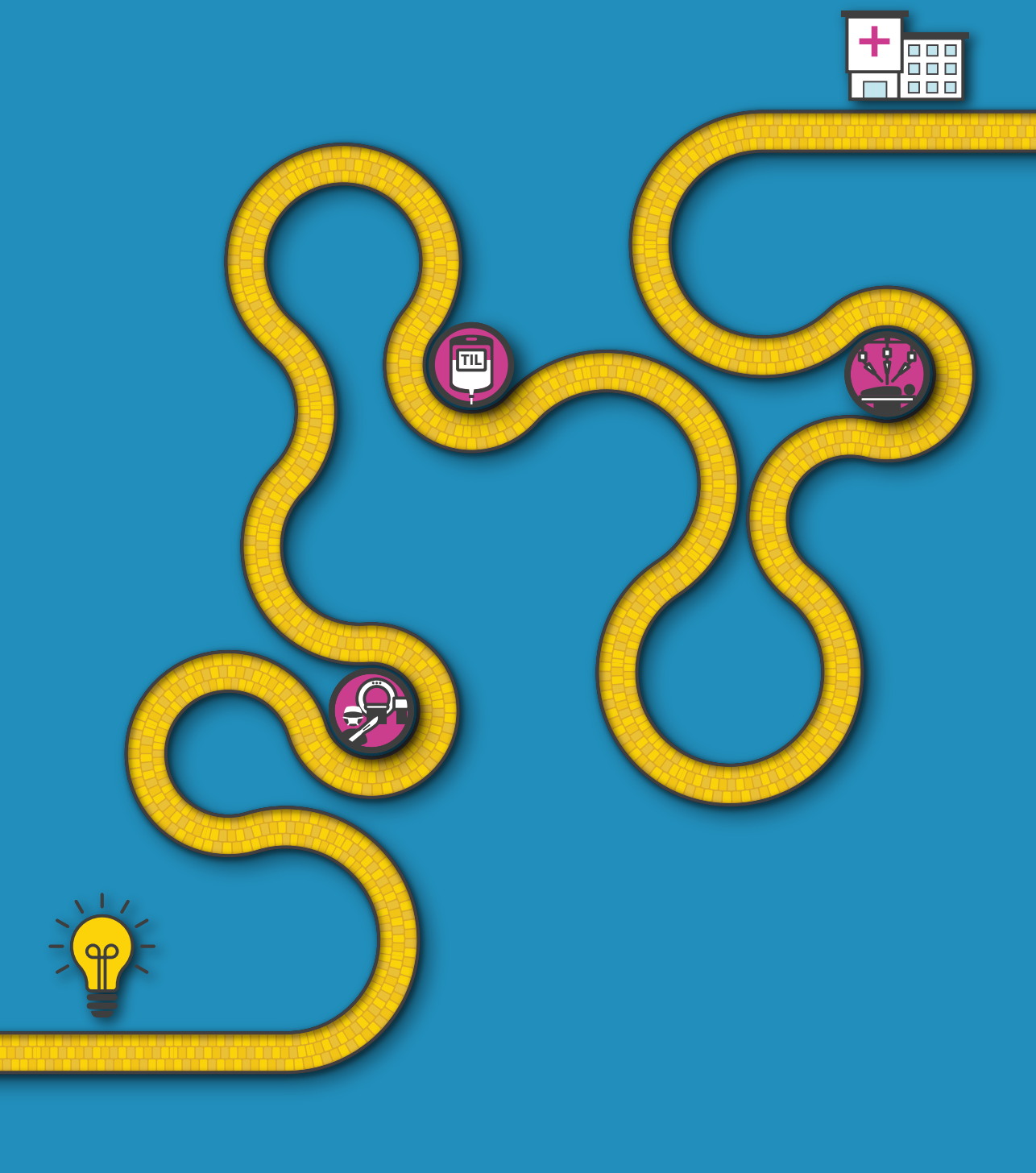
Copromotor
Dr. V.P. Retèl

Leden
Prof. dr. Ir. R.M. Verdaasdonk (Universiteit Twente)
Dr. ir. H. Koffijberg (Universiteit Twente)
Prof. dr. J.P. Ruurda (Universiteit Utrecht)
Prof. dr. A. Klink (Vrije Universiteit)
Prof. dr. A.J.M. van den Eertwegh (Amsterdam UMC)
Dr. ir. N.C. van der Vaart (LUMEN medical)

Paranimfen

Frank Halfwerk
Angela Lindenberg

TABLE OF CONTENTS


TIL

1

Introduction

1

Healthcare costs increased rapidly over the years due to trends in demographic factors, and an impressive launch of new healthcare innovations[1]. Therefore, healthcare budgets are under pressure, resulting in an increased awareness that money has to be spent wisely[2], and health innovations need to prove their value for money before adoption in the clinic.

In the development process of health innovations, several barriers can appear that slow their way to the patient or even make this path inaccessible[3]. These barriers mainly appear at two crucial moments in the translational process, the process between having a conceptual idea of a promising health innovation and adoption in the clinic, which are visualized and described by the "Valleys of death"[4] (Figure 1). The first valley is located between having a conceptual idea and obtaining market access. In this phase, for example obtaining research funding to perform the required translational research / translational medicine could be challenging. The second valley finds itself in between market access, and clinical use and reimbursement[4]. One way to control healthcare budgets is to evaluate whether a developed innovation is – besides safe and effective – cost-effective before market entry. Another option for budget control is to prevent investing (public) money in technologies that fail to come to the market since the failure of developed technology is expensive[5,6].

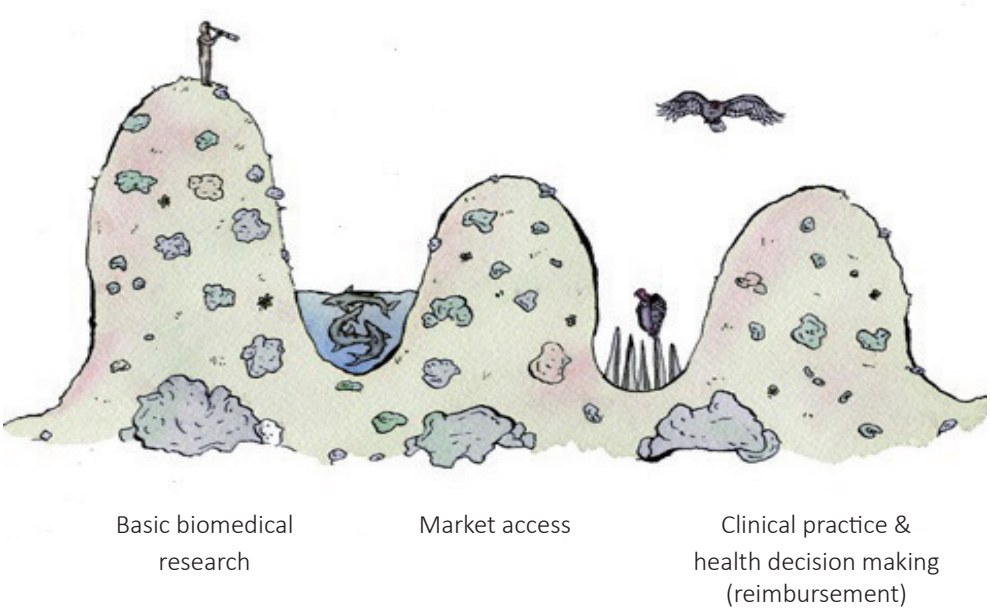


Figure 1. Visualization of the valleys of death. Adapted from Erinn Acland, 2016.

Although calculating reliable failure rates is difficult, many sources speculate that a large percentage of medical devices or pharmaceuticals fail on their way to market launch[7]. In preventing this failure it has been stressed to comprehensively and iteratively evaluate the innovation during its development[6,8]. This evaluation may identify implementation and/or diffusion barriers, based on which actions can be taken. Additionally, it may show that the technology is not sufficiently effective in the targeted population, necessitating to choose a better target population or to evaluate how to improve the effectiveness of the technology.
A proposed method to comprehensively and systematically evaluate these technologies alongside the development process is early health technology assessment (early HTA)[9]. Although the application of early HTA or the iterative use of HTA alongside the product development process has been described in the mid-1990s[9], guidance is still required on its application aiming to inform research and development (R&D) and clinical- and policy-decisions especially in the evaluation of medical devices.

The aim of this dissertation is to contribute to the knowledge on the application of early and mainstream HTA methodologies alongside the translational pathway of medical technologies, aiming to support R&D, and clinical- and policy decision-making.
This dissertation can serve as a start to position early HTA in the comprehensive evaluation of medical technologies during the translational pathway to facilitate effective innovation and adoption.

This chapter starts with further exploring the translational pathway and potential barriers along its way. Second, mainstream HTA, early HTA, and very early HTA and their use alongside the translation pathway are described, which is followed by an introduction of the complex innovations that we target in this dissertation, and the chosen HTA methods. Finally, the research scope and the outline of this dissertation are given.

1

Translational pathway

The path from having a conceptual idea to the actual use of a technology in clinical practice is referred to as the translational pathway, which takes on average 12 to 15 years[10]. This translational pathway is best conceptualized by the phases of translational research. Although many definitions exist, all point in a similar direction; the process starts with phase 1 (T1) in which ideas are brought from basic research to initial testing in humans (pilot studies (clinical trial phase I)), phase 2 (T2) entails the performance of early phase clinical trials (clinical trial phase II). Phase 3 (T3) focusses on implementation and dissemination. In the process between phase 2 and 3, the first "Valley of death" has been recognized. Phase 4 (T4) focusses on outcomes and effectiveness research (clinical trial phase III), where the second "Valley of death" looms. Finally, phase 5 (T0) involves research to create new ideas based on the effectiveness results (e.g. biomarker development when specific subgroups show improved outcomes).[11] This process can be further specified towards specific types of innovation (e.g. a pharmaceutical product, a medical technology).

Since this dissertation mainly focusses on medical technology, a more detailed process focusing on medical technology was used (Figure 2). IJzerman and Steuten described the translational pathway of a medical product as follows: it starts with basic research followed by a proof of principle and further product development. Afterwards, the phase I, II, and III clinical studies will start, aiming to obtain coverage, resulting in the adoption of the technology[6]. In this process, we define adoption as: "The choice to acquire and use a new invention or innovation."[12].

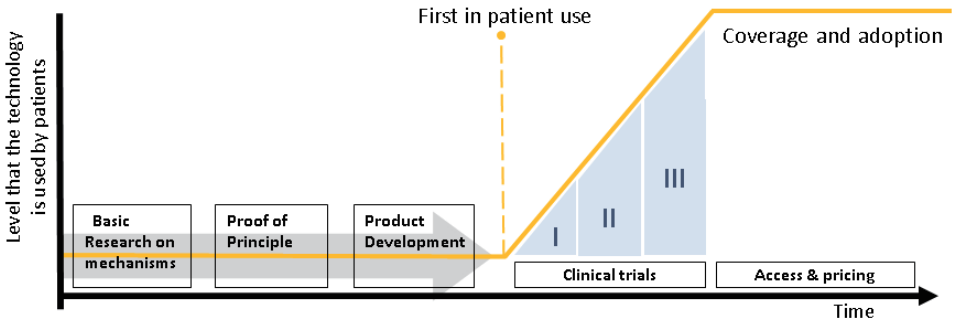


Figure 2. Translational pathway for medical technologies (product life cycle). The translational pathway that is used in this dissertation to identify the different phases of the research and development (R&D) process of complex innovations. Adapted from IJzerman and Steuten[6].

Implementation barriers

Alongside this translational pathway, many barriers can appear that hamper market access and diffusion. In this context, diffusion is described as: "The process by which an innovation is communicated through certain channels over time among the members of a social system."[13]. Implementation research by Rogers, Cain and Mittman identified 12 key attributes that influence the diffusion of a specific technology: (1) relative advantage, (2) compatibility, (3) complexity, (4) trialability, (5) communication channels, (6) homophilous groups (within a group, the tendency of individuals to associate with each other), (7) the pace of innovation/ reinvention, (8) observability, (9) norms, (10) roles and social networks, (11) opinion leaders, and (12) infrastructure[13,14]. For the promising innovations fulfilling the relative advantage attribute that fail to come to the market, one or multiple of the other attributes seem to hamper diffusion. Therefore, these attributes should be incorporated when comprehensively evaluating a complex innovation.

Health technology assessment (HTA)

Health technology assessment (HTA), as mentioned before, has been proposed as a tool to comprehensively evaluate new technologies, aiming to bridge the gap between research and medical decision-making. It aims to systematically evaluate various aspects of new interventions such as medical, economical, organizational, social & patient-related, demographical, and ethical & legal aspects (Figure 3)[15]. A variety of methods, both quantitative and qualitative, are used to evaluate these different aspects.

Generally, HTA is used in mature technologies that proved their safety and effectiveness, to inform pricing and reimbursement decisions by performing cost-effectiveness analyses (CEA)[16,17]. The other aspects (such as organizational) are often left out of the scope of the analysis. In this dissertation, similar to Miquel-Cases et al, we refer to this application of HTA (timing and focus) as "mainstream HTA"[18]. HTA could however be introduced earlier in the process to effectively guide product development, and as mentioned before, prevent failure of a technology. For example, a similar analysis may be performed (e.g. cost-effectiveness analysis) for a technology that was just developed but depending on the stage of development and the data available, the aim of the HTA analysis differs. This iterative use of HTA – performing an early CEA with limited data and after a certain amount of time update this CEA with the most recent data – has been proposed to progressively generate firmer estimates of the cost-effectiveness[19].

1

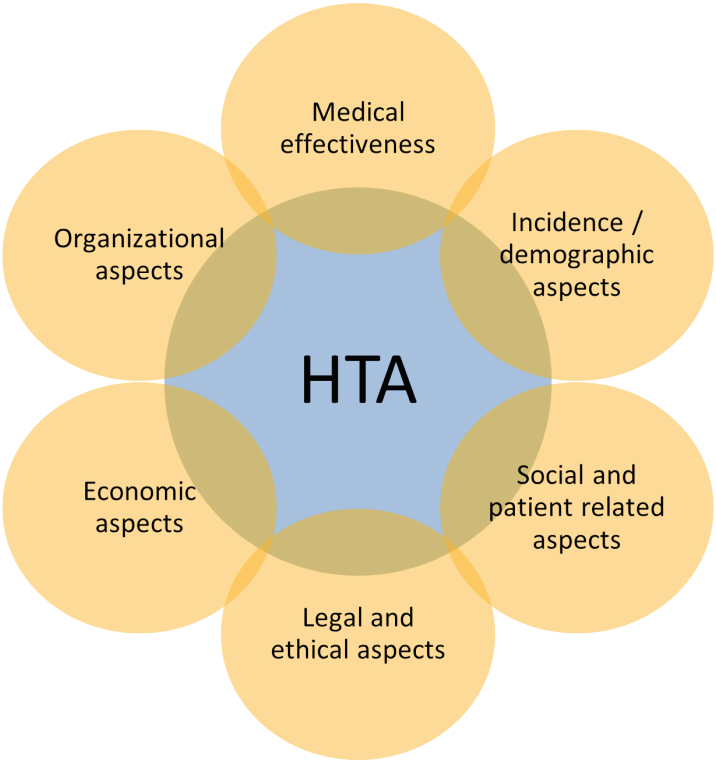


Figure 3. Graphical illustration of the aspects involved in Health Technology Assessment.

Figure 4 presents the three phases of HTA in the translational pathway: "very early HTA", "early HTA", and "mainstream HTA". Furthermore, it lists the potential specific HTA methods per phase based on Miquel-Cases et al. 2017 and Markiewicz et al 2014[18,20]. Depending on the moment of using HTA, a different aim may be applicable. For example, in a proof of principle stage, the aim of performing an HTA is to evaluate whether it is valuable to continue with further validation studies and if they continue, it could aim to inform the future study design. Performing an HTA after a phase II clinical trial may aim to evaluate which characteristics the subsequent study design should have[18]. Finally, "mainstream HTA" can inform decision-makers whether the technology is cost-effective, and thus whether or not the technology should be included in the insurance package. Figure 4 also provides the potential aims for using HTA per translational phase. In the sections below, the three phases of HTA are further explored, starting with "mainstream HTA".

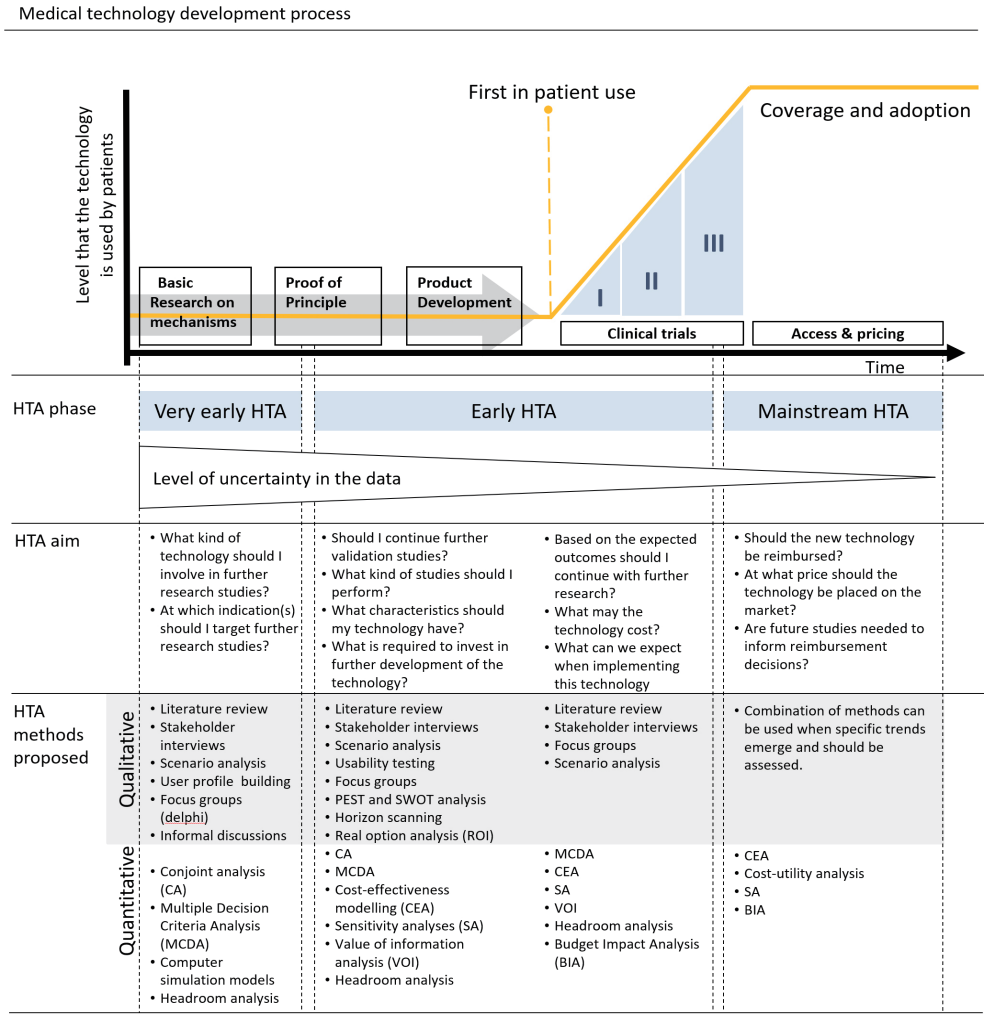


Figure 4. The translational pathway including the HTA phases and proposed HTA methods.

1

Mainstream HTA

"Mainstream HTA" aims to inform reimbursement decisions after a new technology showed its effectiveness and safety. To inform these decisions, mostly a CEA is used in which the incremental costs and incremental effects of a new technology are assessed compared to the current standard of care. This evaluation results in the incremental cost-effectiveness ratio (ICER) (Δ costs / Δ effects), representing the additional amount of money needed per patient to obtain an additional life-year or an additional life-year in perfect health. The latter is described as a quality adjusted life-year (QALY). Based on the results, decision-makers use a certain willingness-to-pay threshold, describing the maximum value that a society is willing to pay for a QALY, to decide to either adopt or reject an innovation[21,22]. These willingness-to-pay thresholds differ per country[22] and may also differ based on the severity of a disease[23], therefore these HTA analyses differ per country and indication.

"Mainstream HTA" has especially been introduced and conducted as part of the authorization process in the assessment of pharmaceuticals. The process of authorization by institutes such as the European medicines agency (EMA) or the US food and drug administration (FDA) has been in place for decades (e.g. EMA since 1995[24]). After authorization of a medicine (EMA or FDA), national and regional authorities have to decide on pricing and reimbursement, these decisions are informed by a CEA often complemented with a budget impact analysis (BIA) to estimate the financial consequences of adoption and diffusion for a specific health-care setting[25]. Until recently, the process of authorization for medical technologies was less strict compared to that of pharmaceuticals. Consequently, in medical technologies, HTA was considerably less applied. In Europe, a new law has been introduced that describes a more formal process for approval of medical technologies[26], which has been delayed for one year due to the Corona pandemic in 2020. This law states that medical technology, including developed software, is required to be verified by a notified body before authorization, which requests more robust clinical evidence and confirmation with high-quality standards, especially in high-risk medical devices. Following these developments, we can expect that HTA and HTA bodies will play a more central role in the process of market authorization and market access of medical technologies.

Early HTA

"Early HTA" covers the actual development process and first clinical trials (phase I and II) of medical technologies. It aims to guide the development and implementation process and inform the design of the first clinical studies[6,9,18,27]. Over the last decade, the value of early HTA is being increasingly acknowledged[28,29]. Especially because it could avoid unnecessary delay in patient access, as an early CEA can directly be updated with the most recent clinical

data (iterative approach[8]). The duration of the process between the publication of the efficacy results and presenting its cost-effectiveness could therefore be shortened.

Besides these advantages of early HTA, one considerable complication is the higher level of uncertainty observed in the data, especially when its use aims to inform policy decisions. This uncertainty is a direct result of the earlier stage of development, for example clinical data originates from expert opinions or pilot studies. In interpreting the results of early analyses, having information on the magnitude of uncertainty and its influence is crucial. This requires specific methodologies, where for example a value of information (VOI) analysis could be used to evaluate the expected value of perfect information, which indicates the maximum investment required to obtain perfect information by performing further research. These analyses can also be used to evaluate which parameters have the highest degree of uncertainty steering the design of future clinical studies. Furthermore, headroom analyses can be used to evaluate the maximum incremental costs for a new technology in a certain indication to inform price setting and design choices. In addition to such quantitative methods, qualitative methods as focus groups, stakeholder interviews, scenario drafting, and, usability testing can be used to support decision making on the further developments and its future implementation. Which specific method or methods are most suitable depend on the phase of the technology and the questions the researchers have (Figure 4).

Type of early HTA: Constructive Technology Assessment

Constructive technology assessment (CTA) is a variant of "early HTA". It has its origin in public policy and technical industry to inform technological development before and during the introduction of the technology[30]. It was introduced to follow and influence the course of technical development and the diffusion of a technology, having a continuous and/or iterative nature as choices about the technology are continuously made.
When applied in healthcare, four main domains should be evaluated: Clinical, Patient-related, Economic, and Organizational[31]. These domains contain several aspects of interest derived from the definition of quality of care from the Institute of Medicine[32], research performed by Poulsen[33], and especially in the organizational domain aspects related to diffusion scenarios by Rogers[13]. The aspects that should be evaluated can differ per innovation and phase of the innovation, as it should focus mainly on aspects that are expected to change over time. To evaluate these aspects, a set of mixed methods can be used, both qualitatively and quantitatively.

1

Early HTA in the formal adoption process

The increasing demand for early access to new (promising) treatments for patients with a high unmet clinical need resulted in the introduction of Managed Entry Agreements (MEA), which can be seen as an application of early HTA in the formal authorization process. MEAs can be defined as "Any agreement between a manufacturer and payer/provider that enables access to a health technology subject to certain conditions."[34]. An example of an MEA is a coverage with evidence development (CED) program, in which a technology is reimbursed for a limited period with a specific requirement to collect further evidence[35–37]. These programs focus mainly on innovations or medical technologies, very rarely on pharmaceuticals, which are expected to result in improved outcomes at high incremental costs[36]. Based on the gathered evidence, a decision needs to be taken on formal reimbursement after conditional reimbursement. To inform this decision, HTA with at least a CEA is a standard component of a CED program.

Very early HTA

"Very early HTA" is used at the initial phase of the development process aiming to support fundamental decisions such as choosing the target population and most optimal technology. As very little data will be available in this early phase, the recommended HTA methods are mainly qualitative (e.g. interviews and focus groups). Some quantitative methods, similar to the ones in early HTA, can be used but are then often informed by expert estimates obtained via expert elicitation or literature. Hilgerink et al described an example of a very early analysis, where the authors evaluated the potential clinical value of a new tool to diagnose breast cancer: photoacoustic imaging[38]. This analysis showed that photoacoustic imaging could substitute the combined use of x-ray mammography and ultrasonography in early breast cancer diagnosis as it is preferred by an expert panel. Furthermore, the analysis identified some areas that would benefit further development, for example, the sensitivity of the detector, the bandwidth, and the number of wavelengths used.

(Complex) innovations

In this dissertation, several (complex) innovations have been studied with HTA methods aiming at facilitating their translational process. These innovations are introduced shortly in this paragraph. Figure 5 presents an overview of the included innovations and their place on the translational pathway. Some innovations are placed in several stages of development (e.g. optical imaging), because they moved on the pathway within the duration of this PhD project, and for some innovations, the development phase differs per targeted indication (e.g. navigated surgery).

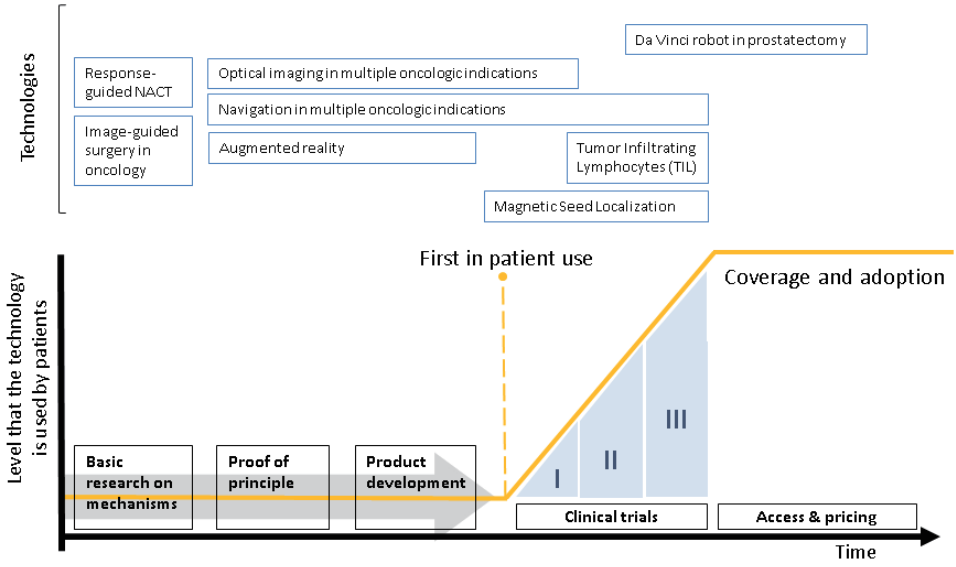


Figure 5. Visualization of the (complex) innovations alongside the translational pathway.

Advanced therapy medicinal products and tumor infiltrating lymphocytes (TIL)

Advanced therapy medicinal products (ATMPs) are currently one of the most promising, personalized strategies for cancer treatment[39]. These products are described as: "medicines for human use that are based on genes, tissues or cells"[40]. They have to comply with the Good Manufacturing Practices (GMP) guideline (2003/94/EC)[41] that translates into a requirement for a solid quality system, suitable investments, and effective logistical preparation. Other than generic pharmaceuticals, ATMPs are often produced by small academic centers that complicate market access due to for example required high upfront investments[42]. Therapy with tumor-infiltrating lymphocytes (TIL) is an example of an ATMP. Promising results of this therapy in advanced melanoma have been shown for the first time in 1988 by Rosenberg et al. at the National Cancer Institute[43]. Hereafter, multiple phase I and II studies have been conducted showing comparable results in treating melanoma. However, to date, TIL-therapy has not been implemented in standard clinical practice. In the Netherlands, TIL-therapy has been introduced at the Netherlands Cancer Institute – Antoni van Leeuwenhoek hospital (NKI-AvL) since 2011. Based on promising results found in a phase II study in the NKI-AVL, TIL-therapy has been included in a CED program since 2015[44]. In this program, the (cost-) effectiveness of TIL-therapy compared to ipilimumab in stage IIIC and IV melanoma is being evaluated in a phase III randomized controlled trial (RCT), while the treatment is conditionally reimbursed by the government (NCT02278887). This study is conducted in collaboration with the Herlev hospital in Denmark and is the first RCT comparing TIL-therapy to another

immunotherapy (ipilimumab). At the beginning of 2021, patient inclusion for the study will be closed, after which it will be evaluated whether TIL-therapy showed significantly improved progression-free survival at 6 months compared to ipilimumab. Based on the effectiveness and cost-effectiveness results, the National Healthcare Institute (Zorginstituut Nederland) will decide whether or not TIL-therapy should be reimbursed in the Netherlands.

(Response-guided) Neoadjuvant chemotherapy in breast cancer patients

Neoadjuvant chemotherapy is a chemotherapy regimen provided before surgery to for example enable breast-conserving surgery instead of mastectomy. Response-guided neoadjuvant chemotherapy (NACT) means that the NACT treatment is guided by early therapeutic response monitoring[45]. Under this scenario, treatment response is measured after a specific number of NACT cycles, and according to this response, further systematic treatment is tailored, i.e. responders continue with the same initial treatment, and non-responders may switch to a presumably non-cross-resistant regimen. In 2013, von Minckwitz et al. published results from the GeparTrio trial in which the authors suggest that response-guided neoadjuvant chemotherapy in hormone receptor positive breast cancer patients might improve survival[45]. The question was whether imaging technologies such as MRI and ^{18}FDG-PET/CT could be used for accurate response monitoring to subsequently change the treatment regimen based on this monitoring.

Currently (approximately 4 years after our research was performed), clinical research is less focused on chancing treatment regimen based on response on NACT, but evaluates whether surgery can be omitted in patients showing complete pathologic response after NACT[46].

Innovative technologies in surgery

In surgery, minimally invasive technologies emerged over the last couple of decades[47]. The adoption of new surgical tools usually takes place without reimbursement consequences, although they often come with additional costs; which lowers the amount of money that a hospital could use for quality improvements. As surgery already accounts for a large part of the annual healthcare costs, new surgical technologies have to prove themselves in terms of value for money[48,49]. This also requests a formal and structured evaluation of the effectiveness which is complicated by factors as learning curves, user interactions, and frequent product modifications[50,51]; factors that are often not assessed in standard efficacy studies. This results in a need for alternative trial designs such as large cohort studies, and case-matching studies[52]. In the following paragraphs, several innovative surgical tools are presented which were studied in this dissertation.

- *Robot-assisted surgery*

In 2000, the Da Vinci® (Intuitive Surgical) robot aiming to support minimally invasive surgery, was launched and approved by the FDA. By using this robot technology, the surgeon is expected to operate more precisely due to better sight (3D) and a greater range of motion. In the operating room, the robot is placed close to the patient with the robot arms above the patient. The robotic arms are controlled by a surgical console (controlled by the surgeon) which is standing in the corner of the operating room. An operating assistant is standing next to the patient and the robot to assist the surgeon by for example cleaning the camera lens or supplying suture materials. When the surgical console is connected to a different Da Vinci robot standing somewhere else in the world, the surgeon can even operate over distance. This robot technology can be used in multiple indications, for instance, gynecology or cardiothoracic surgery, but it was mostly adopted to perform a radical prostatectomy in localized prostate cancer patients[53].

After radical prostatectomy, patients have often complaints of incontinence and erectile dysfunction[54,55]. By using robot technology these complaints may be reduced due to limiting the damage on important nerves. In the Netherlands, the Da Vinci robot was increasingly introduced in prostate cancer surgery since 2003[56], and currently, radical prostatectomies are mostly performed by using this robot. Although many studies compared robot-assisted prostatectomy (RARP) to open and laparoscopic prostatectomy on oncologic and functional outcomes (most studies were performed between 2010 and 2018), the evidence base in favor of RARP was judged too low to receive additional reimbursement[57–59]. Therefore, hospitals are faced with substantial additional costs per patient operated with the Da Vinci robot. After multiple years of experience with RARP in the Netherlands, the question remains whether RARP results in improved functional outcomes compared to its standard of practice, laparoscopic prostatectomy.

- *Image-guided surgery*

In the surgical field, one of the most recent developments is the use of advanced medical imaging technologies during surgery aiming to improve surgical performance, described as image-guided surgery (IGS). Technologies as a CT scanner or an MRI are integrated into or located close to an operating suite that is described as a hybrid operating room. The hypothesized benefit of IGS is an improvement in surgical performance aimed at improving patient outcomes and/or reducing the risk of secondary procedures because it enables intraoperative evaluation of surgical success (e.g., surgical margin status)[60–64]. Currently, IGS has mainly been introduced in trauma interventions and vascular or cardiac surgeries, but it is also expected to be of use in the oncologic field. However, for application in the oncological field, the available imaging techniques in the operating room (C-arm CBCT, Ultrasound, and MRI) have a limited ability to accurately visualize the tumor, its boundaries, and its critical

surroundings. Therefore, medical device companies and several research groups seek for new technical solutions that could provide the image guidance needed in oncologic surgery (e.g., probes, navigation technologies). Since 2014, a specific research group at the Netherlands Cancer Institute aimed to develop surgical tools that would improve surgical performance. Below, the technologies that are evaluated in this dissertation are shortly introduced.

- *Navigated surgery*

The navigation technology developed to use during surgery provides a 3D map of the anatomical area where the surgeon has to operate in, showing the tumor and some of the critical structures that are aimed to spare[65]. This 3D anatomical map is created per patient prior to surgery by image segmentation. In the operating room (OR) this 3D anatomical map is calibrated to the situation in the OR to make sure that the 3D map corresponds to the patient's position in the OR. By using electromagnetic technology, a pointer can be tracked in the surgical field, which can be used to navigate on the 3D anatomical map to the locations of interest. Such technologies were already applied in neurosurgery and other anatomical areas where a fixed bony structure was present, but this navigation system is unique as it also could be used in less fixed anatomical structures. Currently, this technology has been tested in clinical pilot studies in removing breast cancer tumors, colorectal cancer tumors, liver tumors, and lymph nodes[66–68]. Presently, the research group aims to implement this technology in a second hospital in the Netherlands.

- *Optical imaging*

Optical imaging is a summary term for tools that use the characteristics of tissue regarding scattering and absorption of light to detect whether tissue is malign or benign. For example, fatty tissue, often benign, will show a different pattern of scattering and absorption of light than tissue that contains especially water, which is potentially malign. Two main tools are being developed in the NKI, a point measurement tool, and a camera. The point measurement tool can be used during surgery- ideally, this tool would be incorporated in a surgical knife- to check whether a surgeon is resecting in benign or malign tissue. The camera can be used after resection to check whether the resection borders of the resected tissue are free of tumor. Both aim to improve negative surgical margin rates and reduce the risk of local recurrence. It also could decrease the total operating time by obviating pathologic assessment of the resected tissue before ending the surgical procedure. The application of the optical imaging tools has been evaluated (ex-vivo and in-vivo) in resecting breast cancer tumors, colorectal tumors, and head and neck tumors[69–71].

- *Augmented reality*

A definition of augmented reality is: "An enhanced version of reality created by the use of technology to overlay digital information on an image of something being viewed through

a device (such as a smartphone camera)."[72]. At the Netherlands Cancer Institute, this technology is currently still in development and not yet applied in pilot studies[73]. The concept of this technology is that the 3D anatomical map, similar to the one used for navigation is visualized via a Google Hololens or a tablet. This prevents that a surgeon has to switch from the surgical field to the display and back, making the surgery more convenient and potentially more efficient and accurate. This technology could also add value to the planning process of surgery as a surgeon can have a detailed 3D visualization of the targeted area.

- *Magnetic seed localization*

In 2016, the NKI-AVL developed a new localization technology to detect non-palpable breast cancer lesions during surgery using a magnetic seed (MSL)[74]. This technology aims to overcome disadvantages of current localization technologies such as unfavorable incision placement when wire-guided localization (WGL) is used[75,76] and adhering strictly to nuclear safety regulations when using radioactive seed localization (RSL)[77,78]. In principle, MSL is similar to RSL but instead of a radioactive seed, a magnetic seed is used. Using magnetism instead of radioactivity is expected to ease the implementation process and the workflow because no strict regulations need to be followed related to radioactivity.
In using MSL, a magnetic seed is placed in the lesion by the radiologist. Intraoperatively, a magnetic probe providing constant feedback on the location of the seed is used by the surgeon to guide resection of the tumor[74,79]. In 2017 MSL showed to be feasible and safe in detecting non-palpable breast cancer lesions[74]. In 2019 a different group showed that using magnetic localization techniques was as effective as WGL[80]. Currently, Sirius Medical, a company in the Netherlands, aims to launch a commercial version of MSL soon. At the end of 2019 they received the ISO13485:2016 certification, which indicates that the MSL agreed with the strict medical device quality controls by a notified body.

Chosen HTA methodologies

In this dissertation, to inform further development or inform reimbursement decisions of these (complex) innovations, several HTA methods were used. These are described in the section below.

Multiple criteria decision analysis

In healthcare decision making, the perceived added value of a new technology may be described by various attributes. It is therefore important to take all relevant criteria into account, especially in an early phase of development. Multiple criteria decision analysis (MCDA) is a tool to systematically evaluate the importance of the criteria in a decision problem[81]. Additionally, based on the first data or even expert opinions, MCDA also contains methods to evaluate the performance of new technologies on the determined set of decision

criteria. MCDA can therefore be used to identify the attributes for a successful innovation and estimate the expected value of certain innovations. In this dissertation, MCDA was used very early in the development process to prioritize further development and clinical studies of the image-guided surgical technologies. Multiple specific tools are available to perform an MCDA (e.g. MACBETH, Promethee). In our analysis, the Analytic Hierarchical Process (AHP) was selected, because of its intuitiveness and simplicity[82,83].

Cost analysis

Gaining insight into the expected costs of the innovative technology and/or the potential process deviations by using the technology is valuable to researchers, as these costs will affect the final cost-effectiveness of the intervention. A bottom-up cost evaluation by performing an activity-based costing (ABC) analysis is a method evaluating all activities that are performed for a specific intervention, including the need for personnel, the time consumed for a specific activity, and materials used[84,85]. Bottom-up analyses are more transparent and detailed than top-down cost evaluations in which annual expenses per hospital or department are used[86], and increase the generalizability of cost parameters used in cost-effectiveness analyses. In this dissertation we preformed multiple cost analyses, both in (very) early HTA and mainstream HTA; for example the costs of the conventional and hybrid OR, TIL-therapy, the navigation system, and performing a robot-assisted and laparoscopic prostatectomy were evaluated.

Budget impact analysis

A BIA evaluates the financial impact of the adoption of a new technology from a budget holders perspective. Often these analyses are performed on a national level. Input parameters are the total population, the sick population, the targeted population, resource utilization, and costs of illness evaluated for the current environment and estimated for the new environment[25,87]. The difference between the current and new environment is referred to as the budget impact. A BIA is meant to complement a CEA, not to replace it. When on an individual basis only small effects are expected between the interventions, but larger effects are expected on a hospital or national level, a BIA may provide more valuable information than a CEA. In this dissertation, an early BIA was performed for the adoption of MSL in the Netherlands, incorporating implementation costs.

Scenario drafting

Scenarios are used to explore future developments by questioning: “What if?”. Shell global and many other large commercial companies use this method to explore future developments and make informed decisions. Often a panel discussion is held incorporating relevant stakeholders to evaluate all potential dynamics (e.g. a pandemic). In HTA, scenario drafting can be used in a similar way to identify future developments surrounding a specific disease area or a (disruptive) new technology to identify potential barriers and facilitators[88,89].

Incorporation of both believers and non-believers in these evaluations is crucial to obtain a comprehensive overview of perspectives. In a previous study, a scenario analysis identified that providing education for surgeons and obtaining high-level evidence supporting the use of Next-Generation Sequencing (NGS) in the clinic are crucial factors for the adoption of NGS[90]. By identification of this type of information, actions can be taken upon, for example developing an educational plan to facilitate adoption in a later stage. In this dissertation, alongside a clinical trial (early HTA), scenarios were drafted for the adoption of TIL-therapy to identify barriers and facilitators, aiming to facilitate implementation.

Cost-utility and cost-effectiveness analysis

In a CEA the incremental costs are divided by the incremental outcomes resulting in the ICER (incremental cost-effectiveness ratio). In literature, the definitions CEA and cost-utility analysis (CUA) are used interchangeably. However, officially, a CEA evaluates the costs in relation to a natural effect, for example, the avoidance of 1 day feeling depressed, where a CUA evaluates the costs related to a QALY gained. CUAs are used in reimbursement decisions, however, CEAs may provide valuable information in addition to a CUA. CEAs are especially of interest when the intervention brings small improvements on disease-specific complaints, for which the (generic) utility measurement is often not sensitive enough.

Based on literature, available clinical data, historical databases, internal cost data and/or expert elicitation the parameters related to costs, effects, and quality of life are estimated and included in the CEA[91,92]. These parameters may be uncertain, especially in case of an early evaluation. Therefore, each of the parameters receives a distribution surrounding the observed value based on the data and parameter characteristics. The impact of uncertainty in the data is evaluated by simulating a hypothetical cohort of 1000 patients and using a Monte Carlo simulation to draw multiple times (often 1000) a random number from the parameter distribution[93]. This results in 1000 (or more) possible outcomes that are plotted in a cost-effectiveness (CE) plane (Figure 6). On the x-axis the incremental effects are visualized and on the y-axis the incremental costs. When a technology is more effective and less expensive (the lower right quadrant), the new technology should be adopted. When a technology is more effective, but more expensive, adoption depends on a national maximum acceptable ICER or willingness to pay. A similar pay-off is made in the lower left quadrant presenting a technology that is less effective and less expensive. In some cases, below a certain threshold, such a technology may be adopted. Finally, a technology showing higher costs and lower effectiveness is rejected (upper left quadrant). In this dissertation, all cost-effectiveness analyses are performed with CUAs. An early CUA was performed for the navigation technology based on the first clinical data in colorectal cancer patients (phase II study). Besides, a mainstream CUA was performed for the use of the Da Vinci robot in radical prostatectomy based on a large retrospective cohort study.

1

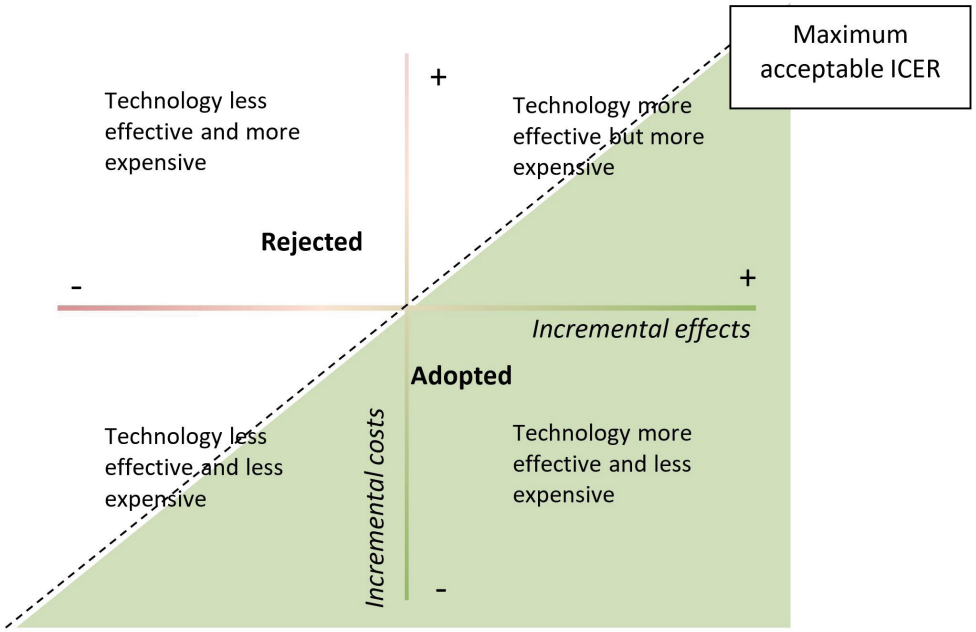


Figure 6. The cost-effectiveness plane including the explanation of each of the quadrants.

Value of information analysis

The results of an early CEA are surrounded by uncertainty due to the limited available data. To support decisions on further research and inform policy-makers, the value of research could be evaluated with a value of information analysis (VOI)[94], this indicates the maximum amount a decision-maker would be willing to spend on further research to obtain perfect information (evaluating the expected value of perfect information (EVPI)). If, based on the EVPI, further research appears to be worthwhile; an expected value of perfect parameter information (EVPPI) analysis can provide insight into the parameters that are most valuable to perform further research on. As a final step, the value of performing research with a specific sample size and a particular design can be evaluated with an expected value of sampling information analysis (EVSI). In the early CUA comparing navigated surgery to standard surgery, an EVPI was performed to inform potential future research activities.

Aims and outline of this dissertation

As described at the beginning of the chapter, this dissertation aims to contribute to the knowledge on the application of early and mainstream HTA methodologies alongside the translational pathway of medical innovations. By doing so, this dissertation serves as a start to position early HTA in the comprehensive evaluation of medical technologies. The research objectives are:

1. To inform multiple stakeholders (e.g. researchers, clinicians, and policy-makers) on further research and development, implementation decisions, and reimbursement decisions to stimulate effective innovation of the technologies of interest.
2. To contribute to the knowledge on the application of early and mainstream HTA methods alongside the translational pathway of medical innovations, by providing real-life examples.
3. Based on real-world examples, to serve as a start, to position early HTA in the comprehensive evaluation of medical innovations during the translational pathway.

The dissertation consists of four parts, distinguished by the phases of HTA and the HTA methods that were identified alongside the translational pathway: “very early HTA” (chapter 2 and 3), “early HTA: up to and including the first clinical studies (phase I)” (chapter 4 to 6), “early HTA: both phase I/II studies” (chapter 7 and 8) and “mainstream HTA” (chapter 9 and 10).

Part I: Very early HTA

Chapter 2 presents the results of a very early evaluation of three image-guided surgery techniques (navigation, optical imaging, and augmented reality) that could be used in five oncologic indications. Based on the EUnetHTA core model, a framework of decision criteria was created determining the success of five oncologic surgical procedures (removal of tongue, breast, rectal, and liver tumors, and lymph node dissections). Employing an AHP analysis, the importance of the decision criteria were evaluated with 18 surgeons in individual face-to-face interviews. Together with the technical developers, the three image-guided surgical technologies were described to enable evaluation of the proposed technologies on the decision criteria. In the second round of interviews, the technologies were presented and using AHP, the expected value of each of the technologies was estimated per indication. Combining the score of the importance of the criteria and the expected performance of a technology on the criteria, the expected value was estimated to steer further R&D activities and inform future clinical trial designs.

Chapter 3 provides an overview of the current literature on the accuracy of using imaging

techniques to monitor the response after NACT. A systematic literature review was performed to evaluate the imaging performance of measuring the response on NACT per breast cancer subtype, as previous studies suggested that imaging performance was dependent on breast cancer subtype. Studies were independently selected based on title and abstract by two researchers. The results of the analysis aimed to inform the further development and testing of the response-guided NACT approach.

Part II: Early HTA: up to and including the first clinical studies (phase I)

In **chapter 4**, the implementation and intervention costs of using magnetic seed localization (MSL) and radioactive seed localization (RSL), and the intervention costs of wire-guided localization (WGL) in breast cancer surgery were evaluated by a bottom-up cost analysis in eight Dutch hospitals. Using the costs of RSL and WGL and the expected costs of MSL, an early BIA adopting a time horizon of 5 years was performed for the gradual implementation of MSL in the Netherlands. As the costs of the magnetic seed were still to be determined, these were included as a range (€100-€500). To guide R&D and inform price-setting decisions, the maximum price of the magnetic seed used in MSL to be cost-efficient was evaluated.

Costs of using an operating room (OR) per minute are, although some attempts based on expense data, unclear and uncertain. In **chapter 5**, we present the costs of a conventional and hybrid OR based on a bottom-up costing analysis in five Dutch hospitals. The cost analysis incorporated the following cost components: construction, personnel, inventory, and overhead. The construction costs were based on key numbers presented by the Dutch advisory board on healthcare housing for building costs in a hospital and the average surface of the ORs. Inventory costs were based on acquisition costs, personnel costs were based on collective labor agreements assuming a standard operating team available during surgery. Overhead costs were calculated according to the Dutch guideline for cost analyses. This chapter aimed to provide insight into the cost drivers of both ORs to inform optimization strategies. Furthermore, these costs can be used in future cost-effectiveness analyses evaluating the added value of interventions in the hybrid OR.

In a clinical study performed in the NKI-AVL, navigated surgery showed improved negative resection margin rates in locally advanced (LARC) and locally recurrent rectal cancer patients (LRRC) compared to standard surgery. In **chapter 6**, we estimate the expected cost-effectiveness of using the image-guided navigation system in LARC and LRRC based on the first clinical results. For this analysis, a Markov model was created evaluating costs and effects over a time horizon of 3 years, containing three health states: Disease-free, Progression, and Death. A deterministic and probabilistic sensitivity analysis was performed to evaluate the impact of uncertainty surrounding the parameters. In a scenario analysis, we evaluated a

situation in which the navigation system would be used optimally, and we evaluated a situation where a hospital has to invest in a hybrid OR before navigated surgery could be introduced in the clinic. Finally, a VOI analysis was performed to evaluate whether it is valuable to invest in further research and inform future research designs. The analysis aimed to inform the researchers on conditions for the cost-effectiveness of navigated surgery to steer further R&D and clinical study designs. Furthermore, the analysis aimed to inform policy-makers or insurance companies on the expected cost-effectiveness, potentially resulting in conditional approval or reimbursement.

Part III: Early HTA: both phase I/II studies

In **chapter 7** we present the potential barriers and facilitators of implementation of ATMPs and specifically for the implementation of TIL-therapy. A CTA was performed to evaluate six relevant domains: clinical, economic, patient-related, organizational, technical, and future among patients, clinicians, and other stakeholders. A set of mixed methods was used including a bottom-up cost analysis, semi-structured interviews with all stakeholders of the TIL-therapy process, a questionnaire among patients eligible for the TIL-therapy study, and a non-systematic literature review evaluating the literature on the implementation of ATMPs. This chapter aimed to guide the implementation of TIL-therapy by informing other research groups on potential barriers and facilitators. Furthermore, the results could inform policy-makers on the complexity of the evaluation of TIL-therapy, potentially resulting in a more flexible coverage with evidence program.

Based on the barriers, facilitators, and expected future directions presented in chapter 7, **chapter 8** presents 14 potential adoption scenarios and their influence on the early cost-effectiveness compared to ipilimumab. The scenarios were drafted based on the results presented in chapter 7, and discussed with the internal expert and research group of the TIL-therapy study. A web-based survey was created to evaluate the likelihood of the scenarios within the coming 5 years among international experts. The survey was distributed via the network of our internal experts, and promoted at a conference. Based on the average likelihood, recent literature, and consensus within the internal expert group, the likely scenarios were chosen to incorporate in an existing cost-effectiveness model[95]. The analysis aimed to inform further R&D as it points to barriers that would decrease the chance of TIL-therapy being cost-effective. Additionally, this analysis aimed to inform decision-makers on the magnitude of uncertainty surrounding this therapy.

1

Part IV: Mainstream HTA

Although in the Netherlands almost every hospital performing radical prostatectomies use robot assistance, although RARP has not been recommended in the European guidelines. **Chapter 9** presents the results from a large retrospective cluster study evaluating the long-term functional outcomes of 1370 prostate cancer patients operated in the Netherlands. Patients were included from 12 Dutch hospitals undergoing either RARP or Laparoscopic Radical Prostatectomy (LRP). 2117 of the 2626 patients that underwent a prostatectomy in the 12 selected hospitals between 2010 and 2012 were invited with an invitation letter accompanied by a questionnaire consisting of multiple standardized questionnaires (e.g. EQ5D, EORTC C30, EPIC26). Of all participants, clinical, pathological, and preoperative parameters were extracted from the medical record. Primary outcomes were the scores on the urinary and sexual domains of the Extended Prostate Cancer Index Composite (EPIC-26). Using a mixed modeling approach, differences between the groups were identified. Additionally, a regression analysis was performed evaluating the impact of hospital volume, age, receiving radiotherapy, and receiving a nerve-sparing procedure on the urinary and sexual domain scores. This chapter shows the first comparison of long-term functional outcomes of RARP and LRP internationally and the first large analysis based on real-world data from the Netherlands. This analysis aimed to inform guidelines and potentially reimbursement decisions.

Chapter 10 presents the cost-effectiveness of RARP compared to LRP based on the results presented in chapter 9. The costs and effects were evaluated over 7.08 years, following the median follow-up period of the clinical study using a decision tree incorporating three health states: being incontinent and impotent, being continent and impotent, and being continent and potent. The intervention costs were estimated by a bottom-up costing analysis performed in five of the twelve participating hospitals. The costs for additional care used when being continent and potent were based on the results presented in chapter 9. Based on a survey distributed among the surgeons that operated between 2010 and 2012 in the selected hospitals, ergonomic effects were evaluated and incorporated by linkage to potential productivity losses. A deterministic and probabilistic sensitivity analysis was performed to evaluate the uncertainty surrounding the input parameters. In a scenario analysis, various scenarios of centralization of care were evaluated (e.g. the effect of using the Da Vinci robot for other indications, and the effect of performing many RARPs per year (>150 procedures/year)). This is the first cost-effectiveness analysis of RARP compared to LRP from a Dutch perspective, aiming to inform reimbursement decisions in the Netherlands and provide input for international debate.

REFERENCES

1. Burns, L. R. Interview: Growth and innovation in medical devices: A conversation with stryker chairman John Brown. *Health Affairs* (2007). doi:10.1377/hlthaff.26.3.w436
2. Drummond, M. Evaluation of health technology: Economic issues for health policy and policy issues for economic appraisal. *Soc. Sci. Med.* (1994). doi:10.1016/0277-9536(94)90059-0
3. Berwick, D. M. Disseminating innovations in health care. *JAMA* **289**, 1969–1975 (2003).
4. Reis, S. E., McDonald, M. C. & Byers, S. J. Crossing the research valleys of death: The University of Pittsburgh approach. *Clin. Transl. Sci.* (2008). doi:10.1111/j.1752-8062.2008.00021.x
5. International Federation of Pharmaceutical Manufacturers and associations. *The pharmaceutical industry and global heatlh: facts and figures.* (2017).
6. IJzerman, M. J. & Steuten, L. M. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Heal. Econ Heal. Policy* **9**, 331–347 (2011).
7. Lowe, D. The Latest on Drug Failure and Approval Rates. *Science Translational Medicine: in the pipeline* (2019).
8. 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* **24**, 459–464 (2008).
9. IJzerman, M. J., Koffijberg, H., Fenwick, E. & Krahn, M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *Pharmacoeconomics* 1–14 (2017). doi:10.1007/s40273-017-0509-1
10. Van Nooten, F. *et al.* Health economics and outcomes research within drug development: Challenges and opportunities for reimbursement and market access within biopharma research. *Drug Discovery Today* (2012). doi:10.1016/j.drudis.2012.01.021
11. Fort, D. G., Herr, T. M., Shaw, P. L., Gutzman, K. E. & Starren, J. B. Mapping the evolving definitions of translational research. *J. Clin. Transl. Sci.* (2017). doi:10.1017/cts.2016.10
12. Hall, B. . & Khan, B. *Adoption of new technology.* (2002).
13. Rogers, E. M. *The Diffusion of Innovations. Diffusion of Innovations* (Simon & Schuster, 2003). doi:10.1029/2001WR001015
14. Cain, M. & Mittman, R. Diffusion of Innovation in Health Care. *Ihealthreports* (2002).
15. World Health Organization. Health Technology Assessment. Available at: http://www.who.int/medical_devices/assessment/en/. (Accessed: 12th May 2020)
16. Banta, D. & Jonsson, E. History of HTA: Introduction. *International Journal of Technology Assessment in Health Care* (2009). doi:10.1017/S0266462309090321
17. Drummond, M. F. *et al.* Key principles for the improved conduct of health technology assessments for resource allocation decisions. *International Journal of Technology Assessment in Health Care* (2008). doi:10.1017/S0266462308080343
18. Miquel-Cases, A. *et al.* (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. *Cancer Treat. Rev.* **52**, 117–127 (2017).
19. Sculpher, M., Drummond, M. & Buxton, M. The iterative use of economic evaluation as part of the process of health technology assessment. *Journal of Health Services Research and Policy* (1997). doi:10.1177/135581969700200107
20. Markiewicz, K., Van Til, J. A. & Ijzerman, M. J. Medical devices early assessment methods: Systematic literature review. *International Journal of Technology Assessment in Health Care* (2014). doi:10.1017/S0266462314000026
21. Grosse, S. D. Assessing cost-effectiveness in healthcare: History of the $50,000 per QALY threshold. *Expert Review of Pharmacoeconomics and Outcomes Research* (2008). doi:10.1586/14737167.8.2.165
22. Cameron, D., Ubels, J. & Norström, F. On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. *Global Health Action* (2018). doi:10.1080/16549716.2018.1447828
23. Versteegh, M. M. *et al.* Severity-Adjusted Probability of Being Cost Effective. Pharmacoeconomics (2019). doi:10.1007/s40273-019-00810-8
24. European Medicines Agency. European Medicines Agency - About us. Available at: https://www.ema.europa.eu/en/about-us/history-ema. (Accessed: 26th June 2020)

25. Mauskopf, J. A. *et al.* Principles of good practice for budget impact analysis: Report of the ISPOR Task Force on Good Research Practices- Budget Impact Analysis. *Value Heal.* (2007). doi:10.1111/j.1524-4733.2007.00187.x
26. Minesterie van Volksgezondheid welzijn en sport. *Nieuwe regelgeving medische hulpmiddelen en in-vitro diagnostica.* (2020).
27. Husereau, D., Henshall, C., Sampietro-Colom, L. & Thomas, S. CHANGING HEALTH TECHNOLOGY ASSESSMENT PARADIGMS? *Int. J. Technol. Assess. Health Care* **32**, 191–199 (2016).
28. Ciani, O. & J.mmi, C. The role of health technology assessment bodies in shaping drug development. *Drug Design, Development and Therapy* (2014). doi:10.2147/DDDT.S49935
29. Lehoux, P., Miller, F. A., Daudelin, G. & Denis, J. L. Providing value to new health technology: The early contribution of entrepreneurs, investors, and regulatory agencies. *Int. J. Heal. Policy Manag.* (2017). doi:10.15171/ijhpm.2017.11
30. Schot, J. & Rip, A. The past and future of constructive technology assessment. *Technol. Forecast. Soc. Change* **54**, 251–268 (1997).
31. Douma, K. F. L., Karsenberg, K., Hummel, M. J. M., Bueno-de-Mesquita, J. M. & van Harten, W. H. Methodology of constructive technology assessment in health care. *Int. J. Technol. Assess. Health Care* **23**, 162–168 (2007).
32. Institute of Medicine. Crossing the quality chasm: a new health system for the 21th century. *IOM* 1–8 (2001).
33. Poulsen, P. B. *Health Technology Assessment and Diffusion of Health Technology*. (1999).
34. Bouvy, J. C., Sapede, C. & Garner, S. Managed entry agreements for pharmaceuticals in the context of adaptive pathways in Europe. *Front. Pharmacol.* (2018). doi:10.3389/fphar.2018.00280
35. Reckers-Droog, V., Federici, C., Brouwer, W. & Drummond, M. Challenges with coverage with evidence development schemes for medical devices: A systematic review. *Heal. Policy Technol.* (2020). doi:10.1016/j.hlpt.2020.02.006
36. Bruegger, U. A review of coverage with evidence development in different countries : what works and what doesnt? *HTAi* (2014).
37. Levin, L. *et al.* Coverage with evidence development: The Ontario experience. *Int. J. Technol. Assess. Health Care* (2011). doi:10.1017/S0266462311000018
38. Hilgerink, M. P., Hummel, M. J. M., Manohar, S., Vaartjes, S. R. & Ijzerman, M. J. Assessment of the added value of the Twente Photoacoustic Mammoscope in breast cancer diagnosis. *Med. Devices Evid. Res.* **4**, 107–115 (2011).
39. Department for Business Innovation & Skills. Eight great technologies. (2013). Available at: https://www.gov.uk/government/speeches/eight-great-technologies. (Accessed: 11th July 2017)
40. European Medicines Agency. Advanced Therapy Medicinal Products. Available at: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000294.jsp&mid=WC0b01ac05800241e0. (Accessed: 30th May 2018)
41. EudraLex. Volume 4- Good manufacturing practice (GMP) Guidelines. *European Commission, Public Health.* (2016). Available at: https://ec.europa.eu/health/documents/eudralex/vol-4_en. (Accessed: 24th July 2017)
42. de Wilde, S. *et al.* Hurdles in clinical implementation of academic advanced therapy medicinal products: A national evaluation. *Cytotherapy* **18**, 797–805 (2016).
43. Rosenberg, S. A. *et al.* Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N. Engl. J. Med.* **319**, 1676–1680 (1988).
44. van Harten, W. H. & Retèl, V. P. Innovations that reach the patient: early health technology assessment and improving the chances of coverage and implementation. *Ecancermedicalscience* **10**, 683 (2016).
45. von Minckwitz, G. *et al.* Response-guided neoadjuvant chemotherapy for breast cancer. *J. Clin. Oncol.* **31**, 3623–3630 (2013).
46. Heil, J. *et al.* Eliminating the breast cancer surgery paradigm after neoadjuvant systemic therapy: current evidence and future challenges. *Annals of Oncology* (2020). doi:10.1016/j.annonc.2019.10.012
47. Kelley, W. E. The evolution of laparoscopy and the revolution in surgery in the decade of the 1990s. *J. Soc. Laparoendosc. Surg.* (2008).
48. Muñoz, E., Muñoz, W. & Wise, L. National and surgical health care expenditures, 2005-2025. *Ann. Surg.* (2010). doi:10.1097/SLA.0b013e3181cbcc9a
49. Birkmeyer, J. D. *et al.* Medicare payments for common inpatient procedures: Implications for episode-based payment bundling. *Health Serv. Res.* (2010). doi:10.1111/j.1475-6773.2010.01150.x

50. Blazeby, J. M. *et al.* Demonstration of the IDEAL recommendations for evaluating and reporting surgical innovation in minimally invasive oesophagectomy. *Br. J. Surg.* **98**, 544–551 (2011).
51. Hirst, A. *et al.* No Surgical Innovation Without Evaluation: Evolution and Further Development of the IDEAL Framework and Recommendations. *Ann. Surg.* (2019). doi:10.1097/SLA.0000000000002794
52. Simpson, A. H. R. W., Howie, C. R. & Norrie, J. Surgical trial design - learning curve and surgeon volume: Determining whether inferior results are due to the procedure itself, or delivery of the procedure by the surgeon. *Bone Joint Res.* **6**, 194–195 (2017).
53. Leung, T. & Vyas, D. Robotic Surgery: Applications. *Am. J. Robot. Surg.* **1**, 1–64 (2014).
54. Lardas, M. *et al.* Quality of Life Outcomes after Primary Treatment for Clinically Localised Prostate Cancer: A Systematic Review. *Eur. Urol.* **72**, 869–885 (2017).
55. O'Leary, E., Drummond, F. J., Gavin, A., Kinnear, H. & Sharp, L. Psychometric evaluation of the EORTC QLQ-PR25 questionnaire in assessing health-related quality of life in prostate cancer survivors: a curate's egg. *Qual. Life Res.* **24**, 2219–2230 (2015).
56. Abrishami, P., Boer, A. & Horstman, K. Understanding the adoption dynamics of medical innovations: Affordances of the da Vinci robot in the Netherlands. *Soc. Sci. Med.* **117**, 125–133 (2014).
57. Ramsay, C., Pickard, R., Robertson, C., Close, A. & Vale, L. Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol. Assess. (Rockv).* **16**, 313 (2012).
58. Health Quality Ontario. Ontario health technology assessment series: Robotic surgical system for radical prostatectomy: A health technology assessment. *Ont. Health Technol. Assess. Ser.* **17**, 1–172 (2017).
59. Ilic, D. *et al.* Laparoscopic and robot-assisted vs open radical prostatectomy for the treatment of localized prostate cancer: a Cochrane systematic review. *BJU Int.* **121**, 845–853 (2018).
60. Singhal, S. The future of surgical oncology image-guided cancer surgery. *JAMA Surgery* **151**, 184–185 (2016).
61. Hertault, A. *et al.* What should we expect from the hybrid room? *J. Cardiovasc. Surg. (Torino).* **58**, 264–269 (2017).
62. Kpodonu, J. Hybrid cardiovascular suite: The operating room of the future. *Journal of Cardiac Surgery* **25**, 704–709 (2010).
63. Kinoshita, T. *et al.* The Survival Benefit of a Novel Trauma Workflow that Includes Immediate Whole-body Computed Tomography, Surgery, and Interventional Radiology, All in One Trauma Resuscitation Room: A Retrospective Historical Control Study. *Ann. Surg.* (2017). doi:10.1097/SLA.0000000000002527
64. Ruers, T. J. M. The future of surgical oncology. *Br. J. Surg.* (2019). doi:10.1002/bjs.11207
65. Nijkamp, J. *et al.* Prospective study on image-guided navigation surgery for pelvic malignancies. *J. Surg. Oncol.* **119**, 510–517 (2019).
66. Janssen, N. *et al.* Real-time wireless tumor tracking during breast conserving surgery. *Int. J. Comput. Assist. Radiol. Surg.* (2017). doi:10.1007/s11548-017-1684-4
67. Ivashchenko, O. *et al.* Intraoperative electromagnetic navigation towards liver tumors. *Int. J. Comput. Assist. Radiol. Surg.* **13**, 1–273 (2018).
68. Kok, E. *et al.* First Clinical Results of Image-Guided Navigation in Locally Advanced Primary and Recurrent Rectal Cancer Patients. *Eur. J. Surg. Oncol.* **46**, e86 (2020).
69. Baltussen, E. J. M. *et al.* Tissue diagnosis during colorectal cancer surgery using optical sensing: An in vivo study. *J. Transl. Med.* (2019). doi:10.1186/s12967-019-2083-0
70. Brouwer de Koning, S. G. *et al.* Toward Assessment of Resection Margins Using Hyperspectral Diffuse Reflection Imaging (400–1,700 nm) During Tongue Cancer Surgery. *Lasers Surg. Med.* (2019). doi:10.1002/lsm.23161
71. Kho, E. *et al.* Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information. *Biomed. Opt. Express* (2019). doi:10.1364/boe.10.004496
72. Merriam webster: augmented reality. Available at: https://www.merriam-webster.com/dictionary/augmented reality. (Accessed: 24th May 2020)
73. Thomson, B. R. *et al.* Hepatic vessel segmentation using a reduced filter 3D U-Net in ultrasound imaging. in (ed. arXiv) eprint 1907.12109 (2019).
74. Schermers, B. *et al.* Feasibility of magnetic marker localisation for non-palpable breast cancer. *Breast* **33**, 50–56 (2017).
75. Jakub, J. W. *et al.* Current status of radioactive seed for localization of non palpable breast lesions. *Am. J. Surg.* **199**, 522–528 (2010).

76. Gray, R. J. *et al.* Randomized prospective evaluation of a novel technique for biopsy or lumpectomy of nonpalpable breast lesions: radioactive seed versus wire localization. *Ann. Surg. Oncol.* **8**, 711–715 (2001).
77. Nuclear Regulatory Commission. Iodine-125 and Palladium-103 Low Dose Rate Brachytherapy Seeds Used for Localization of Non-Palpable Lesions. Available at: https://www.nrc.gov/materials/miau/med-use-toolkit/seed-localization.html. (Accessed: 7th August 2018)
78. THE COUNCIL OF THE EUROPEAN UNION. *Council directive 2013/59/Eurotom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom a*. (2014).
79. Harvey, J. R. *et al.* Safety and feasibility of breast lesion localization using magnetic seeds (Magseed): a multi-centre, open-label cohort study. *Breast Cancer Res. Treat.* **169**, 531–536 (2018).
80. Zacharioudakis, K. *et al.* Is the future magnetic? Magseed localisation for non palpable breast cancer. A multi-centre non randomised control study. *Eur. J. Surg. Oncol.* (2019). doi:10.1016/j.ejso.2019.06.035
81. 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 Heal.* **19**, 1–13 (2016).
82. Saaty, T. l. Saaty T.L. (2003), The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback. *Creat. Decis. Found.* 114 (2003).
83. Saaty, T. L. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **1**, 83 (2008).
84. Lievens, Y., Van Den Bogaert, W. & Kesteloot, K. Activity-based costing: A practical model for cost calculation in radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **57**, 522–535 (2003).
85. Kaplan, R. S. & Anderson, S. R. Time-driven activity-based costing. *Harv. Bus. Rev.* **82**, 131-138,150 (2004).
86. Kaplan, R. S. & Porter, M. E. How to solve the cost crisis in health care. *Harv. Bus. Rev.* (2011).
87. Sullivan, S. D. *et al.* Budget impact analysis - Principles of good practice: Report of the ISPOR 2012 budget impact analysis good practice II task force. *Value Heal.* **17**, 5–14 (2014).
88. Shell international BV. *Scenarios: An Explorer's Guide*. (2008).
89. Enserink, B. & Hermans, L. *Policy Analysis of Multi-Actor Systems chapter 5: Exploring the Future*. (Eleven International Publishing, 2010).
90. Joosten, S. E. ., Retèl, V. P., Coupé, V. M. H., van den Heuvel, M. M. & van Harten, W. H. Implementation scenarios of Next Generation Sequencing for personalized cancer treatment. *Accept. Publ. BMC cancer* (2016).
91. Roberts, M. *et al.* Conceptualizing a model: A report of the ISPOR-SMDM modeling good research practices task force-2. *Value Heal.* **15**, 804–811 (2012).
92. Caro, J. J., Briggs, A. H., Siebert, U. & Kuntz, K. M. Modeling good research practices - Overview: A report of the ISPOR-SMDM modeling good research practices task force-1. *Value Heal.* (2012). doi:10.1016/j.jval.2012.06.012
93. Briggs, A. H. *et al.* Model parameter estimation and uncertainty: A report of the ISPOR-SMDM modeling good research practices task force-6. *Value Heal.* (2012). doi:10.1016/j.jval.2012.04.014
94. Fenwick, E. *et al.* Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Heal.* (2020). doi:10.1016/j.jval.2020.01.001
95. Retèl, V. P. *et al.* Early cost-effectiveness of tumor infiltrating lymphocytes (TIL) for second line treatment in advanced melanoma: a model-based economic evaluation. *BMC Cancer* **18**, 895 (2018).

PART I

Very early HTA


TIL

2

Selecting image-guided technologies in oncology: A surgeon's perspective

Melanie Lindenberg
Valesca P Retèl
Janine van Til
Koert Kuhlmann
Theo Ruers
Wim H. van Harten

Journal of surgical research. 2021; 257; 333-343

ABSTRACT

Background
To improve surgical performance, image-guided (IG) technologies are increasingly introduced. Yet, It is unknown which oncological procedures yield most value from these technologies. This study aims to select the most promising IG technology per oncologic indication.

Methods
An analytic hierarchical process (AHP) was used to evaluate three IG technologies: navigation, optical imaging and augmented reality, in five oncologic indications compared to usual care. Sixteen decision criteria were selected. The relative importance of the criteria and the expected performance of the technologies were evaluated among surgeons. The combination of these scores gives the expected value per technology.

Results
On criteria level, sparing critical tissue (9%-18%) and reducing the risk of local recurrence (11%-27%) were most important. Navigation was preferred in three indications; removal of lymph nodes (42%), liver (47%) and rectal tumors (33%). In removing rectal tumors, optical imaging was equally preferred (34%). In removing breast and tongue tumors, no technology was clearly preferred.

Discussion
In selecting IG technologies, especially optical and navigation technologies are expected to add value in addition to usual care. Further development of those technologies for the preferred indications seems valuable. Multi-attribute analysis showed to be useful in prioritization of conducting clinical studies and steer R&D initiatives.

INTRODUCTION

Effective resection of advanced malignant tumors can be challenging which thereby could result in an increased risk for (local) recurrence depending on tumor origin[1,2]. As (local) recurrences have an impact on the quality of life, health care costs, and chance on survival[3,4], improving surgical performance is of high clinical and economic importance.

Recently, image-guided surgery (IGS), which integrates imaging modalities such as CT and MRI in the surgical environment, was introduced to improve surgical performance. The hypothesized benefit of IGS is an improvement in surgical performance and reduce the risk of secondary procedures by enabling intraoperative evaluation of surgical success (e.g., surgical margin status), and ultimately improvement in clinical outcomes[1,5–8]. The use of IGS in oncology could, for instance, enable intraoperative evaluation of surgical margin status. Considering that the risk of recurrence is associated with positive surgical margins, IGS could play an important role in the future[9–11]. Although it has a high potential for oncology, IGS has not been introduced widely. This may be explained by the limited ability of the available imaging techniques in the operating room (C-arm CBCT, Ultrasound, and MRI) to accurately visualize the tumor, its boundaries, and its critical surroundings. Therefore, medical device companies and several research groups seek for new technical solutions that could provide the image guidance needed in oncologic surgery (e.g., probes, navigation technologies). The development of these technologies and the subsequent evaluation in terms of safety and efficacy per tumor type is time-consuming. This is especially the case as the clinical evaluation is known to be challenged by factors as learning curve, user interactions, and frequent product modifications[12,13]. To stimulate targeted development and guide the clinical evaluation of these technologies, early evaluation of the potential of these technologies for the interventions of interest would be valuable.

Health technology assessments (HTA) can be used to systematically evaluate social, clinical, ethical, organizational, and economical aspects of new technologies and are mainly used to inform decision-making processes (e.g., reimbursement)[14–16]. For HTA on medical devices an iterative approach has been suggested as devices are known to be frequently adapted, may become of use for other indications, and prices are likely to change over time[15,17]. One of the methods used, especially in an early stage, is a multiple criteria decision analysis (MCDA) which is a method used in complex decision making[18–22].

This study aims to select the most valuable image-guided (IG) technology per indication of three promising image-guided technologies targeted at the oncologic setting by means of MCDA to inform ongoing and new research and development (R&D) initiatives.

METHODS

2

Several types of MCDA exist to evaluate new interventions in an early stage. In this analysis, the analytic hierarchical process (AHP) which is often applied to inform healthcare decision problems was selected as an appropriate method[20]. The steps followed are presented in Figure 1[18].

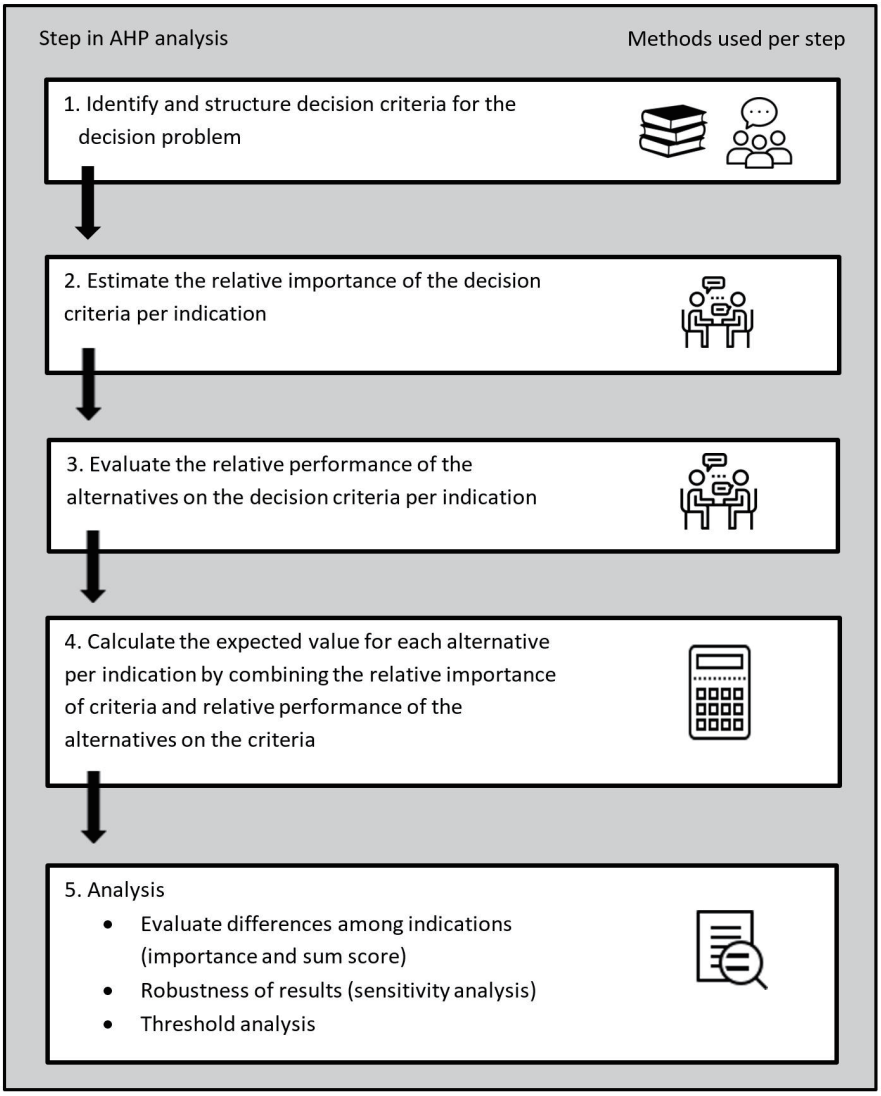


Figure 1. Schematic visualization of the analytic hierarchical process.

In this AHP we focus on five indications: surgical removal of breast, rectal, tongue and liver tumors, and lymph nodes for which three newly developed IG technologies that are expected to improve surgical outcomes are currently being tested in the Netherlands Cancer Institute – Antoni van Leeuwenhoek hospital (NKI). These technologies are: navigation technology[23,24], optical imaging (spectroscopy)[25–27] and augmented reality[28,29]. In addition, the standard of care per indication was included to evaluate the added value of the technologies compared to standard of care. In an AHP, first decision criteria are chosen (section 2.1). Second, the importance of the criteria is evaluated (section 2.2). Third, the expected performance of the four alternatives (technologies) is evaluated per indication on the chosen decision criteria (section 2.3). Finally, combining the importance of the criteria and the performance of the four alternatives on the criteria results in the expected relative value for the alternatives per intervention (section 2.3)[18,30,31]. After discussing our research design with the Institutional Review Board (IRB), they waived formal assessment because no patients were included in our research set-up.

Identification of decision criteria

In step one, the decision criteria were chosen. We identified decision criteria relevant to all indications and used the HTA core model for medical and surgical interventions by EUnetHTA[32]. We identified four domains: effectiveness, efficiency, technical, and organizational. In addition, some sub-criteria were identified such as workflow, training, and capacity. Furthermore, recent clinical guidelines per indication were checked to identify clinically relevant criteria for all indications e.g., the sparing of critical tissue surrounding the tumor and the reduction of the risk of local recurrence. The draft set of criteria was discussed in a face-to-face meeting among a team of surgeons (experienced in one or more of the disease areas) and technical developers/researchers involved in the development of one of the IG technologies on completeness and redundancy[33]. Their feedback was discussed by ML and VR. The set was adapted accordingly when a consensus was reached and finally checked on overlapping criteria[33]. The final set contains 16 criteria that were expected to influence the success of IG technologies in the chosen indications (Figure 2)[18,30,31].

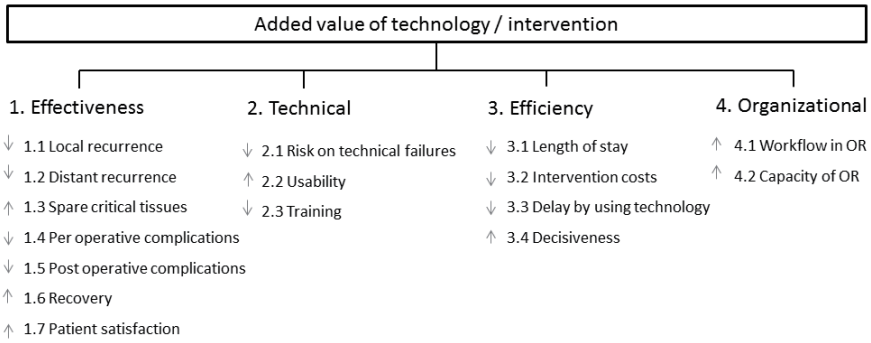


Figure 2. Decision criteria framework. The decision criteria used to evaluate the value of a new image-guided technology in five indications. The framework was based on the HTA core model for medical and surgical interventions[32], literature and group discussions with experts. The arrows describe the direction of the desirable influence which would be classified as added value for the technology. For example, a reduction in local recurrence rate (1.1).

Relative importance of criteria

Due to tumor-specific challenges, the relative importance of the 16 criteria is expected to vary per indication. The importance of the criteria was evaluated per indication among surgeons in the NKI specialized in one or more of the five indications in the beginning of 2018 (Table 1). In a face-to-face interview, they were asked to pair-wise compare the criteria on a reciprocal rating scale[30]. The questionnaire used in this interview contained 37 questions. All criteria were explained to each surgeon before the pairwise comparison to ensure that the interpretation was similar. The definitions are listed in Supplement 2. The individual results from the evaluation were sent back to the surgeon to validate whether these numbers reflected their ideas. If necessary, the initial ratings were adapted accordingly.

After all individual rounds, an extra round of feedback was held among the participating surgeons to reach consensus using an overview of their scores and the scores of their colleagues. This additional round was held to mimic a group meeting which is mainly used in MCDAs to stimulate group discussions and allowing to reach consensus. We highlighted substantial differences between the individual estimates and the group estimates. The surgeons were able to adapt the final values and asked to describe reasons for a change or the initial value in case of no change.

Description of the IG technologies of interest, relative performance and relative value

The technologies of interest are currently only used in a research setting and therefore no comparative data is available yet. For that reason, twenty surgeons having experience with at least one of the technologies were asked to express the expected performance of the three IG technologies and usual care on each of the criteria. The experience with the technologies was thought to be required to evaluate the experienced ease of use and expected advantages[34,35]. To enable comparison, the working principle of each technology was described in detail together with the technical developers. The definitions used are presented in Table 1. For evaluation of the technologies on the criterion "intervention costs", a first non-commercial estimation of the costs per patient was calculated by estimating equipment costs and additional material and labor costs per technology. In this second round of interviews, also in the beginning of 2018, each surgeon was asked to pair-wise compare the technologies and usual care on the 16 decision criteria. After the interview, the retrieved values were fed back to the surgeon to validate the ranking of the alternatives and the differences between the alternatives. If these values were not reflecting their opinion, the results were adapted accordingly.

Based on the individual responses, average importance and performance scores per indication were retrieved using the geometric mean. Combining the importance of the criteria and the expected performance on each criterion resulted in a sum score of the technologies and usual care[30]. The sum score per technology is a relative percentage showing the expected relative value compared to the other technologies and usual care. This sum score was calculated for both the full criteria set and the effectiveness domain only.

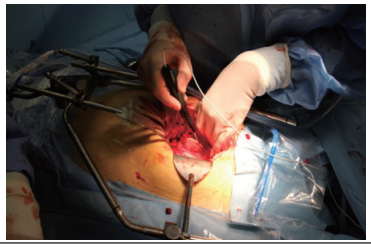
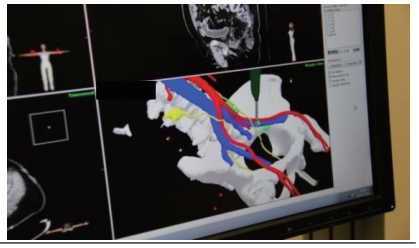
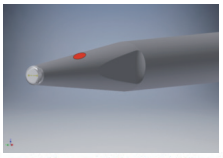
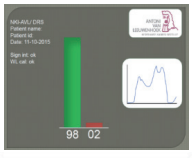
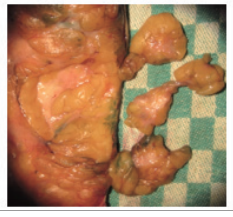
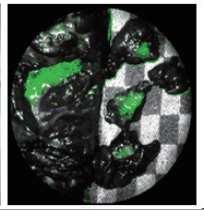
Analysis

To identify differences in the relative importance of the criteria and relative expected value of the technologies between the five interventions, the Standard Error (SE) was used. When there was no overlap between two SEs of one criterion in a certain intervention and the same criterion in the other interventions, that criterion was judged as more or less important. The same approach was used to decide whether one technology was preferred in a specific intervention.

Additionally, sensitivity analyses were conducted. In the first analysis, the importance of each criterion was varied with ± two SEs and measured the influence on the expected relative value by evaluating the robustness of the preferred technology per indication. In the second analysis, the importance of the effectiveness domain was varied over a range of 20% to 80%, the importance of the other domains was adapted according to the initial importance scores.

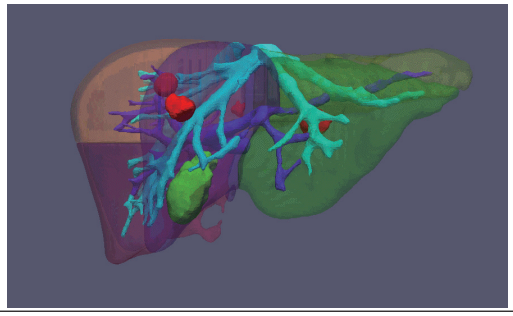
2

Table 1 - Definitions of the three technologies of interest.

Technology	Definition
Usual care	This is defined in the first interviews by the clinicians per intervention. For example, in breast tumor resection the use of the radio-active seed localization is part of current clinical practice.
Navigation	Technology providing a 3D anatomical model that can be linked to the current situation in the operating room by using the CBCT scanner in the OR. The pointer can be located by electromagnetic technology which is linked to the 3D anatomical model. This enables, navigating to the locations of interest. • The model shows an accuracy of at least 5mm • Surgery has to take place in the hybrid operating room • Before or during surgery an extra CT scan is made to link the model to the current situation. This result in a 15-minute delay of the surgery • Additional costs per patient are estimated between €1750 - €3070, - (depending on the utilization of the technology and CBCT)
Pictures	 
Optical imaging (spectroscopy)	Technology that can assess characteristics of tissue (molecular) during a surgery by using the principles of scattering and absorption of light. It can be used in two different clinical situations: (1) as a surgical knife that includes optical fibers that can identify during resection whether the resection takes place in benign or malign tissue. (2) a hyperspectral camera that is used after resecting a certain specimen to check whether the resection margins are clean (ex vivo). This camera can also be used in a laparoscopic setting to also discriminate benign from malign tissue during the intervention (in vivo). • The accuracy to identify the type of tissue correctly is at least 90% • For the in and ex vivo measurements the OR lights should be adjusted to obtain accurate results, this results in at least a delay of 10- 15 minutes • Additional costs per patient are estimated between €700- €1300
Pictures	   

(continued on next page)

Table 1 (continued) - Definitions of the three technologies of interest

Technology	Definition
Augmented reality	Imaging technology in which the reality is enriched by information from preoperative images or an anatomical model. In this phase this technology can be used in the preoperative setting in planning the surgery upfront. The 3D anatomical model that is used in the navigation technology can be presented via a Hololens or a tablet to help to plan the most optimal way to resect the tumor. • The 3D anatomical model is an accurate display of the reality • This technology is used in the days before the surgery takes place • The additional time of this technology is unclear • The additional costs per patient are estimated on €80- €250,-
Pictures	

RESULTS

Relative importance of decision criteria

In total 18 surgeons evaluated the importance of the decision criteria and 12/18 participated in the final feedback round reaching an overall consensus of 83%. In Supplement 1 the characteristics of the surgeons are listed.

Surgeons value the effectiveness domain the highest when evaluating the value of new IG technologies (59%-66%), followed by the technical (13%-21%), efficiency domain (12%-17%), and the organizational domain (4%-12%). Figure 3 and Supplement 3 show the relative importance of each criterion per indication. Among the 16 criteria, “sparing critical tissue” (9%-18%) and “decreasing risk of local recurrence” (11%-27%) are expected to have a high impact on the value of the IG technology in all indications. Specifically for lymph node removal, “decreasing risk of distant metastasis” is expected to have a high influence (13.7%) and for breast cancer surgeries, “improving patient satisfaction” seems to play a major role (16.2%). In addition, among all indications, the influence on the workflow, risk on technical failures, a high ease of use and improved decisiveness were evaluated as important criteria for the success of a new technology. The observed differences among indications are highlighted in Figure 3.

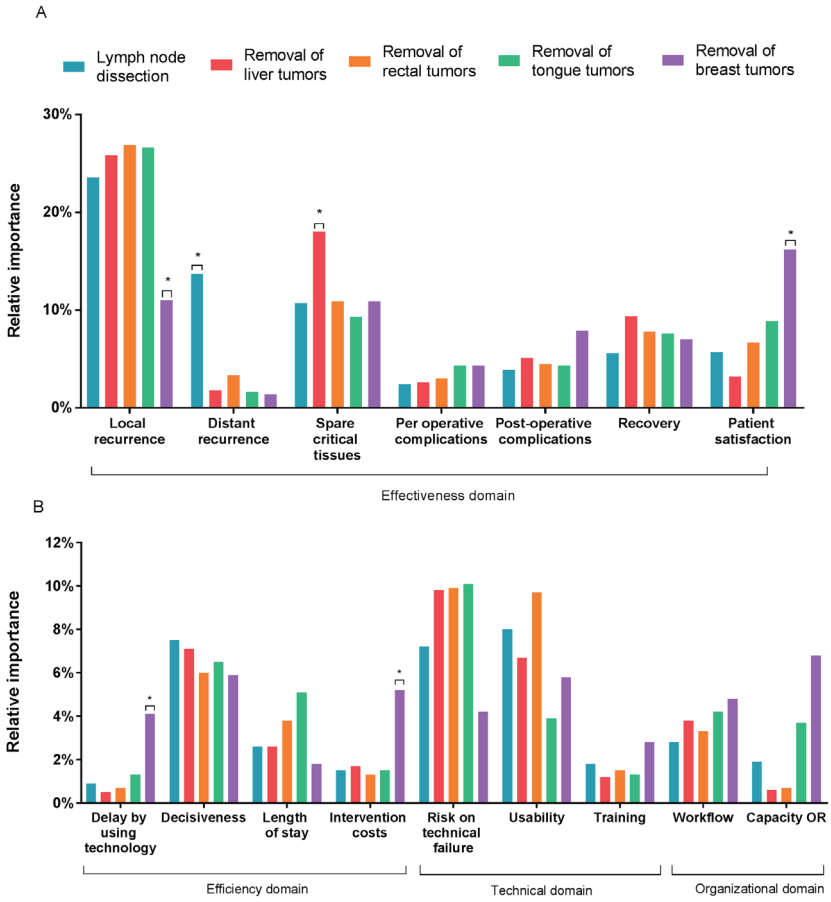


Figure 3. Relative importance of decision criteria per indication. The importance of the decision criteria to evaluate the added value of image-guided technologies shown per indication as evaluated by the surgeons. A shows the results for the effectiveness domain. Figure B shows the results for the efficiency, technical and organizational domain. The shown percentages were retrieved by multiplying the importance of the domain and the importance of the individual criteria in that domain. * show decision criteria that have more or less influence on the expected value of the technology compared to the other indications.

Relative performance and relative expected value of the IG technologies

Seventeen surgeons evaluated the IG technologies and usual care on the decision criteria of which the results are presented in Supplement 4. These spider plots show that surgeons expected that especially navigation could improve their performance in liver, rectal, tongue tumor resection and lymph node dissection on “sparing critical tissue”, “improving patient recovery”, “decreasing the risk of complications”, and “being more decisive”. Additionally, for tongue and rectal tumor resection, optical imaging is expected to “reduce the risk of local

recurrence". In removing breast tumors, usual care is on most of the criteria preferred over the three technologies, but the surgeons expected that navigation could improve "patient satisfaction" and "decisiveness".

Table 2 and Figure 4 show the expected relative value of the IG technologies and usual care per indication when taking all the decision criteria into account. In removal of liver tumors and lymph node dissection, navigation is preferred (47%, 42%) over usual care (16%, 17%), optical imaging (21%, 18%) and augmented reality (17%, 22%). In rectal cancer, both optical imaging (34%) and navigation (33%) are preferred over usual care (16%) and augmented reality (17%). In removing tongue tumors all IG technologies are preferred (26%, 27%, 28%) over usual care (19%) although the difference between usual care and the technologies is small. Finally, in removing breast tumors, navigation and usual care were preferred equally, with 27% and 26% respectively. The expected added value for the use of IG technologies seems thus rather low.

Table 2 - Expected relative added value of each technology per intervention.

	Average score (± two SEs)				
	Lymph node dissection	**Removal of liver tumors**	**Removal of rectal tumors**	**Removal of tongue tumors**	**Removal of breast tumors**
Navigation	0.42 [1] (0.25-0.6)	0.47* [1] (0.37-0.57)	0.33 [2] (0.15-0.51)	0.28 [1] (0.15-0.35)	0.27 [1] (0.17-0.36)
Optical imaging	0.18 [3] (0.10-0.26)	0.21 [2] (0.08-0.33)	0.34 [1] (0.23-0.46)	0.26 [3] (0.10-0.38)	0.23 [4] (0.10-0.27)
Augmented Reality	0.22 [2] (0.11-0.34)	0.17 [3] (0.10-0.24)	0.18 [3] (0.11-0.25)	0.27 [2] (0.13-0.37)	0.24 [3] (0.17-0.31)
Usual care	0.17 [4] (0.09-0.26)	0.16 [4] (0.10-0.22)	0.15 [4] (0.07-0.25)	0.19 [4] (0.13-0.29)	0.26 [2] (0.18-0.31)

This table shows the relative added value scores of each of the alternatives including the range (± two SEs).
* represents that the relative value score and its range showed no overlap with the relative value score and range for usual care (± two SEs). [] show the ranking of the alternatives with 1 representing the favored option for that indication.

For breast cancer, an additional analysis was conducted because the low expected added value may be explained by the composition of the criteria set. During the interview, breast cancer surgeons suggested adding a criterion incorporating the risk of reoperations. Assuming the relative importance of this new criterion equal to "the risk of local recurrence" (11%) and replacing the performance of "risk of local recurrence" by the performance on "risk of reoperations" resulted in a higher expected relative value for navigation (29%) compared to usual care (25%), optical imaging (23%) and augmented reality (23%) than in the initial analysis.

2

Figure 5 shows the expected relative added value of the alternatives when only the effectiveness domain (relative importance and relative performance) is considered. For effectiveness only; navigation was preferred over usual care in removing rectal, tongue, and liver tumors, and lymph nodes and optical imaging were preferred over usual care in the removal of rectal tumors.

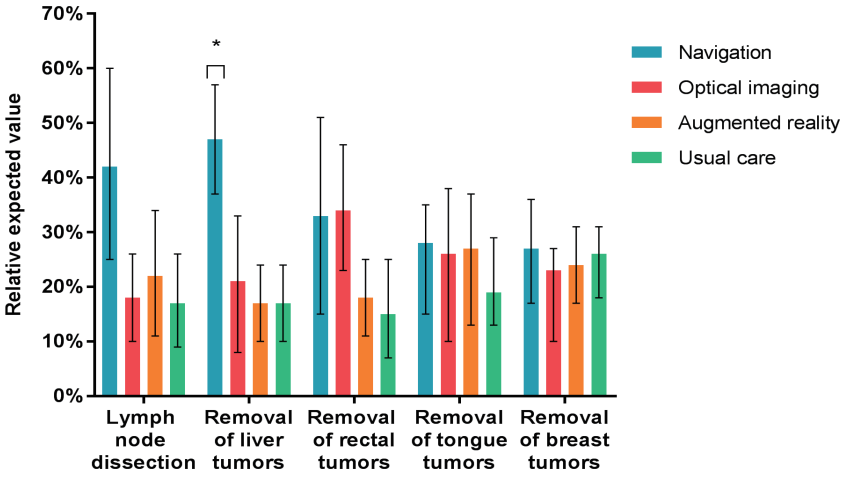


Figure 4. The expected relative value of IG technologies per intervention. Showing the relative expected value of each alternative per intervention. The relative expected value is the result of combining the relative importance of the decision criteria and the expected performance of each alternative on those criteria. * The relative value score and its range showed no overlap with the relative value score and range for usual care (± two SEs).

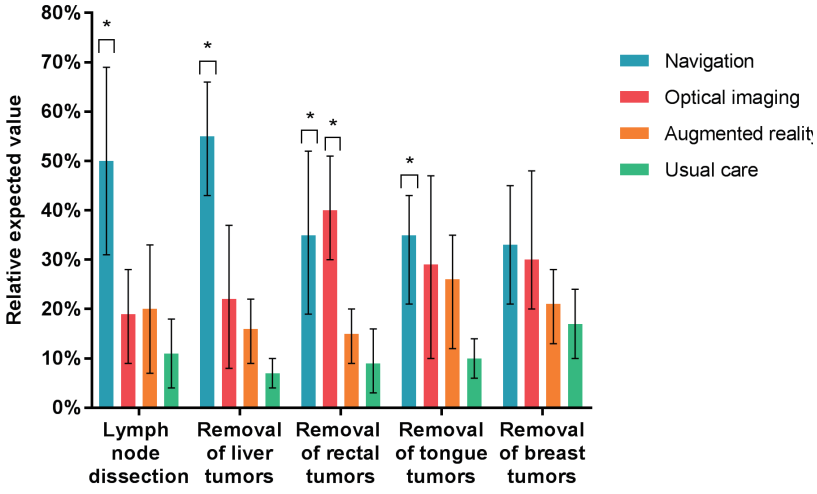


Figure 5. Expected relative value of IG technologies per intervention: effectiveness domain. Showing the expected relative value per technology per intervention when the effectiveness domain would define the total value of the technology. *The relative value score and its range showed no overlap with the relative value score and range for usual care (±2 SEs).

Sensitivity analyses

The uncertainty surrounding the relative importance scores of the criteria showed no major changes in the order of the preferred alternatives per intervention. The preferred option changed only in breast and rectal cancer on some criteria mainly due to a similar initial relative expected value. Ranging the importance of the effectiveness domain shows that it could influence the preferred technology (Supplement 5). For example, when the effectiveness domain explains at least 60% of the value of a technology in tongue cancer, navigation is preferred. Below 60%, augmented reality is preferred.

DISCUSSION

This study provides valuable insight into the potential to select IG surgical technologies used in surgical oncology from a surgeon's perspective.

Based on the relative importance results, the added value of an IG technology is mostly explained by the effectiveness domain (at least 59% importance). Therefore, we may conclude that out of a surgeon's perspective innovative surgical technologies should improve clinical outcomes in order to become successful. Additionally, for the success of technologies, we identified that they should show: no (or a small) influence on the workflow, a small risk of technical failures, high ease of use and an improved decisiveness during surgery (Figure 3 and Supplement 3). The differences in relative importance per indication can be used to steer the development of new and existing IG technologies. Finally, the set up for the current analysis could be used to evaluate the benefit of surgical technologies again in a later stage when more data is available (iterative approach).

Based on the expected relative value per technology, research and clinical research projects can be prioritized. Navigation is expected to add the most value in lymph node dissections and removal of rectal and liver tumors. Hence, our results suggest that the development of navigation technologies should focus on those tumor types. After this, pilot studies and comparative observational studies are needed to evaluate its safety and efficacy. The results on the expected performance (Supplement 4) can inform the choice of relevant outcome parameters. The same applies to optical imaging in resecting tongue and rectal tumors and for augmented reality in tongue tumor resection. Regarding these recommendations, one should keep in mind that not all oncologic specialties were included in the evaluation and therefore we can't guide nor cover the full development process. Besides, related to the phase of development, as the surgeons were – except for the description of the technology – quite unfamiliar with the full potential of the technologies, so the expected relative performance

should be interpreted with caution[34]. For removing breast tumors, our results can't really guide further development as no preferred technology was found. This may be explained by the already highly adequate surgical performance[36,37], by which criteria in favor of usual care (e.g. costs, capacity, workflow) were considered relatively more important than in the other interventions. Besides, the value of IG technologies could be underestimated as shown by the additional analysis incorporating "risk of reoperations".

It should be noted that the strong relative preference for navigation could be biased as the included surgeons had relatively more experience with the navigation technology than with optical imaging and augmented reality[24]. This is because optical imaging is in the preclinical study phase and for augmented reality just recently the first feasibility studies are conducted. Therefore the included surgeons may not have been able to oversee the range of benefits that they may experience when using optical imaging and augmented reality. All participating surgeons had however a similar level of experience with the three technologies, therefore we were unable to check whether the preference for a specific technology is explained by the level of experience. Recently a study showed that surgeons were very enthusiastic about the use of augmented reality in renal cell carcinoma when evaluating surgical performance (e.g., assessment of critical structures)[38]. Additionally, the included head and neck surgeons - who are using 3D models (digital or phantoms) in current clinical practice[39] - were positive on the use of augmented reality in tongue cancer. It seems plausible that the inexperience of the surgeons could thus have underestimated the added value of both optical imaging and augmented reality in this study. In our opinion, this effect was hard to prevent as these technologies are very innovative and therefore incorporating surgeons having experience with all technologies was not possible.

One of the strengths is that the expected added value of these IG technologies was retrieved from the end-users perspective and in perspective of usual care. We especially feel this as a strength with regards to the relative importance results. Additionally, the broad decision criteria framework used was a strength as the set enables evaluation of clinical, process and technical related criteria, which are important determinants in the successful implementation of technologies[12,13,40]. This framework creates a rather negative scenario for the IG technologies by which the technologies have to show substantial benefit to be preferred over usual care. Finally, by using one criteria framework for all indications, our analysis was able to pinpoint at differences between indications that feed R&D of those technologies.

This study has some limitations. The institutional focus limits the generalizability of our results, especially of the relative importance of the decision criteria. Nonetheless, incorporating the uncertainty surrounding the importance scores in the sensitivity analyses did not alter our conclusions. The small number of surgeons included per intervention (≥3) should be noted,

as the sample might not be representative for the Netherlands or other countries. We were however limited to include surgeons having relevant knowledge on the technologies to provide a reasonable estimate of the added value of those IG technologies. To increase the generalizability of the importance scores, in a future study the importance of the decision criteria should be tested among a larger population of surgeons, preferably among an international sample. Furthermore, as previously touched upon, especially the results from the relative performance could be subject of bias, as surgeons were still unfamiliar with the use of optical imaging and augmented reality. However, we feel that this was the best information we could retrieve at this stage in the development process, and may be used with caution to guide further development. The final issue is related to the early nature of this evaluation. Since specific applications of these technologies are yet unknown, a very general evaluation of the technologies by only describing the tumor type and not specifying tumor stage or a specific surgery (e.g., Lower Anterior Resection) was conducted. In future analyses, the technologies should be evaluated for specific cases.

As early stage Health Technology Assessment research can be seen as an iterative process, the future research steps follow the development process of the technologies[41]. The next step will therefore be to evaluate the added value of each technology per indication based on the first clinical results. In a later phase the presented analyses can be updated with the most recent data. For example for the indication lymph node dissection a randomized controlled trial was started comparing surgical success of lymph-node removal with navigation and without (usual care), with the aim to also evaluate the cost-effectiveness of navigation use. Furthermore, in colorectal cancer patients an observational study has been performed showing improved negative resection margin rates with navigated surgery compared to standard surgery in a historical control group[42]. Based on these results an early cost-effectiveness analysis was performed showing that navigation has the potential to become cost-effective in specific clinical subgroups and when the navigation system is used optimally[43]. In colorectal, liver and breast cancer surgery several feasibility studies have been performed for optical imaging applications and augmented reality[44–46]. When clinical studies show improved clinical outcomes, early cost-effectiveness analyses will be performed to evaluate whether the added value weighs up for the additional costs or what could be improved in technology, in technology costs or indications for its use.

This case study showed a specific example of how MCDA could be used in the selection and (early) evaluation of innovative surgical technologies for further research. As surgical innovations are increasingly coming to the market and are likely to result in increased healthcare costs, assessment of the added value of innovations becomes more important. Based on our results we may conclude that for lymph node dissection and liver tumor resection, most is expected from using a navigation system in addition to usual care. For rectal

cancer both navigation and optical imaging seem to be preferred. For removal of tongue and breast tumors no clear preference was identified. In our opinion, this study showed that multi-attribute evaluations can be useful to broadly assess the value before implementation and therefore enables prioritization of clinical research and further development.

ACKNOWLEDGEMENTS

The authors want to thank all the surgeons from the NKI-AVL that participated in the two rounds of interviews.

REFERENCES

1. Singhal, S. The future of surgical oncology image-guided cancer surgery. *JAMA Surgery* **151**, 184–185 (2016).
2. Aliperti, L. A., Predina, J. D., Vachani, A. & Singhal, S. Local and systemic recurrence is the achilles heel of cancer surgery. *Annals of Surgical Oncology* **18**, 603–607 (2011).
3. Guyot, F. *et al.* Time trends in the treatment and survival of recurrences from colorectal cancer. *Ann. Oncol.* **16**, 756–761 (2005).
4. Imkampe, A., Bendall, S. & Bates, T. The significance of the site of recurrence to subsequent breast cancer survival. *Eur. J. Surg. Oncol.* **33**, 420–423 (2007).
5. Hertault, A. *et al.* What should we expect from the hybrid room? *J. Cardiovasc. Surg. (Torino).* **58**, 264–269 (2017).
6. Kpodonu, J. Hybrid cardiovascular suite: The operating room of the future. *Journal of Cardiac Surgery* **25**, 704–709 (2010).
7. Kinoshita, T. *et al.* The Survival Benefit of a Novel Trauma Workflow that Includes Immediate Whole-body Computed Tomography, Surgery, and Interventional Radiology, All in One Trauma Resuscitation Room: A Retrospective Historical Control Study. *Ann. Surg.* **269**, 370–376 (2019).
8. Ruers, T. J. M. The future of surgical oncology. *BJS* **106**, 663–664 (2019).
9. den Dulk, M. *et al.* Risk Factors for Adverse Outcome in Patients With Rectal Cancer Treated With an Abdominoperineal Resection in the Total Mesorectal Excision Trial. *Ann. Surg.* **246**, 83–90 (2007).
10. Liao, C.-T. *et al.* Analysis of Risk Factors of Predictive Local Tumor Control in Oral Cavity Cancer. *Ann. Surg. Oncol.* **15**, 915–922 (2008).
11. Hussein, A. A. *et al.* Ten-Year Oncologic Outcomes Following Robot-Assisted Radical Cystectomy: Results from the International Robotic Cystectomy Consortium. *J. Urol.* **202**, 927–935 (2019).
12. Blazeby, J. M. *et al.* Demonstration of the IDEAL recommendations for evaluating and reporting surgical innovation in minimally invasive oesophagectomy. *Br. J. Surg.* **98**, 544–551 (2011).
13. Hirst, A. *et al.* No Surgical Innovation Without Evaluation. *Ann. Surg.* **269**, 211–220 (2019).
14. World Health Organization. Health Technology Assessment. Available at: http://www.who.int/medical_devices/assessment/en/. (Accessed: 12th May 2020)
15. IJzerman, M. J. & Steuten, L. M. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Heal. Econ Heal. Policy* **9**, 331–347 (2011).
16. Miquel-Cases, A. *et al.* (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. *Cancer Treat. Rev.* **52**, 117–127 (2017).
17. Drummond, M., Griffin, A. & Tarricone, R. Economic Evaluation for Devices and Drugs—Same or Different? *Value Heal.* **12**, 402–404 (2009).
18. 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 Heal.* **19**, 1–13 (2016).
19. Diaby, V. & Goeree, R. How to use multi-criteria decision analysis methods for reimbursement decision-making in healthcare: A step-by-step guide. *Expert Review of Pharmacoeconomics and Outcomes Research* **14**, 81–99 (2014).
20. Schmidt, K., Aumann, I., Hollander, I., Damm, K. & Von Der Schulenburg, J. M. G. Applying the Analytic Hierarchy Process in healthcare research: A systematic literature review and evaluation of reporting. *BMC Medical Informatics and Decision Making* **15**, (2015).
21. Garrison, L. P. *et al.* A Health Economics Approach to US Value Assessment Frameworks—Summary and Recommendations of the ISPOR Special Task Force Report [7]. *Value Heal.* **21**, 161–165 (2018).
22. Friedmann, C., Levy, P., Hensel, P. & Hiligsmann, M. Using multi-criteria decision analysis to appraise orphan drugs: a systematic review. *Expert Rev. Pharmacoecon. Outcomes Res.* **18**, 135–146 (2018).
23. Nijkamp, J., Kuhlmann, K., Sonke, J.-J. & Ruers, T. Image-guided navigation surgery for pelvic malignancies using electromagnetic tracking. in *Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling* (eds. Webster, R. J. & Yaniv, Z. R.) 97862L (2016). doi:10.1117/12.2216213
24. Nijkamp, J. *et al.* Prospective study on image-guided navigation surgery for pelvic malignancies. *J. Surg. Oncol.* **119**, 510–517 (2019).
25. Langhout, G. C. *et al.* Differentiation of healthy and malignant tissue in colon cancer patients using optical spectroscopy: A tool for image-guided surgery. *Lasers Surg. Med.* **47**, 559–565 (2015).

26. Kho, E. *et al.* Hyperspectral Imaging for Resection Margin Assessment during Cancer Surgery. *Clin. Cancer Res.* **25**, 3572–3580 (2019).
27. Baltussen, E. J. M. *et al.* Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery. *J. Biomed. Opt.* **24**, 1–9 (2019).
28. Khor, W. S. *et al.* Augmented and virtual reality in surgery—the digital surgical environment: applications, limitations and legal pitfalls. *Ann. Transl. Med.* **4**, 454–454 (2016).
29. Andress, S. *et al.* On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial. *J. Med. Imaging* **5**, 1–12 (2018).
30. Saaty, T. l. Saaty T.L. (2003), The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback. *Creat. Decis. Found.* 114 (2003).
31. Saaty, T. L. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **1**, 83 (2008).
32. EUnetHTA. HTA Core Model Handbook,. *EUnetHTA Jt. Action 2, Work Packag. 8. HTA Core Model ® version 3.0 (Pdf); Available from www.htacoremodel.info/BrowseModel.aspx* **2**, 408 (2016).
33. 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 Heal.* **19**, 125–137 (2016).
34. Rogers, E. M. The Diffusion of Innovations. *Diffus. Innov.* 576 (2003).
35. Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **13**, 319 (1989).
36. Ahmed, A., Gibbs, P., Pickles, M. & Turnbull, L. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. *J. Magn. Reson. Imaging* **38**, 89–101 (2013).
37. Janssen, N. N. Y. *et al.* Radioactive seed localization in breast cancer treatment. *Br. J. Surg.* **103**, 70–80 (2016).
38. Wellens, L. M. *et al.* Comparison of 3-Dimensional and Augmented Reality Kidney Models With Conventional Imaging Data in the Preoperative Assessment of Children With Wilms Tumors. *JAMA Netw. Open* **2**, e192633 (2019).
39. Kraeima, J. *et al.* Multi-modality 3D mandibular resection planning in head and neck cancer using CT and MRI data fusion: A clinical series. *Oral Oncol.* **81**, 22–28 (2018).
40. 40. Fleuren, M. Determinants of innovation within health care organizations: Literature review and Delphi study. *Int. J. Qual. Heal. Care* **16**, 107–123 (2004).
41. 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* **24**, 459–464 (2008).
42. Kok, E. N. D. *et al.* Association of image-guided navigation with complete resection in patients with locally advanced primary and recurrent rectal cancer: a nonrandomized trial. *JAMA Netw. Open* **3**, e208522–e208522 (2020).
43. Lindenberg, M. *et al.* Image-guided navigation for locally advanced primary and locally recurrent rectal cancer: Evaluation of its early cost-effectiveness. in (J Clin Oncol 38: 2020 (suppl; abstr e16015), 2020).
44. Baltussen, E. J. M. *et al.* Tissue diagnosis during colorectal cancer surgery using optical sensing: an in vivo study. *J. Transl. Med.* **17**, 333 (2019)
45. Thomson, B. R. *et al.* Hepatic vessel segmentation using a reduced filter 3D U-Net in ultrasound imaging. in (ed. arXiv) eprint 1907.12109 (2019).
46. Kho, E. *et al.* Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information. *Biomed. Opt. Express* **10**, 4496 (2019).

SUPPLEMENTARY MATERIAL

Supplement 1 - Number of participants and characteristics

	1st round: Importance of criteria	Delphi round to reach consensus on importance	2nd round: Performance of technologies
Surgeons(n)	18*	12	17*
Per intervention			
Lymph node dissection			
Surgeons (n)	6	3	4
Gender: F/M	1/5	1/2	1/3
Average years of experience	13.2	9.7	14.8
Removal of liver tumours			
Surgeons (n)	3	1	3
Gender: F/M	0/3	0/1	0/3
Average years of experience	9.5	15.0	16.3
Removal of rectal tumours			
Surgeons (n)	4	1	3
Gender: F/M	0/4	0/1	1/2
Average years of experience	13.8	10.0	8.3
Removal of tongue tumours			
Surgeons (n)	5	4	4
Gender: F/M	1/4	1/3	1/3
Average years of experience	13.6	11.5	16
Removal of breast tumours			
Surgeons (n)	5	3	3
Gender: F/M	3/2	2/1	2/1
Average years of experience	14.0	13.7	13.7

* Some of surgeons participated in two or three interventions as they were experienced in more than one of our interventions.

Supplement 2 - Definitions of criteria

Domain: Effectiveness	
Local recurrence	The chance on a local or regional recurrence (for instance caused by positive resection margins).
Distant recurrence	The chance on distant recurrence.
Spare critical tissues	The chance that critical tissues are spared.
Per operative complications	The chance on complications that occur during the surgery.
Post-operative complications	The chance on complications that occur within 30 days after surgery.
Recovery	The patient's ability of performing daily activities and having a good quality of life, after being discharged of the hospital.
Patient satisfaction	The outcomes in terms of patient experience / satisfaction and cosmetic outcomes.
Domain: Technical	
Malpractice	Chance on technical failures either by technology itself or misuse of technology.
Usability	Usability of the technology for the future user, and intuitiveness of the technique.
Training	The expected amount of time that should be invested in training the staff to successfully use the technology.
Domain: Efficiency	
Delay by using technology	The expected delay of using this technology during surgery.
Decisiveness	The expected reduction in duration of the OR by being more decisive because of the technology that is used during the surgery.
Length of stay	Number of days a patient has to recover in the hospital.
Intervention costs	The expected costs of the proposed intervention including technology and disposables and extra staff necessary, excluding information on the longer OR duration.
Domain: Organizational	
Workflow	The expected influence of the technology on the workflow in the OR.
Capacity OR complex	Expected influence on the capacity of the OR, including OR personal but also special technical support, by implementing this technology.

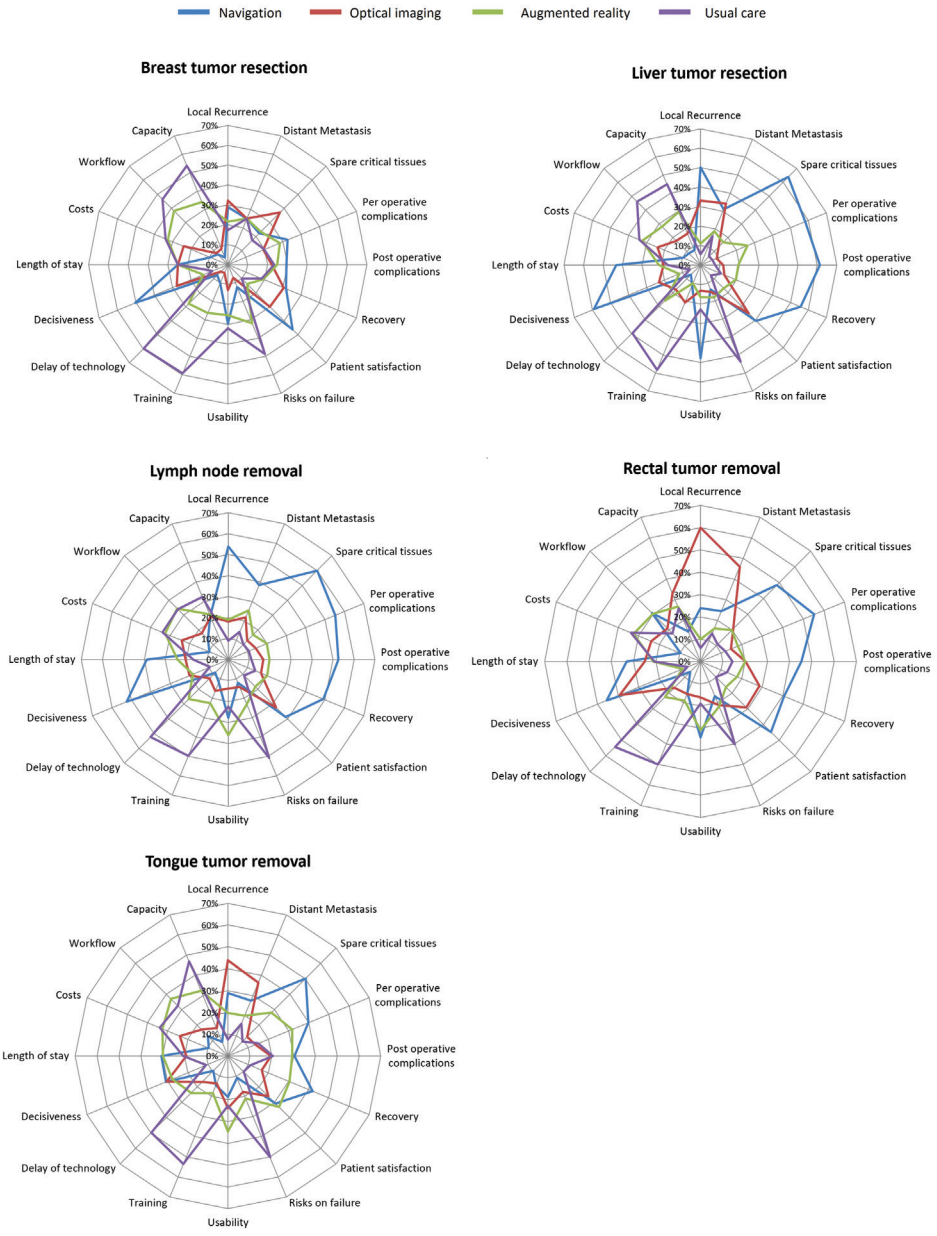
Supplement 3 - Importance of the decision criteria per intervention

Criteria	Lymph node removal (n=6) (±2SE)	Liver tumor resection (n=3) (±2SE)	Tongue tumor resection (n=5) (±2SE)	Breast tumor resection (n=5) (±2SE)	Rectal tumor resection (n=4) (±2SE)
Domain: Effectiveness	65.7% (.60-.67)	66.0% (.63-.65)	62.5% (.55-.64)	58.70% (.53-.59)	63.1% (.58-.68)
Local recurrence	23.6% (.14-.28)	25.8% (.18-.32)	26.6% (.20-.27)	11.0% (.08-.18)	26.9% (.22-.28)
Distant recurrence	**13.7%*(.09-.20)**	1.8% (.01-.02)	1.6% (.01-.02)	1.4% (.01-.02)	3.3% (.00-.12)
Spare critical tissues	10.7% (.06-.14)	18.0%* (.10-.25)	9.3% (.06-.13)	10.9% (.05-.17)	10.9% (.06-.16)
Per-operative complications	2.4% (.01-.03)	2.6% (.01-.04)	4.3% (.02-.08)	4.3% (.02-.06)	3.0% (.00-.08)
Post-operative complications	3.9% (.02-.05)	5.1% (.04-.05)	4.3% (.02-.07)	7.9% (.03-.13)	4.5% (.01-.08)
Recovery	5.6% (.03-.09)	9.4% (.06-.11)	7.6% (.05-.10)	7.0% (.03-.12)	7.8% (.03-.13)
Patient satisfaction	5.7% (.01-.15)	3.2% (.01-.01)	8.9% (.05-.16)	**16.2%* (.10-.18)**	6.7% (.04-.07)
Domain: Technical	17.0% (.10-.15)	17.7% (.12-.26)	15.3% (.10-.24)	12.8% (.09-.17)	21.1% (.16-.28)
Risk of technical failure	7.2% (.03-.10)	**9.8% (.03-.21)**	**10.1% (.05-.02)**	4.2% (.01-.11)	**9.9% (.04-.20)**
Usability	8.0% (.03-.09)	6.7% (.01-.17)	3.9% (.02-.09)	5.8% (.02-.11)	9.7% (.04-.19)
Training	1.8% (.01-.02)	1.2% (.01-.02)	1.3% (.01-.03)	2.8% (.01-.05)	1.5% (.01-.02)
Domain: Efficiency	12.6% (.13-.24)	11.9% (.05-.21)	14.3% (.09-.21)	17.0% (.11-.26)	11.8% (.11-.12)
Delay by using technology	0.9% (.01-.02)	0.5% (.00-.01)	1.3% (.00-.04)	4.1%* (.01-.07)	0.7% (.00-.01)
Decisiveness	7.5% (.07-.15)	7.1% (.03-.14)	6.5% (.03-.11)	5.9% (.01-.11)	6.0% (.05-.07)
Length of stay	2.6% (.02-.07)	2.6% (.01-.08)	5.1% (.02-.09)	1.8% (.00-.16)	3.8% (.02-.05)
Intervention costs	1.5% (.01-.02)	1.7% (.01-.04)	1.5% (.01-.03)	5.2%* (.02-.10)	1.3% (.01-.02)
Domain: Organizational	4.7% (.03-.07)	4.4% (.04-.04)	8.0% (.06-.11)	11.6%* (.08-.17)	4.0% (.04-.04)
Workflow	2.8% (.01-.06)	3.8% (.03-.04)	4.2% (.02-.08)	4.8% (.01-.11)	3.3% (.03-.03)
Capacity OR complex	1.9% (.01-.04)	0.6% (.01-.01)	3.7% (.02-.07)	6.8% (.03-.14)	0.7% (.01-.01)
	100%	100%	100%	100%	100%

The underlined and bold values are representing the three most important criteria per intervention. * means that the weight of the criteria is more or less important compared to the other interventions taking into account the range surrounding the values.

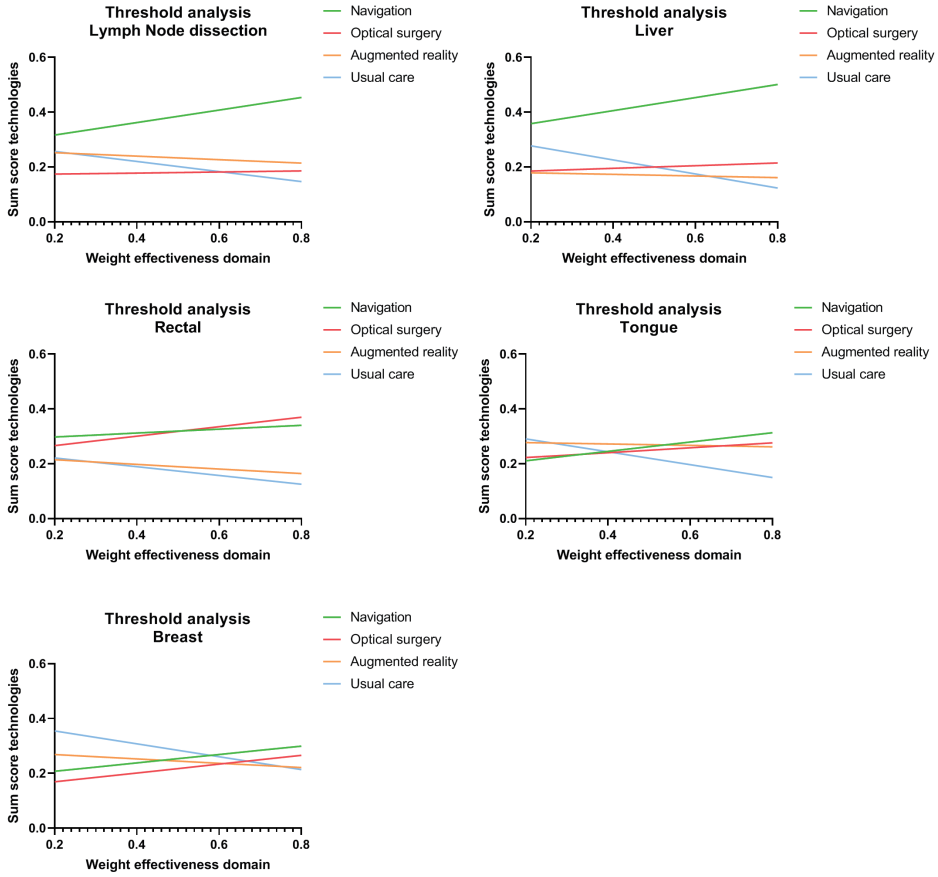
2

Supplement 4 - Overview of relative performance per technology per intervention

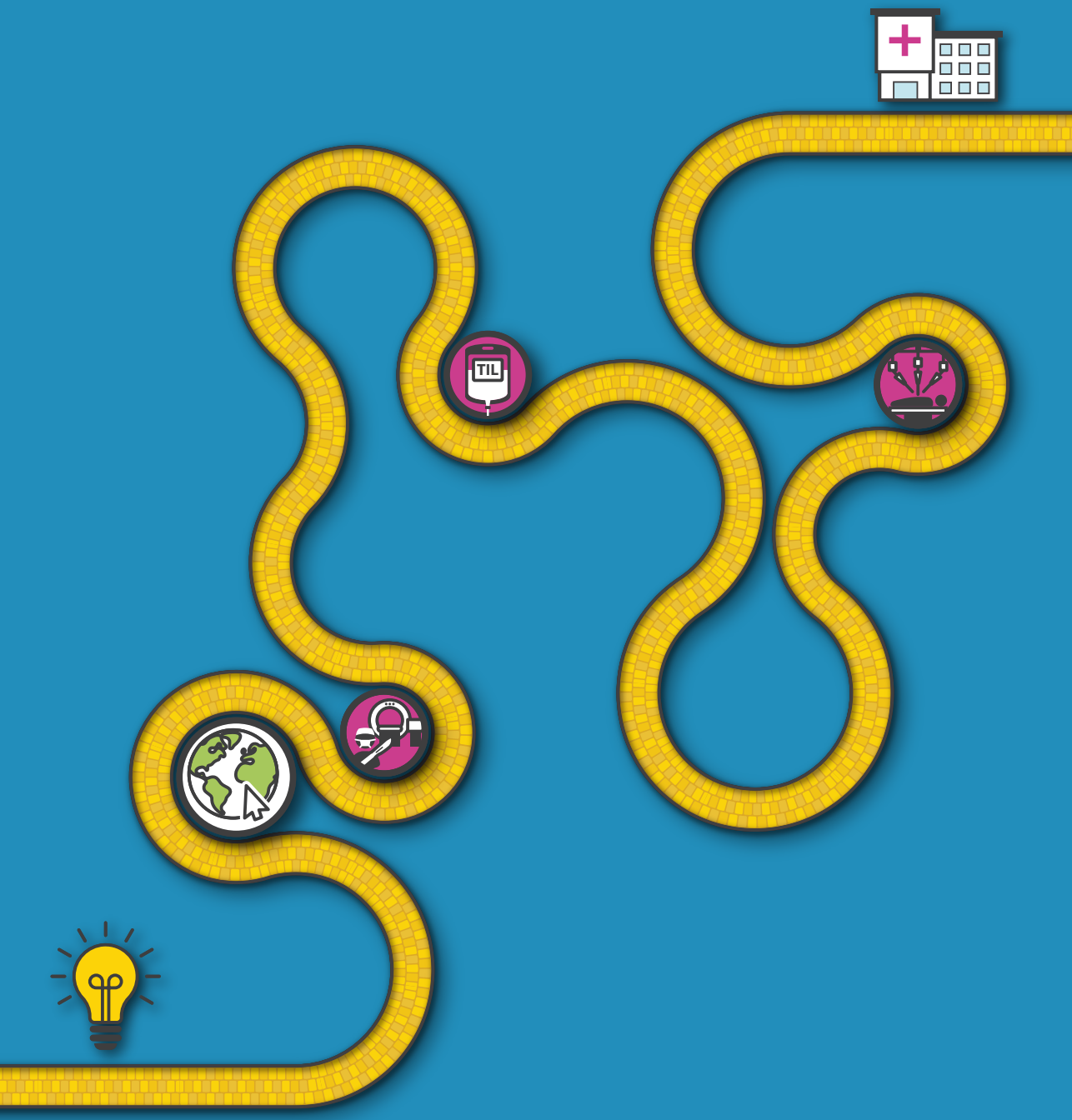


Spider plots showing the relative expected performance of each of the alternatives (without incorporation of the importance of the criteria).

Supplement 5 - Threshold analysis per intervention



The results of a sensitivity analysis in which the importance of the effectiveness domain was varied from 20% to 80%. The importance of the other three domains were adjusted according their initial values. For liver tumor and lymph node removal, navigation remained the preferred option when the importance is smaller than 20%. In removal of rectal tumors, both navigation and optical imaging remained preferred. For tongue tumor removal, augmented reality was preferred over navigation when the importance of the effectiveness domain was <60% and for breast tumor resection usual care was preferred when the importance of the effectiveness domain was <60%.


TIL

3

Imaging performance in guiding response to neoadjuvant therapy according to breast cancer subtypes: A systematic literature review

Melanie Lindenberg
Anna Miquel-Cases
Valesca P Retèl
Gabe S. Sonke
Jelle Wesseling
Marcel P. M. Stokkel
Wim H. van Harten

Crit Rev Oncol Hematol. 2017; 112; 198-207

ABSTRACT

Background

Monitoring therapeutic response to neoadjuvant chemotherapy (NAC) is likely to improve NAC effectiveness in breast cancer (BC). Imaging performance seems to vary per tumour subtype (by ER and HER2 status), therefore we performed a systematic review on subtype specific imaging performance in monitoring NAC in BC.

Methods

Studies examining imaging performance in predicting pathologic complete response(pCR) during NAC in BC subtypes were selected. Per study, negative- and positive predictive value, sensitivity (se) and specificity (sp), AUC and accuracy were derived.

Results

Fifteen/106 articles were included. Inter-study variability was revealed in: monitoring interval, response and pCR definitions. In ER-positive/HER2-negative BC, ^{18}F-FDG-PET/CT showed se/sp of 38%-89%/74%-100%, MRI showed se/sp of 35%-37%/87%-89%. In triple negative BC, ^{18}F-FDG-PET/CT showed se/sp of 0%-79%/95%-100%. ^{18}F-FDG-PET/CT showed in ER-positive/HER2-positive BC se/sp of 59%/80% and in ER-negative/HER2-positive 27%/88%.

Discussion

Evidence on imaging performance in monitoring NAC according BC subtypes is lacking. Consensus should be reached in: definitions of pCR, response and monitoring interval before starting well-designed studies.

INTRODUCTION

In 2012, 1.7 million new cases of breast cancer were diagnosed worldwide. Despite research and improvements in breast cancer treatment, breast cancer is still: one of the most prevalent cancers overall, the most prevalent cancer among women, and one of the main causes of death[1]. Research on new treatment approaches is thus of evident interest.

Neoadjuvant chemotherapy (NAC) showed to be at least equally effective as adjuvant chemotherapy[2] while having additional advantages[3,4], such as the ability to monitor therapeutic response during treatment[5]. Early therapeutic response assessed by imaging seems to be a predictor of pathologic complete response (pCR)[6], usually defined as absence of any residual invasive tumour cells in the original tumour bed and axilla[7]. PCR itself predicts long-term survival, especially in HER2positive and triple negative (TN) tumours[8,9], monitoring early therapeutic response may be used to guide systemic treatment, which is called a response-guided NAC approach[10]. Under this scenario, patients could be monitored after a specific number of NAC cycles, and according to their response at imaging, their further systematic treatment could be tailored, i.e. responders continue with the same initial treatment, and non-responders can be switched to a presumably non-cross-resistant regimen (Figure 1)[10].

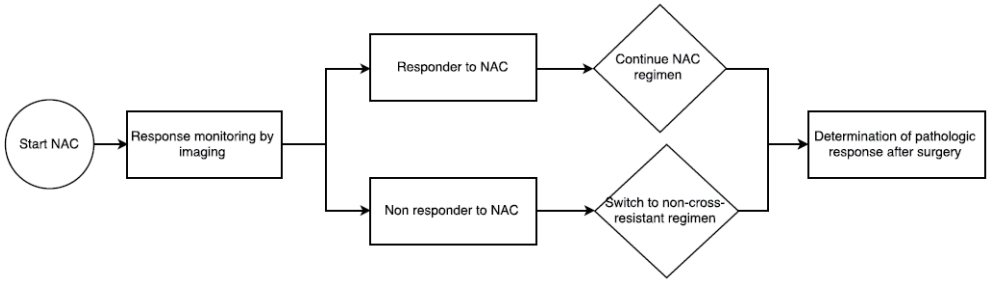


Figure 1. Response-guided neoadjuvant (NAC) approach. Patients start with first-line NAC treatment and after a specific number of cycles, they are monitored by imaging. Patients considered responders of NAC at imaging (according to a pre-defined threshold) continue the same initial treatment, whereas non-responders are switched to a presumably non-cross resistant treatment. Upon NAC finalization, pathologic response is determined at surgery, which is used to determine whether there the imaging results were correct.

Currently, there is no definite guideline to assess response to NAC during treatment. Previous authors proposed physical examination plus mammography and ultrasound, but their performance seems to be limited[11–13]. Therefore, performance examination of more advanced techniques, i.e. magnetic resonance imaging (MRI) and PET – Computed Tomography (PET/CT) is of interest. So far, meta-analyses have shown sensitivities and specificities of 68% and 91% for dynamic contrast-enhanced (DCE)-MRI[14], 93% and 82% for diffusion-weighted (DW)-

MRI[14] and 84% and71% for ^{18}F-FDG-PET/CT[15] respectively. On the basis of these findings, MRI is currently the technique mainly used in clinical practice. While these techniques seem to already have better performance, recent studies have shown that breast cancer subtype affects imaging performance[16–18]. Hence, personalizing the use of imaging techniques based on subtypes may further improve their performance in evaluating therapeutic response[16,19].

As there is no subtype-specific guidance on imaging techniques to monitor therapeutic response during NAC to guide in further treatment regimen, this paper aims to create an overview of current knowledge on the performance of imaging techniques in breast cancer subtypes based on expression of ER and HER2.

METHODS

We performed a systematic literature search to find studies reporting on the performance of imaging in assessing pCR during NAC for breast cancer subtypes.

Search strategy

For PubMed the terms: “breast cancer" (MeSH: Breast neoplasm); “imaging” (i.e. MRI, PET/CT); “outcome” (pathologic complete response, clinical response); “Neoadjuvant chemotherapy” and “breast cancer subtype”(oestrogen receptor (ER), progesterone receptor(PR), luminal, triple negative (TN) and human endocrine receptor 2 (HER2)) were combined for the systematic search (supplement 1). Snowballing was used to find additional relevant publications.

Selection criteria

The search was limited to studies written in English and published between January 2000 and March 2015. Case studies were excluded. Studies were included if performance data of the imaging technique(s) was reported: before and during NAC, specified to at least one receptor status (ER/HER2) and controlled with pCR as primary outcome. As secondary outcomes the neoadjuvant response index (NRI)[20] and residual cancer burden[21] were accepted as response definition. Finally, studies using FDG-PET without CT were excluded, as this technology is no longer recommended in daily practice.

Data extraction

The first selection was performed based on abstract information and following the inclusion

and exclusion criteria by two independent reviewers (AMC and ML). The selected studies were fully read by the same reviewers and were again assessed based on the inclusion and exclusion criteria. Disagreements were first discussed between the two reviewers, and if no agreement was reached, a third reviewer was approached (VR). For each article, the following items were extracted: author, sample size, study design, treatment regimen, breast cancer subtype, clinical stage, age, monitoring technique, cut-off value or response definition at imaging, interval time i.e. number of NAC cycles between baseline and response monitoring, technical settings, pCR definition: pCR or partial response, performance results, i.e. sensitivity, specificity, accuracy, negative and positive predictive values (NPV, PPV) and Area Under the Curve (AUC) in a Receiver Operating Curve (ROC) and if available false and true positives and negatives. Finally, we requested to authors of included studies in which information was presented towards only ER or HER2, whether performance information on further stratified groups existed.

Quality assessment

Three research design criteria were defined to assess quality of the included articles. Articles were considered of sufficient quality if they complied with 2 of the 3 following criteria: 1) no treatment switch during NAC; 2) score higher than 8 on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)[22]; and 3) sample size higher or equal to 20. If more than one subtype was presented in the article, criteria 2 and 3 were assessed per subtype.

Performance of imaging

For articles directly reporting on the number of true/false negative/positive (TN,FN,TP,FP) patients, and studies in which these were possible to derive, 2 x 2 contingency tables were constructed. These were used to calculate sensitivity(ability of imaging to identify non-responders with residual tumour tissue after NAC i.e.,TP/TP+FN), specificity (ability of imaging to identify responders achieving a pCR after NAC i.e.,TN/TN+FP), NPV (TN/TN+FN), PPV (TP/TP+FP) and accuracy (TP+TN/all patients). Pooling of these sensitivity and specificity values would be the preferable method to compare different imaging modalities. However, due to substantial heterogeneity across the included studies it was inappropriate to use this method.

Preferred imaging technique per subtype

To score and compare the performance of the imaging techniques a scale was developed. The scale runs from A (perfect performance) to D (insufficient performance), and was applied to the various performance concepts i.e., ROC-AUC value, accuracy and sensitivity/specificity (Table 2). Based on the coding shown in Table 3, the performance results per breast cancer

subtype were placed in order in Table 5, and, if sufficient results were available for different imaging techniques per subtype, the preferred imaging technique in each subtype was chosen.

RESULTS

Of the initially 229 identified articles, 30 were selected for full reading after removing duplicates. Sixteen articles were further excluded because: 1) response monitoring was performed before or after NAC, 2) did not report performance data or did not specify their results to subtypes and 3) FDG-PET was used without CT. After snowballing one extra article was included, which made a total of 15 articles (Figure 2).

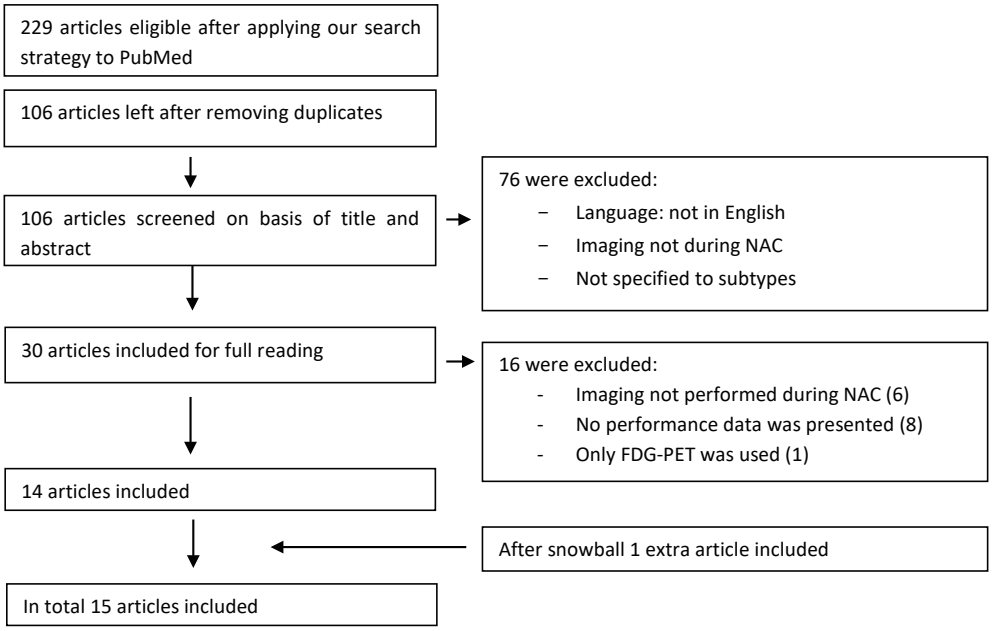


Figure 2. Flow diagram of selection process. Of the 106 identified articles through PubMed, 15 articles were finally included.

Study characteristics

Study populations' size ranged from seven to 246 patients (median: 31), and the overall mean age was 50 years. Studies enrolled patients prospectively (9 studies) and retrospectively (6 studies). One of the five contacted authors replied with additional data towards receptor status[23]. Nine articles presented results for the subgroup of ER-positive/HER2-negative patients[16,23–30], nine for TN patients[16,19,24,25,27,29–32], nine for the whole HER2-positive group[16,19,24,25,27,29,30,33,34] and one for HER2-positive stratified by ER receptor status[35]. The NAC regimen differed per subtype, with ER-positive/HER2-negative patients mainly receiving doxorubicin and cyclophosphamide (AC) and switched to docetaxel and capecitabine (DC) in case of an unfavourable intermediate response[16,23–25], TNBC patients mainly receiving epirubicin and cyclophosphamide followed by docetaxel (EC-D)[26,30–33] or one of the following regimens: intensified EC-D (SIM)[31,32], fluorouracil plus EC (FEC)[19,36], FEC-D[19,36]. ER-negative/HER2-positive patients received mainly EC(-D) followed by a combination of trastuzumab and paclitaxel or docetaxel[30,34,35].

Three definitions of pCR were identified[19,23–25,28,31–33,35,36], and are shown in Table 1. These three definitions will be referred to as pCR in rest of the article. The fourth identified category, described as partial response, was: considerable or partial reduction in tumour cells in breast after completion of neoadjuvant chemotherapy[16,26,27,29,30].

Table 1 - Categorization of different pathologic complete response definitions (pCR).

Category	Classifications and scales used in literature
Category 1	
Complete absence of invasive tumour cells and ductal carcinoma in situ (DCIS) in breast and axillary lymph nodes after completion of neoadjuvant chemotherapy	Chevalier classification grade 1 [37] ypT0 ypN0
Category 2	
Complete absence of invasive tumour cells in the breast and axillary lymph nodes after completion of neoadjuvant chemotherapy	Chevalier classification grade 2 [37] ypT0/is ypN0 Miller and Payne grade 5 and NRG A or D [38]
Category 3	
Complete absence of invasive tumour cells in the breast after completion of neoadjuvant chemotherapy	Miller and Payne grade 5 [38] YpT0/is ypT0/is ypN0/+
Category 4	
Considerable or partial reduction in tumour cells in breast after completion of neoadjuvant chemotherapy	Sataloff classification T-A [39] Sataloff classification T-B [39] Miller and Payne grade 4 [38]

Table 2 - Scale to score diagnostic performance.

Performance		Sensitivity / specificity (α)	ROC-AUC value (β)	Accuracy (γ)
A	Good	Both > 80%	0.80 – 1.00	80% – 100%
B	Sufficient	Both > 60% and < 80% or one result > 60% and < 80% and one result > 80%	0.60 – < 0.80	60% – < 80%
C	Limited	Both > 40% and < 60% or one result > 40% and < 60% and one result > 60%	0.40 – < 0.60	40% – < 60%
D	Insufficient	Both > 0 % and < 40% or one result > 0% and < 40% and one result > 40%	0.00 – < 0.40	0% – < 40%

Each performance concept has its sensitivity and specificity data described as (α), ROC-AUC values were presented as (β) and accuracy results as (γ). The performance scales used per concept are presented in the last three columns of the table, and these are in turn categorized from perfect (A) to insufficient (D) performance by the first column of the table. **Abbreviations:** ROC-AUC = area under the receiver operator curve.

Of the included articles, three were on MRI and 12 on ^{18}F-FDG-PET/CT. A summary of the main settings used in the assessment with ^{18}F-FDG-PET/CT and MRI are presented in Table 3. The study characteristics described are presented in Supplement 2. Regarding quality assessment, three subgroups showed a small sample size[27,29,30], four subgroups had a study design that allowed a switch in treatment during NAC[16,23–25], and no study showed a score below 8 on the QUADAS list (Supplement 3). Since each subgroup of each article satisfied 2 of the 3 criteria described in the method section, no study or subgroup was excluded from further analysis (Table 4).

Table 3 - Main technical settings of imaging techniques used in response assessment summarized per imaging technique.

Imaging technique	Technology	Contrast (dosage)	Settings	Position
MRI[16,23,28]	Philips magnetom vision[16,23] 1.5T and 3.0T magnet[16,23,28]	Gadolinium (14ml of 0.1mmol/kg)[16,23]	-	Use of breast coils[16,23,28]
^{18}F-FDG-PET/CT[19,24–27,29–35]	Philips[19,24–26,30–33,35,40] GE medical[27,29,35] Siemens [35]	^{18}F-FDG (3.5 MBq/kg – 7.4 MBq/kg)[19,24–27,29–35] Fasted 6 hours before injection[19,24–27,29–35]	Scan performed 60 to 70 min after contrast injection CT: 120kV and 100mAs[19,24–27,29–35]	Hanging breast method[24,25]

More details are described in the study characteristics table (Supplement 2). **Abbreviations:** MBq = Megabecquerel, mAs = Milliamperesecond, kV = Kilovolt; T = Tesla.

Performance of imaging techniques per subtype

Results on performance of the various imaging techniques per breast cancer subtype are summarized in the section below and in Table 5. Table 5 shows also the number of NAC cycles between baseline monitoring and response monitoring, the cut-off value that was used, and

pCR definition according Table 1. In addition to Table 5, the sensitivity and specificity values identified per subgroup are shown in an AUC plot (Figure 3).

ER-positive/HER2-negative

Six studies assessed performance of ^{18}F-FDG-PET/CT and three of MRI. The use of ^{18}F-FDG-PET/CT showed AUC-ROC values of: 0.61 (CI 0.37-0.86) after 1 NAC cycle)[24], 0.87 (CI 0.69–1.00) after 3NAC cycles[24], 0.77 (CI 0.68–0.87) after 3 NAC cycles[25] and 0.88 after 2 NAC cycles[30]. An Italian research group described performance of ^{18}F-FDG-PET/CT in 2 articles, showing a sensitivity of 38% and specificity of 100% after 2 NAC cycles using the difference in Standardized Uptake Value (ΔSUVmax)[27,29]. Another study showed sensitivity of 62% and specificity of 78% also after 2 NAC cycles[26]. Using the difference in Total Lesion Glycolysis (ΔTLG), ^{18}F-FDG-PET/CT showed a sensitivity of 89%, sensitivity of 74%, and an AUC value of 0.81 after 2 NAC cycles[26] and 0.96 in case partial responders were included as responder[30]. MRI showed sensitivity of 35%-37%, specificity of 87%-89%, accuracy of 39%-45%, NPV of 10%-22% and PPV of 93%-98% after 3 NAC cycles in two different studies[23,28]. Although the trial of Charehbili et al. was in HER2-negative patients, its results were included in this subtype because the majority of patients showed an ER-positive expression (187/222)[28]. The final included MRI study did not report specific performance results, but showed no significant association between tumour size decrease and Breast Response Index (BRI; part of the NRI outcome measure[20])(p=0.07) after 3 NAC cycles[16].

Triple negative

Eight studies assessed performance of ^{18}F-FDG-PET/CT and one of MRI. The use of ^{18}F-FDG-PET/CT showed AUC-ROC values of 0.76 (CI 0.55-0.96) after 1 NAC cycle[24], 0.87 (CI 0.73-1.00) after 3 NAC cycles[24] and 0.85 (CI 0.68–1.00) also after 3 NAC cycles[25]. Two additional studies assessed performance of ^{18}F-FDG-PET/CT and showed high sensitivity of 71%-79%, specificity of 95%-100% and accuracy of **80%-85%** after 2 NAC cycles[31,32]. Lowering the cut-off value from 50% to ≥-42% ΔSUVmax improved specificity to 100%, but decreased sensitivity to 58% and 64%[31,32]. Two additional studies showed sensitivity of 0% and specificity of 100% after 2 NAC cycles since true or false non-responders were not discovered[27,29]. Of the two final studies, one study showed no significant association (p=0.50) between ΔSUV and pCR after 1 NAC cycle[19], and another showed no improvement (p>0.05) in predictive value by using ΔTLG as imaging measurement parameter[30]. The only study assessing performance of MRI presented no specific performance results but showed a significant association between tumour size decrease and BRI (p <0.001)[16].

Table 4 - Quality assessment based on three criteria.

	Subtype	Sample size	Criteria 1 Treatment is not switched during NAC	Criteria 2 No risk of bias is present	Criteria 3 Sample size is ≥ 20 patients	Include?
[28]	ER-positive/HER2-negative	194	+	+	+	Yes
[35]	ER-negative/HER2-positive	43	+	+	+	Yes
	ER-positive/HER2-positive	34	+	+	+	Yes
[31]	TN	20	+	+	+	Yes
[26]	ER-positive/HER2-negative	64	+	+	+	Yes
[33]	HER2-positive	30	+	+	+	Yes
[32]	TN	50	+	+	+	Yes
[30]	ER-positive/HER2-negative	26	+	+	+	Yes
	TN	13	+	+	-	Yes
	HER2-positive	12	+	+	-	Yes
[19]	ER-positive/HER2-negative	53	+	+	+	Yes
	TN	25	+	+	+	Yes
	HER2-positive	37	+	+	+	Yes
[34]	HER2-positive	57	+	+	+	Yes
[24]	ER-positive/HER2-negative	50	-	+	+	Yes
	TN	31	+	+	+	Yes
	HER2-positive	26	+	+	+	Yes
[25]	ER-positive/HER2-negative	45	-	+	+	Yes
	TN	25	+	+	+	Yes
	HER2-positive	25	+	+	+	Yes
[16]	ER-positive/HER2-negative	103	-	+	+	Yes
	TN	47	+	+	+	Yes
	HER2-positive	38	+	+	+	Yes
[29]	ER-positive/HER2-negative	16	+	+	-	Yes
	TN	9	+	+	-	Yes
	HER2-positive	7	+	+	-	Yes
[23]	ER-positive/HER2-negative	246	-	+	+	Yes
[27]	ER-positive/HER2-negative	31	+	+	+	Yes
	TN	15	+	+	-	Yes
	HER2-positive	14	+	+	-	Yes

1. The treatment was not switched during NAC, 2. Study does not score below 8 on the quality assessment tool for diagnostic accuracy studies (QUADAS), 3. The sample size is above 20 patients.

HER2-positive

Eight studies assessed the performance of ^{18}F-FDG-PET/CT and 1 of MRI. The use of ^{18}F-FDG-PET/CT showed AUC-ROC values of 0.61 (CI 0.33-0.89) after 3 administrations of the 1st NAC cycle (8 in total)[24], 0.59 (CI 0.34-0.85) after the 1st completed NAC cycle[24] and 0.41 (CI 0.16–0.67) after 1 NAC cycle[25]. Two studies also assessed the performance of ^{18}F-FDG-PET/CT and showed sensitivity of 17%-20%, specificity of 100%[27,29] and accuracy of **29%** after 2 NAC cycles[29]. Three other studies also showed performance results in terms of sensitivity and specificity of ^{18}F-FDG-PET/CT. The first study showed sensitivity of 86% and specificity of 75% after 2 NAC cycles and 86% and 63% also after 2 NAC cycles but using another pCR category (Table 5)[33]. In this study the absolute level of residual SUVmax at PET2 showed even a higher accuracy (AUC=0.91). The second study showed sensitivity and specificity of 83% and 53% after 1 NAC cycle[34] and the third, sensitivity, specificity and accuracy of 64%, 83% and **76%** also after 1 NAC cycle[19]. Using ΔTLG showed no improvement in predictive value compared to ΔSUVmax[30]. The MRI study presented a significant association with the BRI (p=0.05) (after 1 NAC cycle)[16].

ER-positive/HER2-positive

One study assessed the performance of ^{18}F-FDG-PET/CT which showed sensitivity of 38%, specificity of 71%, accuracy of **44%,** NPV of 20% and PPV of 86% after 2 weeks, and improved results with a higher cut-off value after 6 weeks: sensitivity of 59%, specificity of 80%, accuracy of **62%**, NPV of 24% and PPV of 95%[35].

ER-negative/HER2-positive

One study showed sensitivity of 27%, specificity of 88%, accuracy of **64%**, NPV of 65% and PPV of 60% for the use of ^{18}F-FDG-PET/CT after 2 weeks and sensitivity of 18%, specificity of 76%, accuracy of **54%**, NPV of 59% and PPV of 33% after 6 weeks[35].

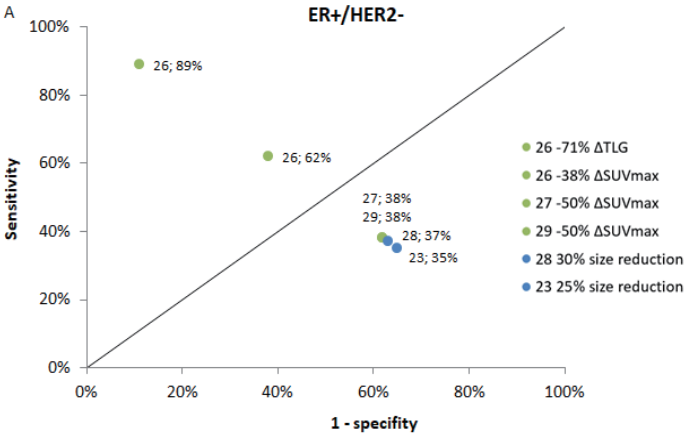
Within-study comparisons

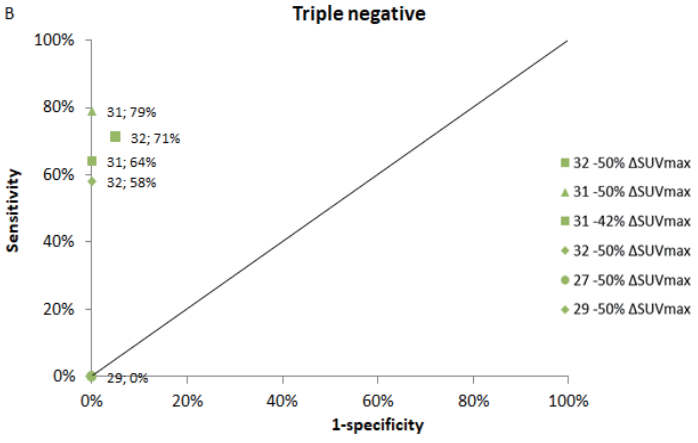
Seven of the 15 included studies analysed imaging performance in more than one BC subtype, the full results are shown in Table 5, in this paragraph we highlight some of these findings.
Koolen et al. analysed the performance of ^{18}F-FDG-PET/CT in all three groups in which the best results were found in the ER-positive/HER2-negative and triple negative groups[24,25]. Martoni et al. and Zucchini et al. described also ^{18}F-FDG-PET/CT performance in all groups which showed low sensitivity/specificity values in all groups[27,29] and Humbert et al. described performance of ^{18}F-FDG-PET/CT in TN and HER2-positive only, in which for the HER2-positive

group a sufficient performance was found[19]. Hat et al. described ^{18}F-FDG-PET/CT performance with several imaging measurement parameters in the three subgroups. Only in ER-positive/HER2-negative it showed improved performance when imaging measurement parameters ΔTLG and ΔMetabolic Active Tumour Volume were used[30]. Finally, MRI performance was described by Loo et al. in all groups, showing only a significant association in the triple negative group between BRI and tumour decrease (P<0,001)[16].

Preferred imaging technique per subtype

We aimed to find the preferred imaging technique per subtype, however due to limited performance results for different imaging techniques within subtypes it was not possible to conclude on the preferred imaging technique.





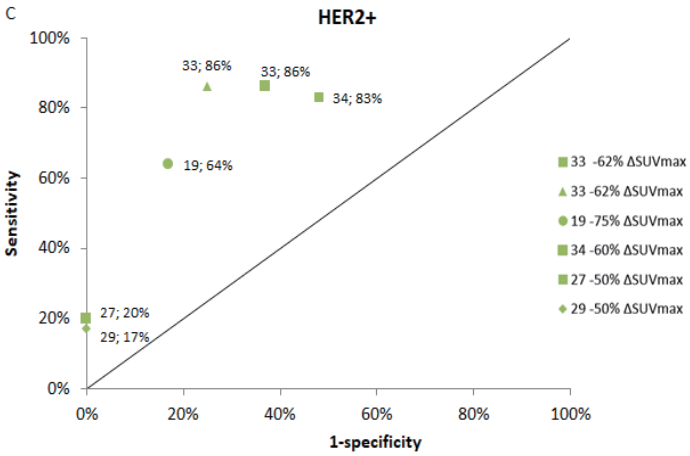


Figure 3. Scatterplots of the sensitvity and specificity results per breast cancer subtype. The identified sensitivity and specificity values as shown in Table 5 are presented in an Area Under the Curve plot per subtype. The corresponding imaging settings (threshold, interval time, and pCR definition) are described in Table 5. Figure 3a. shows sensitivity and specificity values in the subgroup: ER-positive and HER2-negative. The green dots show the values for ^{18}FDG-PET/CT and the blue ones for MRI. Figure 3b. shows sensitivity and specificity values in the triple negative subgroup. In this subgroup only sens/spec data was available for ^{18}FDG-PET/CT (green). Figure 3c shows sensitivity and specificity values in the HER2positive subgroup. In this subgroup only sens/spec data was found for ^{18}FDG-PET/CT (green).

Table 5 - Performance of imaging techniques per subtype.

Article	Monitoring technique	Monitoring interval	Cut-off value	pCR definition
ER-positive/HER2-negative				
30	^{18}F-FDG-PET/CT	After 2 cycles	Several imaging measurement parameters	IV
24	^{18}F-FDG-PET/CT	After 3 cycles	ΔSUVmax	II
26	^{18}F-FDG-PET/CT	After 2 cycles	-71% ΔTLG	IV
26	^{18}F-FDG-PET/CT	After 2 cycles	-38% ΔSUVmax	IV
25	^{18}F-FDG-PET/CT	After 3 cycles	ΔSUVmax	III
24	^{18}F-FDG-PET/CT	After 1 cycle	ΔSUVmax	II
27	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	IV
29	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	IV
28	DCE MRI	After 3 cycles	30% size reduction	III
23	DCE MRI	After 3 cycles	25% size reduction	III
16	DCE MRI	After 3 cycles	30% size reduction	IV
Triple negative				
32	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	II
25	^{18}F-FDG-PET/CT	After 3 cycles	ΔSUVmax	III
24	^{18}F-FDG-PET/CT	After 3 cycles	ΔSUVmax	II
31	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	II
31	^{18}F-FDG-PET/CT	After 2 cycles	-42% ΔSUVmax	II
32	^{18}F-FDG-PET/CT	After 2 cycles	-42% ΔSUVmax	II
24	^{18}F-FDG-PET/CT	After 1 cycle	ΔSUVmax	II
27	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	IV
29	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	IV
19	^{18}F-FDG-PET/CT	After 1 cycle	-75% ΔSUVmax	II
30	^{18}F-FDG-PET/CT	After 2 cycles	Several outcome parameters	IV
16	DCE MRI	After 3 cycles	30% size reduction	IV

Sens	Spec	Acc	NPV	PPV	AUC	Performance score
					ΔSUVmax: 0.88	$A^{(\beta)}$
					ΔTLG: 0.96	$A^{(\beta)}$
					ΔMATV: 0.98	$A^{(\beta)}$
					0.87 (0.69-1.00)	$A^{(\beta)}$
89%	74%		31%	98%	0.81	$B^{(\alpha)}A^{(\beta)}$
62%	78%		12%	97%	0.73	$B^{(\alpha)}B^{(\beta)}$
					0.77 (0.68-0.87)	$B^{(\beta)}$
					0.61 (0.37-0.86)	$B^{(\beta)}$
38%	100%		24%	100%		$D^{(\alpha)}$
38%	100%	50%	27%	100%		$D^{(\alpha)}C^{(\gamma)}$
37%	87%	45%	22%	93%	0.55 (0.45-0.65)	$D^{(\alpha)}C^{(\beta)}C^{(\gamma)}$
35%	89%	39%	10%	98%		$D^{(\alpha)}D^{(\gamma)}$
Association between BRI and tumour decrease was not significant (p = 0.07)						-
71%	95%	80%	67%	96%	-	$A^{(\alpha)}A^{(\gamma)}$
					0.85 (0.69-1.00)	$A^{(\beta)}$
					0.87 (0.73-1.00)	$A^{(\beta)}$
79%	100%	85%	67%	100%	0.881	$B^{(\alpha)}A^{(\beta)}A^{(\gamma)}$
64%	100%	75%	55%	100%	0.881	$B^{(\alpha)}A^{(\beta)}B^{(\gamma)}$
58%	100%	74%	59%	100%	-	$C^{(\alpha)}B^{(\gamma)}$
					0.76 (0.55-0.96)	$B^{(\beta)}$
0%	100%		27%	0%	-	$D^{(\alpha)}$
0%	100%	33%	33%		-	$D^{(\alpha)}D^{(\gamma)}$
No significant correlation between early metabolic response and pCR						-
Use of different parameters did not improve predictive value of ΔSUVmax						-
Association between BRI and largest tumour diameter was significant (p= < 0.001)						-

(continued on next page)

Table 5 (continued) - Performance of imaging techniques per subtype.

Article	Monitoring technique	Monitoring interval	Cut-off value	pCR definition
HER2-positive				
33	^{18}F-FDG-PET/CT	After 2 cycles	-62% ΔSUVmax	III
33	^{18}F-FDG-PET/CT	After 2 cycles	-62% ΔSUVmax	II
19	^{18}F-FDG-PET/CT	After 1 cycle	-75% ΔSUVmax	II
34	^{18}F-FDG-PET/CT	After 1 cycle	-60% ΔSUVmax	II
24	^{18}F-FDG-PET/CT	After 3/8 of 1st cycle$^{\#*}$	ΔSUVmax	II
27	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	IV
24	^{18}F-FDG-PET/CT	After 1 cycle$^{\#}$	ΔSUVmax	II
25	^{18}F-FDG-PET/CT	After 1 cycle$^{\#}$	ΔSUVmax	III
29	^{18}F-FDG-PET/CT	After 2 cycles	-50% ΔSUVmax	IV
30	^{18}F-FDG-PET/CT	After 2 cycles	Several imaging measurement parameters	IV
16	DCE MRI [(2)]	After 1 cycle$^{\#}$	30% size reduction	IV
HER2-positive and ER-positive				
35	^{18}F-FDG-PET/CT	After 6 weeks	-25% ΔSUVmax	III
35	^{18}F-FDG-PET/CT	After 2 weeks	-15% ΔSUVmax	III
HER2-positive and ER-negative				
35	^{18}F-FDG-PET/CT	After 2 weeks	-15% ΔSUVmax	III
35	^{18}F-FDG-PET/CT	After 6 weeks	-25% ΔSUVmax	III

Response definition: ^{I} response category 1; ^{II} response category 2; ^{III} response category 3; ^{IIII} response category 4; **Cut-off values:** ^{1} = cut-off value 25% size reduction, ^{2} = cut-off value= 30% size reduction, ^{3} = cut-off value -15% ΔSUVmax, ^{4} = cut-off value -25% ΔSUVmax, ^{5} = cut-off value: -38% ΔSUVmax, ^{6} = cut-off value: -42% ΔSUVmax, ^{7} = cut-off value: -50% ΔSUVmax, ^{8} = cut-off value -60% ΔSUVmax, ^{9} = cut-off value -62% ΔSUVmax, ^{10} = cut-off value -75% ΔSUVmax, ^{11} = cut-off value: -71% ΔTLG, **Outcome parameters:** * = ΔSUVmax, $^{\Delta}$ = Different outcome parameters, **Performance score:** $^{\alpha}$ = Sensitivity and specificity results, $^{\beta}$ = AUC values, $^{\gamma}$ = Accuracy values, **Other:** $^{\#}$ = in the original article it was described as administrations instead of cycles, **Abbreviations:** AUC = area under the curve, NPV = negative predictive value, PPV = positive predictive value, SUV = standard uptake value, TLG = total lesion glycolysis, MATV = metabolic active tumour value.

Sens	Spec	Acc	NPV	PPV	AUC	Performance score
86%	63%	73%	84%	67%	0.86	$B^{(\alpha)}A^{(\beta)}B^{(\gamma)}$
86%	75%	80%	86%	75%	0.86	$B^{(\alpha)}A^{(\beta)}A^{(\gamma)}$
64%	83%	76%	79%	69%	0.73	$B^{(\alpha)}B^{(\beta)}B^{(\gamma)}$
83%	52%		84%	50%	0.70 (0.55-0.85)	$C^{(\alpha)}B^{(\beta)}$
					0.61 (0.33-0.89)	$B^{(\beta)}$
20%	100%		33%	100%		$D^{(\alpha)}$
					0.59 (0.34-0.85)	$C^{(\beta)}$
					0.41 (0.16-0.67)	$C^{(\beta)}$
17%	100%	29%	17%	100%		$D^{(\alpha)}D^{(\gamma)}$
Use of different parameters did not improve predictive value of ΔSUVmax						-
Association between BRI and largest tumour diameter was significant (p= 0.05)						-
59%	80%	62%	24%	95%	-	$C^{(\alpha)}B^{(\gamma)}$
38%	71%	44%,	20%	86%	-	$D^{(\alpha)}C^{(\gamma)}$
27%	88%	64%	65%	60%	-	$D^{(\alpha)}B^{(\gamma)}$
18%	76%	54%	59%	33%	-	$D^{(\alpha)}C^{(\gamma)}$

DISCUSSION

In view of the potential of response-guided NAC to improve breast cancer survival, we aimed to create an overview of current knowledge on imaging performance to monitor NAC according to breast cancer subtype.

Our results suggest that due to the differences in imaging performance across subtypes, personalizing the monitoring step of response-guided NAC based on these is of relevance. However after reviewing the 15 included articles, we revealed that there is a lack of evidence with enough statistical power to conclude on the preferred imaging technique per subtype. Although, we did identify studies reporting on the performance of MRI and ^{18}F-FDG-PET/CT specified to breast cancer subtypes, all studies were observational, showed a lot of inter study variability, and described only performance of one imaging modality. Thereby, our results should be seen as preliminary and thus be interpreted with caution. This information can nonetheless serve to pinpoint areas of further research.

In the ER-positive/HER2-negative subtype, the best performing technique was ^{18}F-FDG-PET/CT after 2 NAC cycles[26], while the use of DCE MRI was limited and insufficient[16,23,28]. Worth mentioning is that the performance of ^{18}F-FDG-PET/CT was even higher with the imaging measurement parameter ΔTLG and MATV, than with the standard ΔSUVmax[26,30]. However, the performance results of ^{18}F-FDG-PET/CT are based on thresholds that were derived after analysis which might have led to overestimation of these performance results.

In TNBC, ^{18}F-FDG-PET/CT showed also a good performance[24,25,31,32], in which the best results were found after 2 NAC cycles using cut-off value ≥-50% ΔSUVmax (performance:$A^{(\alpha)}A^{(\gamma)}$)[32]. The use of MRI seems also promising in this subtype, as size decrease showed a correlation with BRI[16].

In the overall HER2-positive group, ^{18}F-FDG-PET/CT showed promising results[19,24,33,34], especially after 2 NAC cycles, using cut-off value ≥-62% ΔSUVmax (performance: $B^{(\alpha)}A^{(\beta)}B^{(\gamma)}$)[33]. However the study of Gebhart et al., which splits this HER2-positive group by ER status, showed limited and insufficient performance results. This might be explained by the use of a lower cut-off value and a different monitoring interval compared to the other ^{18}F-FDG-PET/CT studies[35]. MRI showed also in the overall HER2 positive group an association between tumour size decrease and BRI[16]. Therefore, further investigation towards the performance of MRI during NAC in TNBC and HER2-positive breast cancer seems relevant.

Previous publications that described and reviewed literature on the assessment of response

by imaging during NAC in specific subtypes were in line with our findings. For instance, Lobbes and colleagues showed that MRI was more accurate in HER2-positive tumours than in HER2-negative tumours[41]. Humbert et al. and Groheux et al. presented a good performance of ^{18}F-FDG-PET/CT in HER2-positive breast cancer patients using the difference in SUV uptake[40,42]. Also in TNBC ^{18}F-FDG-PET/CT showed promising performance results by both ΔSUVmax and ΔTLG with AUC values of 0.86 and 0.88 respectively[42] and an overall accuracy of 75%[43]. For ^{18}F-FDG-PET/CT, the potential of ΔTLG as an imaging measurement parameter was confirmed by other research groups, who showed its correlation with survival[42,44]. In addition the use of absolute values of SUVmax and SUVpeak instead of their difference was suggested for its better performance in predicting pCR compared to ΔSUVmax[34,42,45]. Furthermore, besides ^{18}F-FDG-PET/CT and MRI, FES-PET/CT and DWI-MRI might be of interest for response prediction as well. Since, FES-PET/CT seems a useful tool for response prediction in ER-positive tumours[46], and DWI-MRI seems to provide complementary information to DCE-MRI[47]. Both are being investigated in a trial (NCT02398773;NCT01564368). Finally, two reviews addressed also the importance of reaching consensus about early NAC evaluation: the first was already published in 2012 and focused on early prediction of pathologic response on NAC by MRI, which did not specify to breast cancer subtypes. They pointed at the importance of reaching consensus on pCR definitions and thresholds for response definition, which shows the lack of progress in standardizing methodology over the years[6]. The second review, published in 2016, described a need for standardization when using ^{18}F-FDG PET/CT to evaluate response to NAC in BC patients on: monitoring interval, BC subtype and type of treatment[45].

With regard to the response-guided NAC approach, we identified two studies. One RCT for ER-negative/HER2positive in which patients were being scanned by ^{18}F-FDG-PET after 1 cycle of NAC and bevacizumab was randomly added to the first-line treatment in non-responders (≤ -70% ΔSUVmax) in a 2:1 ratio[48]. This study was however excluded from our analysis for using PET. The second study (uncontrolled) used MRI in ER-positive/HER2-negative patients to guide treatment regimen. In patients considered as non-responders a switch from AC to DC showed improved tumour size reduction.[23] Since the NPV value in this study was 10%, only 10% of the non-responders were correctly identified. If the study had used ^{18}F-FDG-PET/CT instead, according to our results, NPV could have been increased to 31% which consequently would have increased therapeutic response[26]. This is of course, assuming that 1) the switch to non-cross resistant therapy DC would be beneficial in non-responders to AC, 2) pCR would correlate to survival in this subtype, and 3) the optimal way to predict therapeutic response had been chosen. This hypothetical scenario illustrates that improved effectiveness of the response-guided NAC approach can be achieved with improved imaging performance, more effective treatments or the combination of both.

This review included a few studies, mainly underpowered, and of heterogeneous study

designs and outcome measures. Variability mainly occurred due to differences in interval time between imaging at baseline and monitoring, cut-off values to define treatment response, and pCR definitions, which are consequence of the lack of consensus on imaging settings and protocols. As we were aware of this, and of its possible influence on our results, we carefully described study differences in our results section. Regarding the cut-off values it should be mentioned that there is a difference in its use: some are set upfront which enable analysing sensitivity and specificity of this imaging modality under this cut-off value, when others were derived after analysis in order to identify the most optimal sensitivity and specificity which certainly could overestimate accuracy. Another issue is the identification of mainly studies examining ^{18}F-FDG-PET/CT and some on MRI. Although the performance of MRI in pathologic response prediction is often investigated, we were unable to include these articles, since they combined performance results of response assessment during and after NAC in their analysis. The lack of results on MRI in the majority of the subtypes made it impossible to compare its performance to ^{18}F-FDG-PET/CT and consequently to conclude on the preferred imaging technique per subtype.

The final issues that should be discussed are 1) the inclusion of studies only describing performance results according to one receptor status, as it is known that performance could be affected by the other unknown receptor status. 2) The quality assessment, since we included studies either describing a very small sample size (<20) or studies in which a switch in neoadjuvant treatment regimen has been made, could have biased the performance results and finally 3) the way pathologic complete response was used in the identified articles, since the value of reaching a pCR is different according to subtypes: firstly, some identified articles used a pCR definition for all subtypes (mainly chosen afterwards), some articles describe different pCR definitions per subtype. For instance in ER+/HER2- subtype is complete pathologic response rare, therefore complete and partial pathologic response are mainly being pooled. Secondly, the three identified pCR categories are not comparable because of varying inclusion of: response in both axillary lymph nodes and breast tissue, and absence of DCIS and invasive tumour cells (Table 1). Besides, in the ER-positive/HER2-negative group we did not differentiate into luminal A and B tumours, despite knowing that in luminal A tumours pCR does not correlate with survival[9]. Therefore, our conclusions for this subtype may be unlikely. Nonetheless, they serve to illustrate the urgency to reach consensus for a reliable alternative for pCR in this subgroup.

A major limitation of an analysis as presented here, is the small fraction and the insufficient statistical power of the included studies. It shows however, what is needed to decide on the most effective imaging technique per subtype; consensus on several aspects that affect study comparability. Specifically, on 1) the definition for pathologic response, 2) the thresholds to define complete-, near-, partial-, or no- response during NAC in both ^{18}F-FDG-PET/CT and

MRI, 3) the required interval time between baseline and response monitoring, per subtype and imaging technique, and 4) imaging settings. Only then, meaningful well-designed studies which account for various breast cancer subtypes and imaging techniques can be conducted. Whereupon, RCTs such as the AVATACXER trial[48], which mimics the response-guided NAC approach, could be set to also inform on suitable treatment switches. Further, we suggest to conduct further research to: 1) less investigated techniques such as FES-FDG/PET and DWI-MRI, 2) potential predictive biomarkers that could further personalize the response-guided NAC approach i.e. Ki67 and P53 and 3) the association between NAC treatments and imaging performance. Finally, a cost-effectiveness analysis could be interesting to explore the health-economic consequences of the various scenarios for this response-guided NAC approach.

This literature review is unique in the way that it focuses on imaging performance of NAC monitoring specified to breast cancer subtypes. We conclude that current evidence is too low to draw on subtype-specific imaging recommendations, and that these can only occur when consensus on imaging settings and work regulations are reached. Thus, further research on these are necessary to eventually build protocols and use them to conceive comparable study outcomes.

REFERENCES

1. WHO. Breast Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012. Available at: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. (Accessed: 11th November 2015)
2. Mauri, D., Pavlidis, N. & Ioannidis, J. P. A. Neoadjuvant versus adjuvant systemic treatment in breast cancer: A meta-analysis. *J. Natl. Cancer Inst.* **97**, 188–194 (2005).
3. Fisher, B. *et al.* Effect of preoperative chemotherapy on local-regional disease in women with operable breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-18. *J. Clin. Oncol.* **15**, 2483–2493 (1997).
4. van der Hage, J. A. *et al.* Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. *J. Clin. Oncol.* **19**, 4224–4237 (2001).
5. Gralow, J. R. *et al.* Preoperative therapy in invasive breast cancer: Pathologic assessment and systemic therapy issues in operable disease. *J. Clin. Oncol.* **26**, 814–819 (2008).
6. Marinovich, M. L. *et al.* Early prediction of pathologic response to neoadjuvant therapy in breast cancer: Systematic review of the accuracy of MRI. *Breast* **21**, 669–677 (2012).
7. Kaufmann, M. *et al.* Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: an update. *J. Clin. Oncol.* **24**, 1940–1949 (2006).
8. Chollet, P. *et al.* Prognostic significance of a complete pathological response after induction chemotherapy in operable breast cancer. *Br. J. Cancer* **86**, 1041–1046 (2002).
9. Von Minckwitz, G. *et al.* Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J. Clin. Oncol.* **30**, 1796–1804 (2012).
10. von Minckwitz, G. *et al.* Response-guided neoadjuvant chemotherapy for breast cancer. *J. Clin. Oncol.* **31**, 3623–3630 (2013).
11. Yeh, E. *et al.* Prospective comparison of mammography, sonography, and MRI in patients undergoing neoadjuvant chemotherapy for palpable breast cancer. *Am. J. Roentgenol.* **184**, 868–877 (2005).
12. Hamisa, M., Dabess, N., Yosef, R., Zakeria, F. & Hammed, Q. Role of breast ultrasound, mammography, magnetic resonance imaging and diffusion weighted imaging in predicting pathologic response of breast cancer after neoadjuvant chemotherapy. *Egypt. J. Radiol. Nucl. Med.* **46**, 245–257 (2015).
13. Londero, V. *et al.* Locally advanced breast cancer: comparison of mammography, sonography and MR imaging in evaluation of residual disease in women receiving neoadjuvant chemotherapy. *Eur Radiol* **14**, 1371–1379 (2004).
14. Wu, L.-M. *et al.* Can diffusion-weighted MR imaging and contrast-enhanced MR imaging precisely evaluate and predict pathological response to neoadjuvant chemotherapy in patients with breast cancer? *Breast Cancer Res. Treat.* **135**, 17–28 (2012).
15. Cheng, X. *et al.* 18F-FDG PET/CT and PET for evaluation of pathological response to neoadjuvant chemotherapy in breast cancer: a meta-analysis. *Acta Radiol.* **53**, 615–27 (2012).
16. Loo, C. E. *et al.* Magnetic resonance imaging response monitoring of breast cancer during neoadjuvant chemotherapy: Relevance of Breast Cancer Subtype. *J. Clin. Oncol.* **29**, 660–666 (2011).
17. Hayashi, Y. *et al.* Analysis of complete response by MRI following neoadjuvant chemotherapy predicts pathological tumor responses differently for molecular subtypes of breast cancer. *Oncol. Lett.* **5**, 83–89 (2012).
18. Ko, E. S. *et al.* Analysis of factors that influence the accuracy of magnetic resonance imaging for predicting response after neoadjuvant chemotherapy in locally advanced breast cancer. *Ann. Surg. Oncol.* **20**, 2562–2568 (2013).
19. Humbert, O. *et al.* Changes in 18F-FDG tumor metabolism after a first course of neoadjuvant chemotherapy in breast cancer: influence of tumor subtypes. *Ann. Oncol.* **23**, 2572–2577 (2012).
20. Rodenhuis, S. *et al.* A simple system for grading the response of breast cancer to neoadjuvant chemotherapy. *Ann. Oncol.* **21**, 481–487 (2010).
21. Symmans, W. F. *et al.* Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J. Clin. Oncol.* **25**, 4414–4422 (2007).
22. Bossuyt, P. M. *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Vet. Clin. Pathol.* **36**, 8–12 (2007).

23. Rigter, L. S. *et al.* Neoadjuvant chemotherapy adaptation and serial MRI response monitoring in ER-positive HER2-negative breast cancer. *Br. J. Cancer* **109**, 2965–2972 (2013).
24. Koolen, B. B. *et al.* Sequential (18)F-FDG PET/CT for early prediction of complete pathological response in breast and axilla during neoadjuvant chemotherapy. *Eur. J. Nucl. Med. Mol. Imaging* **41**, 32–40 (2014).
25. Koolen, B. B. *et al.* FDG PET/CT during neoadjuvant chemotherapy may predict response in ER-positive/HER2-negative and triple negative, but not in HER2-positive breast cancer. *Breast* **22**, 691–7 (2013).
26. Groheux, D. *et al.* Estrogen receptor-positive/human epidermal growth factor receptor 2-negative breast tumors: Early prediction of chemosensitivity with 18F-fluorodeoxyglucose positron emission tomography/computed tomography during neoadjuvant chemotherapy. *Cancer* **119**, 1960–1968 (2013).
27. Zucchini, G. *et al.* Potential utility of early metabolic response by 18F-2-fluoro-2-deoxy-D-glucose-positron emission tomography/computed tomography in a selected group of breast cancer patients receiving preoperative chemotherapy. *Eur. J. Cancer* **49**, 1539–45 (2013).
28. Charehbili, A. *et al.* Accuracy of MRI for treatment response assessment after taxane- and anthracycline-based neoadjuvant chemotherapy in HER2-negative breast cancer. *Eur. J. Surg. Oncol.* **40**, 1216–1221 (2014).
29. Martoni, A. A. *et al.* Early (18)F-2-fluoro-2-deoxy-d-glucose positron emission tomography may identify a subset of patients with estrogen receptor-positive breast cancer who will not respond optimally to preoperative chemotherapy. *Cancer* **116**, 805–813 (2010).
30. Hatt, M. *et al.* Comparison between 18F-FDG PET image-derived indices for early prediction of response to neoadjuvant chemotherapy in breast cancer. *J. Nucl. Med.* **54**, 341–9 (2013).
31. Groheux, D. *et al.* Triple-negative breast cancer: early assessment with 18F-FDG PET/CT during neoadjuvant chemotherapy identifies patients who are unlikely to achieve a pathologic complete response and are at a high risk of early relapse. *J. Nucl. Med.* **53**, 249–254 (2012).
32. Groheux, D. *et al.* Early assessment with 18F-fluorodeoxyglucose positron emission tomography/computed tomography can help predict the outcome of neoadjuvant chemotherapy in triple negative breast cancer. *Eur. J. Cancer* **50**, 1864–1871 (2014).
33. Groheux, D. *et al.* HER2-overexpressing breast cancer: FDG uptake after two cycles of chemotherapy predicts the outcome of neoadjuvant treatment. *Br. J. Cancer* **109**, 1157–64 (2013).
34. Humbert, O. *et al.* HER2-positive breast cancer: (1)(8)F-FDG PET for early prediction of response to trastuzumab plus taxane-based neoadjuvant chemotherapy. *Eur. J. Nucl. Med. Mol. Imaging* **41**, 1525–1533 (2014).
35. Gebhart, G. *et al.* 18F-FDG PET/CT for early prediction of response to neoadjuvant lapatinib, trastuzumab, and their combination in HER2-positive breast cancer: results from Neo-ALTTO. *J Nucl Med* **54**, 1862–1868 (2013).
36. Humbert, O. *et al.* Prognostic relevance at 5 years of the early monitoring of neoadjuvant chemotherapy using 18F-FDG PET in luminal HER2-negative breast cancer. *Eur. J. Nucl. Med. Mol. Imaging* **41**, 416–427 (2014).
37. Chevallier, B., Roche, H., Olivier, J. P., Chollet, P. & Hurteloup, P. Inflammatory breast cancer. Pilot study of intensive induction chemotherapy (FEC-HD) results in a high histologic response rate. *Am. J. Clin. Oncol.* **16**, 223–228 (1993).
38. Ogston, K. N. *et al.* A new histological grading system to assess response of breast cancers to primary chemotherapy: Prognostic significance and survival. *Breast* **12**, 320–327 (2003).
39. Sataloff, D. M. *et al.* Pathologic response to induction chemotherapy in locally advanced carcinoma of the breast: a determinant of outcome. *J. Am. Coll. Surg.* **180**, 297–306 (1995).
40. Humbert, O. *et al.* Role of Positron Emission Tomography for the Monitoring of Response to Therapy in Breast Cancer. *Oncologist* **20**, 94–104 (2015).
41. Lobbes, M., Prevos, R. & Smidt, M. Response monitoring of breast cancer patients receiving neoadjuvant chemotherapy using breast MRI – a review of current knowledge. *J. cancer Ther. Res.* **1**, (2012).
42. Groheux, D. *et al.* Early Metabolic Response to Neoadjuvant Treatment: FDG PET/CT Criteria according to Breast Cancer Subtype. *Radiology* **277**, 358–371 (2015).
43. Humbert, O. *et al.* Identification of Biomarkers Including 18FDG-PET/CT for Early Prediction of Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Clin. Cancer Res.* **21**, 5460–5468 (2015).
44. Ulaner, G. a *et al.* Prognostic value of quantitative fluorodeoxyglucose measurements in newly diagnosed metastatic breast cancer. *Cancer Med.* **2**, 725–33 (2013).

45. Groheux, D., Mankoff, D., Espié, M. & Hindié, E. 18F-FDG PET/CT in the early prediction of pathological response in aggressive subtypes of breast cancer: review of the literature and recommendations for use in clinical trials. *Eur. J. Nucl. Med. Mol. Imaging* **43**, 983–993 (2016).

46. Currin, E., Linden, H. M. & Mankoff, D. A. Predicting breast cancer endocrine responsiveness using molecular imaging. *Curr. Breast Cancer Rep.* **3**, 205–211 (2011).

47. O'Flynn, E. a M. & DeSouza, N. M. Functional magnetic resonance: biomarkers of response in breast cancer. *Breast Cancer Res.* **13**, 204 (2011).

48. Coudert, B. *et al.* Use of [(18)F]-FDG PET to predict response to neoadjuvant trastuzumab and docetaxel in patients with HER2-positive breast cancer, and addition of bevacizumab to neoadjuvant trastuzumab and docetaxel in [(18)F]-FDG PET-predicted non-responders (AVATAXHER): *Lancet. Oncol.* **15**, 1493–1502 (2014).

SUPPLEMENTARY MATERIAL

Supplement 1 - Methods section: systematic search strategy

Database	**PubMed**
Time span	from January 2000 until March 2015
Search in	Title and abstract
Category	**Keywords**
"Breast cancer"	breast neoplasms[mesh] OR breast neoplasm OR breast cancer OR breast tumour OR breast tumor OR breast malignan
"Imaging"	diagnostic imaging[mesh] OR imaging* OR MRI OR magnetic resonance imaging OR PET OR PET/CT OR PET-CT OR ultrasonograph* OR mammograph* OR PET/MRI OR PET-MRI OR positron emission tomograph* OR computed tomograph* OR image OR images
"Neo adjuvant therapy"	neoadjuvant therapy[mesh] OR preoperative therapy[MeSH] OR ((neoadjuvant therapy[mesh] OR neo-adjuvant OR neoadjuvant) AND (neoadjuvant therapy[mesh] OR preoperative therapy[MeSH] OR ((neoadjuvant therapy[mesh] OR neo-adjuvant OR neoadjuvant) AND (chemo OR chemotherap* OR chemo therap*)) OR ((pre-operative OR preoperative) AND (chemo OR chemotherap* OR chemo therap*)
"Outcome"	disease-free survival[mesh] OR surviv* OR survival rate[mesh] OR survival analysis[mesh] OR effective* OR cost-effective* OR treatment response* OR treatment outcome[mesh] OR complete pathologic response* OR complete pathological response* OR pathologic complete response* OR pathological complete response* OR pathologic response OR Ki67 OR Ki-67 OR MKI67
"Breast cancer subtype"	HER2 positive OR HER2/neu positive OR HER2neu positive OR HER2-neu positive OR non-luminal OR ((human epidermal growth factor receptor 2 OR receptor, erbB-2 [mesh] OR receptor, epidermal growth factor [mesh]) AND (positive)) OR (estrogen receptor-positive OR hormone receptor-positive OR estrogen receptor-positive OR oestrogen receptor-positive OR ER-positive OR hormone positive OR positive hormone receptor OR positive estrogen) OR Luminal OR triple negative OR TN OR TNBC OR ER-negative PR-negative HER2-negative OR basal-like OR basal like

Supplement 2 - Results section: study characteristics

Table supplement 2 - Results section: study characteristics.

Author, year	Sample size per subtype	Age	Study design	Enrolled	Clinical stage	Monitoring technique	Monitoring interval
Charehbili, 2014	HER2 – (194) (ER+ (187); ER- (35))	Mean 49 years; postmenopausal (88); premenopausal (146)	Retrospectively	July 2010 – April 2012	II and III, T1:(2), T2:(128), T3/4:(92), N-:(99), N+:(123)	DCE MRI 1.5 and 3.0 T	Baseline and after three cycles
Gebhart, 2013	HER2+/ HR- (43) HER2+ / HR+ (34)	-	Prospective	Jan 2008 – May 2010	Metabolic lymph nodes (52) and distant lesions (9)	FDG-PET/CT	Baseline, week 2 and 6
Groheux, 2012	TN(20)	-	Prospective	Enrolled within 30 months	II (9) and III (11)	FDG-PET/CT	Baseline, after two cycles
Groheux, 2012	ER+/HER2 - (64)	Mean: 52; postmenopausal (41); Premenopausal: (22)	Prospective	July 2007 to Oct 2011	T1(1), T2(21), T3(25), T4(17) ;N0(24), N1(29), N2(8), N3(5)	FDG-PET/CT	Baseline, after two cycles
Groheux , 2013	HER2+ (30)	-	Prospective	-	II:(14) and III:(16)	FDG-PET/CT	Baseline, after two cycles
Groheux, 2014	TN (50)	-	Prospective	Nov 2007 to Sept 2012	II:(21) and III:(29)	FDG PET/CT	Baseline, after two cycles

Neoadjuvant therapy	Response definition monitoring	pCR definition (category)	pCR rate	AUC (95% CI)	Sens, Spec, NPV, PPV, Accuracy	Setting imaging
TAC with (107) or without (R) (115) zoledronic acid	>30% decrease of tumour size	Miller-Payne grade 5 or ypT0/is; III	17%	0.55 (0.45-0.65)	37%, 87%, 22%, 93%, 45%	DCE-MRI; 1.5 and 3.0T
(R) Lapatinib or Trastuzumab or both. All received paclitaxel	After 2 weeks ≥ 15% reduction of SUVmax; after 6 weeks ≥ 25%	Absence of invasive cancer in the breast; III	61% 18%	-	After 2 weeks: 27%, 88%, 65%, 60%, 64% After 2 weeks: 38%, 71%, 20%, 86%, 44%	GE / Philips or Siemens PET/CT; fasted 6h before injection; 3.7-7.4MBq/kg; scan at least 50 min after injection; same scanner and parameters in each institution
EC-D (14) or SIM (6)	≥-42% ΔSUVmax and ≥-50% ΔSUVmax	No evidence of residual invasive cancer in both breast tissue and lymph nodes; II	30%	ΔSUV = 0.88	≥-42% ΔSUVmax 64%, 100%, 55%, 100%, 75%	Gemini XL PET/CT; fasted 6h before injection; scan 60 min after injection; 5MBq/kg; CT: 120kV; 100mAs; 16 slices; 2 min per bed position
EC-D	≥- 38% ΔSUVmax and ≥-71% ΔTLG	Sataloff TA-TB; NA-NB-NC considered as responder and partial responder; IIII	6%	ΔSUVmax 0.73; ΔTLG 0.81	ΔSUVmax: 62%, 78%, 12%, 98%,-	Gemini XL Philips; fasted 6h before; scan 60 min after injection: 5MBq/kg; 2 min per bed position; CT:120 kV; 100mAs;
EC-D and trastuzumab	Reduction ≥ 62% ΔSUVmax	No residual invasive disease in tumour and lymph nodes; II	53%	ΔSUVmax = 0.86	86%, 75%, 86%, 75%, 80%	Gemini XL PET/CT; fasted 6h before injection; scan 60 min after injection: 5MBq/kg; CT: 120kV; 100mAs; 2 min per bed position
EC-D (20) or SIM (30)	≥-42% ΔSUVmax and ≥-50% ΔSUVmax	No evidence of residual invasive cancer in breast tissues and lymph nodes; II	38%	ΔSUVmax 0.80 for EC-D and 0.86 for SIM	≥-42% ΔSUVmax: 58%, 100%, 59%, 100%, 74%	Gemini XL PET/CT; Fasted 6h before injection; scan started after 60 min after injection; 5MBq/kg; from mid-thigh to skull with arms raised; resolution (3D): 4x4x4 mm^3 CT: 16 slices; 120kV; 100mAs; 2 min per position

(continued on next page)

Table supplement 2 (continued) - Results section: study characteristics.

Author, year	Sample size per subtype	Age	Study design	Enrolled	Clinical stage	Monitoring technique	Monitoring interval
Hatt, 2013	TN(13);	-	Retrospective	July 2007-May 2009	II:(24) and III:(27)	FDG PET/CT	Baseline, after two cycles
	ER+/HER2-(26)						
	HER2+(12)						
Humbert, 2012	TN (25)	≤50 (61) and >50 (54); mean: 51 years	Prospective	-	T1-2:(62), T3:(42), N-:(35), N+:(79)	FDG PET/CT	Baseline and just before second course NACT
	ER+/HER2-(53)						
	HER2+(37)						
Humbert, 2014	HER2+(57) Majority ER positive	≤50 (36) and >50 (21); postmenopausal (21); premenopausal (35)	Prospective	Nov 2006 – Oct 2012	I and II:(26), III:(28)	FDG PET/CT	Baseline and after first course NAC
Koolen, 2013	ER+/HER2-(45)	Median:47 (range: 25-68)	Retrospective	Since Sept 2008	T1:(8), T2:(59), T3:(24), T4:(7), N0:(14), N1:(57), N2:(2), N3:(25)	FDG PET/CT	Baseline and after first course of NAC
	HER2+(25)						
	TN (25)						

Neoadjuvant therapy	Response definition monitoring	pCR definition (category)	pCR rate	AUC (95% CI)	Sens, Spec, NPV, PPV, Accuracy	Setting imaging
EC-D and in HER2+ EC-D plus trastuzumab	Optimal cut-off values: ΔSUVmax: -48% ΔTLG: -56% ΔMATV: -42%	Staloff scale: TA-B with NABC are considered as responder and partial responder; IIII	23%	Use of different parameters did not improve predictive value of SUVmax		Gemini XL Philips; fasted 6h before injection; 5MBq/kg; after 60 min mid-thigh to skull with arms raised; resolution: 4x4x4; CT: 16 slices; 120kV; 100mAs;
			0%	ΔSUVmax: 0.88 SUVpeak: 0.84 ΔSUVmean: 0.69 ΔTLG: 0.96 ΔMATV: 0.98		
			33%	Use of different parameters did not improve predictive value of SUVmax		
FEC 100 (25); FEC 100 plus docetaxel (39); Docetaxel followed by Epirubicin and docetaxel (8); CEX (6)		Chevallier's classification grade 1 and 2; II	36%	No correlation between early metabolic and final pathological response		C-PET Plus scanner and Gemini GXL scanner; fasted 6h before injection of F-FDG; whole body scan 60 min after injection; 2MBq/kg (C-PET)and 5MBq/kg (Gemini); Prone position started 80-90 min after administration
			1.9%	-		
TH +/- carboplatin (37)	ΔSUVmax of -75%		38%	0.73	64%, 83%, 79%, 69%, 76%	
TH	ΔSUVmax ≥ 60%	No residual invasive cancer in the breast and nodes though in-situ breast residuals were allowed (ypT0/is ypN0); II	44%	AUC: 0.70 (0.55-0.85)	83%, 52%, 84%, 50%,-	Gemini GXL and TF Philips; fasted 6 hours before injection:5MBq/kg (GXL) 3.5MBq/kg (TP); brain to mid-thigh after 60 min; prone position after 90 min
AC (48); AC-CTC (4); AC-CD (20); CD (1); PTC (25)	Change in FDG uptake	Complete absence of residual tumour cells at microscopy, irrespective of DCIS; III	11%	0.77 (0.68-0.87)		Gemini TF Philips, Fastes 6 h before injection; 180-240MBq depending on BMI; scanning after +/- 70min; hanging breast method; 3.0min per bed position; resolution: 2x2x2mm CT: low dose; 40mA s, 2mm slices
			68%	0.41 (0.16-0.67)		
			61%	0.85 (0.69-1.00)		

(continued on next page)

Table supplement 2 (continued) - Results section: study characteristics.

Author, year	Sample size per subtype	Age	Study design	Enrolled	Clinical stage	Monitoring technique	Monitoring interval
Loo, 2011	ER+/ HER2- (103) HER2+ (38) TN (47)	Mean: 46 (range: 23-76)	Retro-spective	Between 2000- 2008	T1:(6), T2:(97), T3:(62), T4:(23) N0:(28), N1:(125), N3:(11), Nx:(24)	DCE MRI 1.5T or 3.0T	Baseline and after three courses NAC and in HER2+ after eight administra-tions (first cycle)
Martoni, 2010	ER+/ HER2-: (16) HER2+: (7) TN (9)	Median:48 years (31-72)	Prospec-tive	-	II:(15), III:(13), IV:(6)	FDG PET/CT	Baseline and after second and fourth cycle
Rigter, 2013	ER+/ HER2- (246)	Median 48 (range 18-68)	Retro-spective	Oct 2004 – March 2012	T1:(21), T2:(91), T3:(43) T4:(9); Na:(49), Nb:(40), Nc:(50), Nd:(98), Ne:(9)	DCE MRI 1.5T or 3.0T	Baseline after three and six courses
Zucchini , 2013	ER+/ HER2-(31) HER2+ (14) TN (15)	Median: 49 years	Prospec-tive	July 2004 – March 2011	II:(30), III(23), IV:(7)	FDG PET/CT	Baseline and after second PCT cycle

Abbreviations: R = randomized, CI = confidence interval, NS = not specified, SUV = standardized uptake value, pCR = pathologic complete response, AUC = area under receiver operating curve, AC = doxorubicin and cyclophosphamide, CD = capecitabine and docetaxel, CTC = cyclophosphamide, thiotepa, carboplatin, PTC = paclitaxel, trastuzumab, carboplatin, TAC = doxorubicin followed by cyclophosphamide and docetaxel, TCaH = taxol, carboplatin, herceptin. AbCaH = abraxane, carboplatin, Herceptin, AbCaAv = abraxane, carboplatin, avastin, TCA = taxol, carboplatin, FEC = fluorouracil, epirubicin and cyclophosphamide, EC-D = epirubicin, cyclophosphamide followed by docetaxel, SIM = epirubicin and cyclophosphamide (1200 mg/m^2), TH = docetaxel and trastuzumab.

Neoadjuvant therapy	Response definition monitoring	pCR definition (category)	pCR rate	AUC (95% CI)	Sens, Spec, NPV, PPV, Accuracy	Setting imaging
AC (90); AC – CD (45); CD or AD (15); Trastuzumab based (38)	Change in largest diameter	Complete absence of residual tumour cells or small number of scattered cells at microscopy; IIII	7%	No association between residual tumour and change in largest diameter		Magnetom Vision scanner 1.5T; 3.0T Philips Achieva scanner; prone position; breast coil; gadolinium (14ml/0.1mmol/kg); 5 series at 90s interval; FOV: 310 (1.5T); 360 (3.0T)
			40%	Residual tumour after NAC associated with change in largest diameter (p<0.05)		
			34%	Residual tumour after NAC associated with change in largest diameter (p<0.001)		
Anthracycline based and taxane based PCT	≥-50% ΔSUVmax	Miller and Payne; 4 and 5 with NRG A and D; IIII	19%	-	After 2nd cycle 38%, 100%, 27%, 100%, 50%	GE medical system; Discovery LS; Fasted 6h before scanning; scan after 60-70 min after injection; 5.3MBq/kg; 4min per bed position; CT: 120kV 60mA. Slices 4 a 5 mm thick
			14%		After 2nd cycle 17%, 100%, 17%, 100%, 29%	
			33%		After 2nd cycle 0%, 100%, 33%,-, 33%	
6 x ddAC (164); 3 x ddAC – 3 x DC (82)	Difference in largest diameter	ypT0/is ypN0 /+ ypT0/is ypN0 and ypT0ypN0; III	3%	-	35%, 89%, 10%, 98%, 39%	Magnetom Vision scanner 1.5T; 3.0 T Philips Achieva scanner; prone position; breast coil; gadolinium (14ml/0.1mmol/kg); 5 series at 90s interval; FOV: 310 (1.5T); 360 (3.0T)
6 x Anthracycline taxane regimen (9); 8 x Anthracycline taxane regimen (45) 4-8 x taxane and trastuzumab (6)	≥-50% ΔSUVmax	Miller and Payne; TRG 4 and 5 with NRG A and D; IIII	16%	-	38%, 100%, 24%, 100%,-	GE medical system; Discovery LS; Fasted 6h before scanning; scan after 60-70 min after injection; 5.3MBq/kg; 4min per bed position; CT: 120kV 60mA. Slices 4 a 5 mm thick
			29%		20%, 100%, 33%, 100%,-	
			27%		0%, 100%, 27%, 0%,-	

3

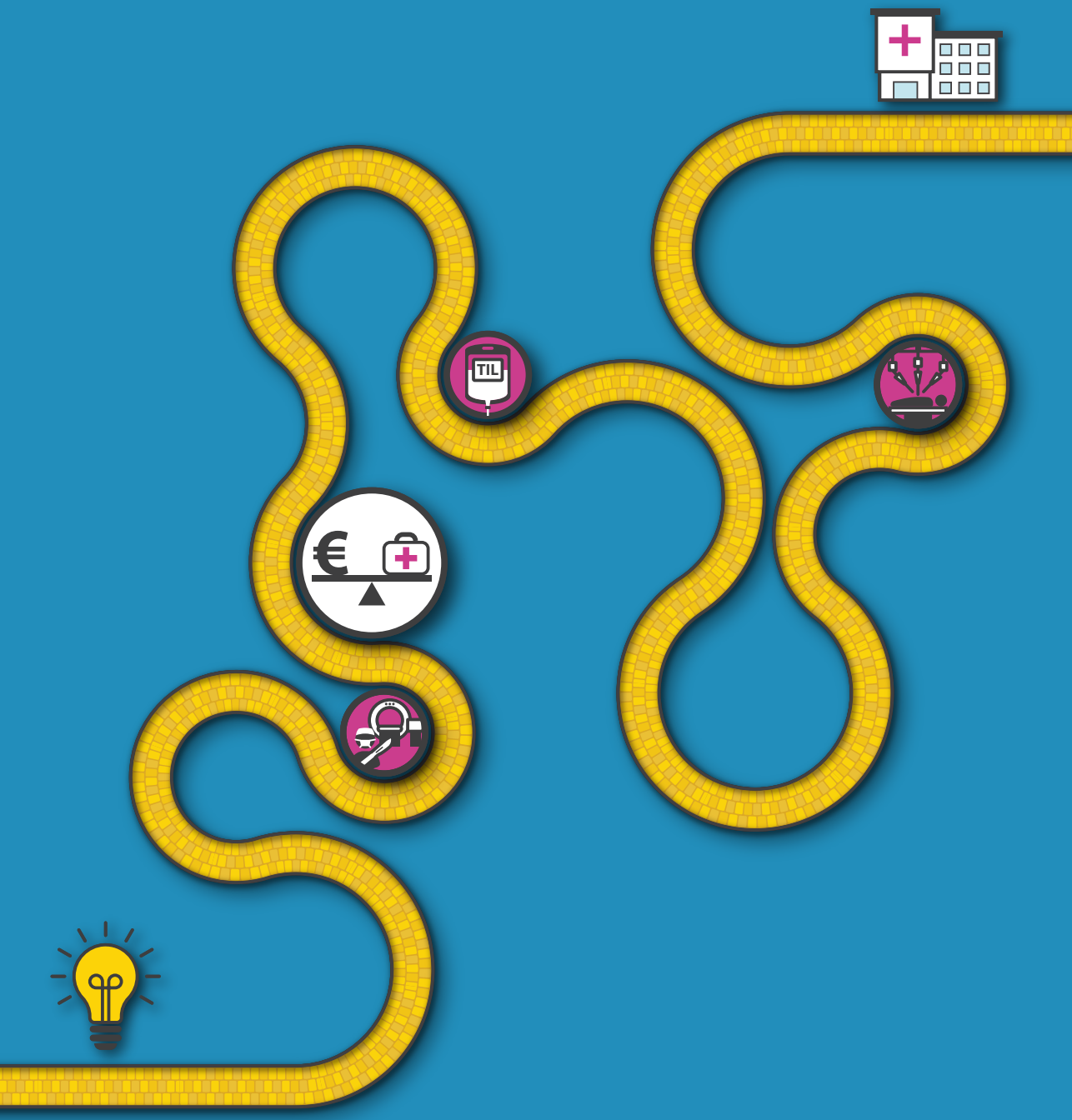
Supplement 3 - Results section: QUADAS criteria

Author	Year	1. Respresentative spectrum of patients	2. Selection criteria clear	3. Reference standard reliable	4. Time interval between MRI/PET and pathology	5. Whole or random sample received verification	6. Same reference standard	7. Reference standard independent of the index test
Charehbili	2014	yes	yes	yes	yes	yes, but not random	yes	yes
Gebhart	2013	yes	yes	yes	yes	yes, but not random	yes	yes
Gorheux	2013	yes	yes	yes	unclear	yes, but not random	yes	yes
Groheux	2012	yes	yes	yes	unclear	yes, but not random	yes	yes
Groheux	2013	yes	yes	yes	unclear	yes, but not random	yes	yes
Groheux	2014	yes	yes	yes	unclear	yes, but not random	yes	yes
Hatt	2013	yes	yes	yes	unclear	yes, but not random	yes	yes
Humbert	2012	yes	yes	yes	yes	yes, but not random	yes	yes
Humbert	2014	yes	yes	yes	yes	yes, but not random	yes	yes
Koolen	2014	yes	yes	yes	unclear	yes, but not random	yes	yes
Koolen	2013	yes	yes	yes	unclear	yes, but not random	yes	yes
Loo	2011	yes	yes	yes	unclear	yes, but not random	yes	yes
Martoni	2010	yes	yes	yes	yes	yes, but not random	yes	yes
Rigter	2013	yes	yes	yes	unclear	yes, but not random	yes	yes
Zucchini	2013	yes	yes	yes	yes	yes, but not random	yes	yes

8. Description execution of MRI/PET	9. Description execution of pahtology	10. Interpretation of PET/MRI blinded from reference	11. Interpretation of reference blinded from MRI/PET	12. Same clinical data available	13. Uninterpretable test results reported	14. Withdrawals explains	Score (max = 14; yes = 1)
yes	yes	yes	unclear	yes	yes	unclear	12
yes	yes	yes	unclear	yes	yes	yes	13
yes	yes	yes	unclear	yes	unclear	unclear	10
yes	yes	yes	unclear	yes	unclear	unclear	10
yes	yes	yes	unclear	yes	unclear	unclear	10
yes	yes	yes	unclear	yes	unclear	unclear	10
yes	yes	unclear	unclear	yes	unclear	unclear	9
yes	yes	yes	unclear	yes	unclear	unclear	11
yes	yes	unclear	unclear	yes	yes	yes	12
yes	yes	unclear	unclear	yes	unclear	unclear	9
yes	yes	yes	unclear	yes	yes	yes	12
yes	yes	yes	unclear	yes	unclear	unclear	10
yes	yes	yes	unclear	yes	unclear	unclear	11
yes	yes	yes	unclear	yes	yes	unclear	11
yes	yes	yes	unclear	yes	unclear	unclear	11

PART II

Early HTA: up to and including the first clinical studies (phase I)


TIL

4

Early budget impact analysis on magnetic seed localization for non-palpable breast cancer surgery

Melanie Lindenberg
Anne van Beek
Valesca Retèl
Frederieke van Duijnhoven
Wim H. van Harten

PLoS ONE. 2020; 15 (5); e0232690

ABSTRACT

Background

Current localization techniques used in breast conserving surgery for non-palpable tumors show several disadvantages. Magnetic seed localization (MSL) is an innovative localization technique aiming to overcome these disadvantages. This study evaluated the expected budget impact of adopting MSL compared to standard of care.

Methods

Standard of care with wire-guided localization (WGL) and radioactive seed localization (RSL) use was compared with a future situation gradually adopting MSL next to RSL or WGL from a Dutch national perspective over 5 years (2017-2022). The intervention costs for WGL, RSL and MSL and the implementation costs for RSL and MSL were evaluated using activity-based costing in eight Dutch hospitals. Based on available list prices the price of the magnetic seed was ranged €100-€500.

Results

The intervention costs for WGL, RSL and MSL were respectively: €2,617, €2,834 and €2,662 per patient and implementation costs were €2,974 and €26,826 for MSL and RSL respectively. For standard of care the budget impact increased from €14.7m to €16.9m. Inclusion of MSL with a seed price of €100 showed a budget impact of €16.7m. Above a price of €178 the budget impact increased for adoption of MSL, rising to €17.6m when priced at €500.

Conclusion

MSL could be a cost-efficient localization technique in resecting non-palpable tumors in the Netherlands. The online calculation model can inform adoption decisions internationally. When determining retail price of the magnetic seed, cost-effectiveness should be considered.

INTRODUCTION

Breast-conserving surgery (BCS) for non-palpable tumors requires appropriate localization technologies to resect the malignancy effectively[1,2]. Currently, mainly two localization technologies are used in the Netherlands: wire-guided localization (WGL) and radioactive seed localization (RSL). RSL aimed to overcome challenges in the use of WGL: challenging hospital planning, potential wire migration, and unfavorable incision placement[3,4]. RSL was shown to be at least non-inferior to WGL on important outcome measures such as re-excision rates and positive surgical margins[5–9]. Moreover, in some studies, RSL has demonstrated improved patient convenience[10–12] and greater ease of use during surgery[12].

RSL however has a considerable disadvantage as its radioactive nature requires adherence to strict nuclear safety regulations[13,14]. This results in a complex implementation process and substantial upfront costs which may explain the relative slow adoption of RSL. In addition, the treatment process may be affected due to the time limitation for an iodine seed to remain in situ (e.g. in the US). To overcome these challenges but retain the advantages of RSL, non-radioactive technologies such as magnetic seed localization (MSL) have been developed. It has been shown that MSL is safe and effective in localizing and excising non-palpable breast tumors[15,16]. Therefore MSL seems to be a realistic alternative for RSL and WGL.

A recent study compared WGL with MSL and concluded that WGL was equally effective as MSL[17]. Several studies have shown that RSL is not superior to WGL in clinical outcomes[5,18]. Therefore, it has been hypothesized that its effectiveness is similar to that of RSL and WGL. When the effectiveness of all three localization modalities are comparable, the widespread adoption of MSL depends on superiority on other aspects such as financial impact and usability.

This study aims to inform the adoption decision of MSL by evaluating the financial impact of gradually adopting MSL as a localization technology for guiding breast conserving tumor excision in the Netherlands health system compared to standard of care (SoC) by means of a Budget Impact Analysis (BIA) incorporating treatment and implementation costs. Secondly, a threshold analysis was conducted to estimate the maximum price level for MSL to become the most cost efficient technology. Finally, the BIA model was made available in a tool to enable translation of the results to other countries or specific hospital settings (Supplement 4).

METHODS

Budget impact analysis

For the analysis, the BIA framework of the International Society for Pharmacoeconomics and Outcomes Research was followed[19]. The analysis was conducted from a Dutch population perspective using a 5-year time horizon (2017-2022). In the Netherlands, RSL and WGL accounted for over 90% of current localization techniques, therefore these were assumed to be the only localization techniques in the current situation[20]. The BI model compares SoC in which both RSL and WGL are used in its present relative "market shares"[20] and a future situation in which MSL is gradually being adopted over time by the Dutch hospitals (Figure 1). The interventions are described in Box 1. Key inputs for the BI model were: size of the target population, utilization of the localization technologies, intervention costs, and the yearly implementation costs for hospitals transferring to either RSL or MSL. The implementation costs were calculated over the first five years that a localization technique is used, starting in the year before its adoption, meaning the first time a technology is used in clinical practice.

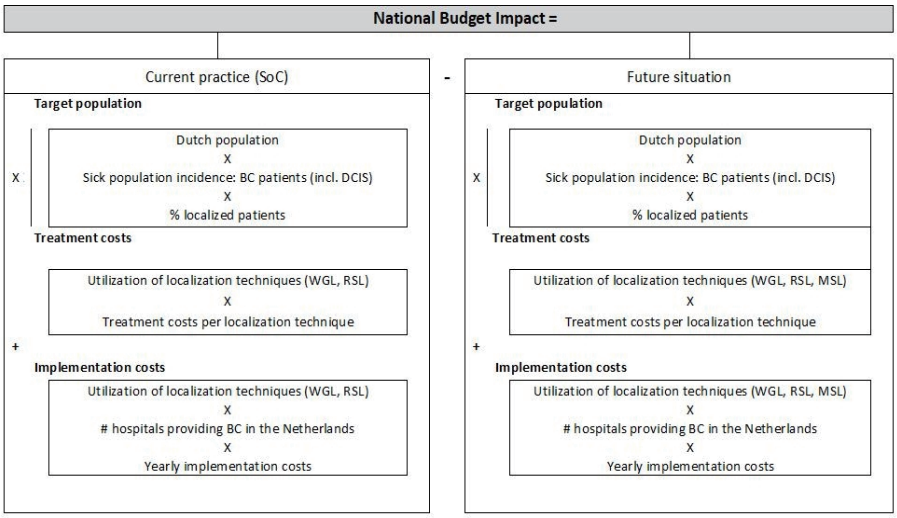


Figure 1. Structure of the Budget Impact model. This national BIA model compares the costs of the current use of localization techniques (RSL and WGL) in the Dutch population with the costs of a future situation in which MSL is adopted. **Abbreviations:** BC = breast cancer, DCIS = ductal carcinoma in situ, WGL = wire-guided localization, RSL = radioactive seed localization, MSL = magnetic seed localization.

Box 1 - Description of the interventions.

In WGL, a metal wire with a hooked tip is placed in the lesion at the radiology department. This placement needs to be performed on the same day as surgery which complicates scheduling of the surgery. Intraoperatively, the surgeon removes the lesion guided by the wire.[21,22]

With the use of RSL, a small radioactive iodine-125 seed is placed in the lesion by the radiologist. The timing of the placement of the seed in the Netherlands is flexible, and not limited to a few days before surgery. Intraoperatively, a gamma probe providing continuous audible feedback is used by the surgeon to detect the seed. After surgery, the Iodine-125 seed must be removed from the excised specimen in the pathology lab to be safely disposed.[4] We found that using RSL requires an additional availability of staff (e.g. pathology analyst, radiation expert, nuclear medicine staff, RNC assistant) for ~37.5 min per patient compared to WGL. Using RSL also brings the risk of having incidents with radioactive material. From the five hospitals using RSL participating in our study, we identified that each year, on average once or twice a year, the following incidents occur: transection, seed loss, and "near incidents". A "near incident" is for example a situation in which it is thought to have lost an iodine seed which needs formal follow-up in view of radioactivity regulations. For this reason the availability of a radiation expert or staff from the nuclear department is required. Transection and seed loss results in at least 2 days of work and "near incidents" in approximately 6 hours.

With the use of MSL, a magnetic seed is placed in the lesion by the radiologist. The signal will not decay over time, therefore the timing of placement is flexible (in feasibility studies the seed has been in situ for a limited period of 2-30 days prior to surgery[15,16]). Intraoperatively, a magnetic probe providing constant feedback on the location of the seed is used by the surgeon to guide resection of the tumor.[15,16] Although the workflow is similar to RSL, no additional activities are required for intake, pathologic analysis and disposal of the seed. Supplement 1 shows the process of each localization technology.

Model inputs

Patient population

In 2017, 17,207 patients were diagnosed with BC (including DCIS)[23]. The National Institute for Public Health estimates a 15% increase in the incidence of BC over a period of 2015 to 2040[24]. This increase was assumed to be constant. The number of BC patients receiving localization was estimated on registry data from 2014[20]. This rate (31.8%) was assumed to be stable over time as no prospective data was available. A proportion (19%) of the target population received neo-adjuvant chemotherapy (NACT)[25]. Receiving NACT was included in the BIA model because it has an influence on the workflow and thus on the intervention costs. When receiving NACT an additional marker is placed for response monitoring when MSL and WGL is used, because the magnetic seed is not compatible with MRI response measurements and the wire can migrate over time. In RSL no additional marker is needed because in the European setting, the iodine seed can be used for response monitoring as the iodine seed is allowed to be in situ for a long period (>30 days). A constant annual increase of 1.3% in receiving NACT was assumed based on historical trends. The input parameters are listed in Table 1.

Table 1 - Input parameters for the budget impact analysis model

Parameter	Values	Source
Patient population		
Breast cancer incidence in the Netherlands:		
2017	17,207	Dutch registries[23,24]
2018	17,308	
2019	17,409	
2020	17,510	
2021	17,611	
2022	17,712	
Percentage of these BC patients:		
that receives localization (%)	31.8%	[20]
that receives neo-adjuvant chemotherapy (%)	19%	[25]
annual increase of % receiving NACT	1.3%	Based on historic trends[20,25]
# Hospitals that provide BC care in 2017	105	[26]
Future utilization of localization techniques in Dutch patients		
Standard of care scenario (without MSL)		
Assumption: All hospitals that would adopt MSL, adopt RSL before 2022		
Assumption: All hospitals have a similar share in BC patients to calculate the number of hospitals per technique		
Assumption: A new localization technology has a redemption period of 5 years (implementation costs)		
	localization technique used in % patients per year (# of hospitals that use a certain technology)	
	WGL / RSL / MSL	
2017	79% (83) / 21% (22) / 0% (0)	[20]
2018	71% (74) / 29% (31) / 0% (0)	Adoption curve of Rogers
2019	60% (63) / 40% (42) / 0% (0)	Adoption curve of Rogers
2020	52% (55) / 48% (50) / 0% (0)	Adoption curve of Rogers
2021	40% (42) / 60% (63) / 0% (0)	Adoption curve of Rogers
2022	30% (31) / 70% (74) / 0% (0)	Assumption
Standard of care with the introduction of MSL		
Assumption: Adoption follows the adoption curve of Rogers[27]		
Assumption: All hospitals have a similar share in BC patients to calculate the number of hospitals per technology		
Assumption: A new localization technique has a redemption period of 5 years (implementation costs)		
	localization technique used in % patients per year (# of hospitals that use a certain technology)	
	WGL / RSL / MSL	
2017	79% (83) / 21% (22) / 0% (0)	[20]
2018	76% (80) / 24% (25) / 0% (0)	Adoption curve of Rogers
2019	72% (76) / 27% (28) / 1% (1)	Adoption curve of Rogers
2020	65% (68) / 32% (34) / 3% (3)	Adoption curve of Rogers
2021	53% (56) / 36% (38) / 11% (11)	Adoption curve of Rogers
2022	30% (31) / 40% (42) / 30% (32)	Expert opinion NKI-AvL

(continued on next page)

Table 1 (continued) - Input parameters for the budget impact analysis model

Parameter	Values		Source
Intervention costs			
WGL	€2,617		
Personnel costs	€279	(11%)	[28,29]
Material costs	€43	(2%)	Hospital specific purchase costs (interviews)
Intervention costs (surgery and imaging)	€2,173	(83%)	[30] and NKI-AvL
Equipment costs	-	(0%)	
Overhead	€ 123	(5%)	[29]
RSL	€2,834		
Personnel costs	€321	(11%)	[28,29]
Material costs	€118	(4%)	Hospital specific purchase costs (interviews)
Intervention costs (surgery and imaging)	€2,173	(77%)	[30] and NKI-AvL
Equipment costs	€53	(2%)	Hospital specific purchase costs (interviews)
Overhead	€168	(6%)	[29]
MSL	€2,662		Without the costs of the magnetic seed
Personnel costs	€279	(9%)	[28,29]
Material costs	€12	(17%)	Hospital specific purchase costs (interviews)
Intervention costs (surgery and imaging)	€2,173	(73%)	[30] and NKI-AvL
Equipment costs	€49	(2%)	Hospital specific purchase costs (interviews)
Overhead	€149	(5%)	[29]
Costs of the magnetic seed	+ €100 – €500		Assumption
Additional costs for patients receiving neoadjuvant chemotherapy			
Using WGL (material and overhead)	€146		Hospital specific purchase costs (interviews);[29]
Using MSL (material and overhead)	€146		
Average inflation ratio to account for an increase in costs in the future	1.0116		Assumption based on Dutch inflation rates of the past 5 years[31]
Implementation costs			
WGL	N.A.		
RSL (yearly costs)	€26,826 (€5,553)		Based on costs of 2017[28,29] and NKI-AvL
Personnel	€18,629		Process analysis by interviews in 5 hospitals.
Overhead	€8,197		On average 332.75 hours of work
MSL (yearly costs)	€ 2,794 (€578)		Based on costs of 2017[28,29] and NKI-AvL
Personnel	€1,940		Process estimation based on interviews in the NKI-AvL estimated hours of work: 24 for training and writing protocols
Overhead	€854		

Abbreviations: WGL = wire-guided localization, MSL = magnetic seed localization, RSL = radio-active seed localization, BC = Breast Cancer. Supplement 2 and 3 contain specific details on the cost components incorporated in the intervention costs (including actual costs) presented here.

Expected utilization of localization techniques

In 2017, RSL and WGL were used in respectively 21% and 79% of the BC patients, due to hospital differences[20]. To simulate future uptake, we assumed that WGL is not implemented in the coming years but that RSL or MSL will be implemented as a new technology in the future.

The potential future uptake of RSL and MSL in 2022 was estimated by experts working in the Netherlands Cancer Institute (NKI) where MSL is used in a research setting next to RSL[15]. Since, theoretical models describe that having knowledge on the innovation and the degree of relative advantage are important factors in the adoption decision, we did not consider it likely that very fast implementation would occur[27,32]. The uptake of MSL was estimated to be 30% in 2022 and the total usage of RSL and WGL 40% and 30% respectively. For SoC in 2022, the uptake of RSL and WGL was estimated at 70% and 30% respectively, assuming that hospitals wiling to adopt MSL (30%) adopt RSL instead.

The classic diffusion theory by Rogers was used to estimate the adoption speed of RSL and MSL[27,33]. The annual uptake of MSL and RSL was estimated by using the "S"-shaped curve proposed by Rogers. This is shown in Figure 2 and Table 1. According to the projected diffusion curves, the adoption rates for 2023 were estimated to allocate the implementation costs of RSL and MSL in 2022.

Intervention and implementation costs

In the Netherlands, reimbursement for the use of localization techniques during breast cancer surgery is part of a budget allocated for a specific combination of diagnosis and treatment. Therefore, specific costs for using a localization technology apart from e.g. hospital stay, are not specified. By means of Activity Based Costing (ABC) costs for using the localization technologies were estimated. This method takes into account all activities consumed within a process and allocates costs to the resources required for these activities[34].

Clinical process per localization technology

First the processes had to be drafted for using WGL, RSL and MSL. The processes were evaluated by clinical expert interviews in eight hospitals in 2017 (five using RSL, three using WGL), "real-time" observations, literature and hospital treatment protocols. Since MSL was only used in one Dutch hospital, the MSL process was based on interviews held in that institute (NKI-AvL). The expert interviews assessed also the implementation process for RSL (evaluated in five hospitals) and MSL (evaluated in the NKI-AvL only). Supplement 1 shows the workflow per localization technology.

In drafting the processes, the incidents associated with the use of radioactive seeds were

included in the process. Based on literature, the duration of seed placement (45 min) and surgery (90 min) were assumed to be similar between the localization techniques[4,12] and migration of magnetic seeds was assumed to be negligible[15,16]. Furthermore, based on multiple studies comparing WGL and RSL, minor complications e.g. wound infection, and displacement of the wire or seed were neglected as they were assumed to be uncommon and equal for the three localization technologies[10,12,35]. For the implementation processes we evaluated the numbers of staff involved and their number of hours invested in processes as: drafting protocols, performing a risk analysis, training, obtaining a license and internal procedures.

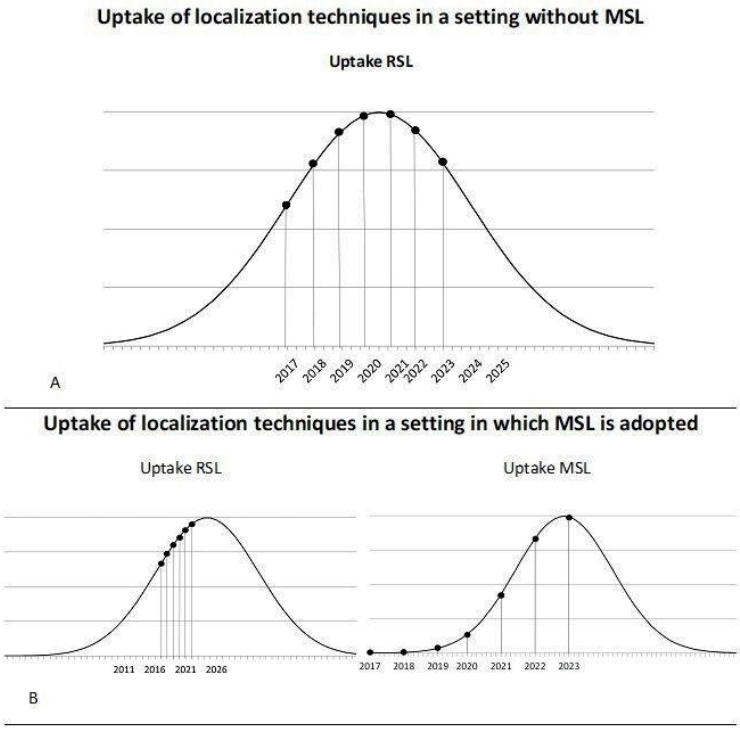


Figure 2. The expected uptake of RSL and MSL for both standard of care (SoC) and SoC with the adoption of MSL. (A) Shows the uptake of RSL when MSL is not implemented based on the adoption curve of Rogers and the assumed uptake of localization techniques in 2022: 70% RSL, 30% WGL. (B) Shows the uptake curves for RSL and MSL based on the adoption curve of Rogers and the assumed uptake of localization techniques in 2022: 30% MSL, 40% RSL. These curves were used to identify the number of patients per year receiving one of the technologies and to identify the hospitals that transfer from one technology to another. **Abbreviations:** WGL = wire-guided localization, RSL = radioactive seed localization, MSL = magnetic seed localization.

Cost calculation

To calculate the costs of each process step, the Dutch manual for cost calculations was used[29]. Personnel costs were calculated by multiplying the reference costs or gross salaries according to the collective agreement for hospitals of 2017 to the amount of time a staff member was occupied per process step[28]. Those costs were also used to calculate the implementation costs by multiplying the costs for the involved staff and the number of hours spent for implementation. The costs for surgery, pathology assessment and seed/wire placement were based on internal hospital prices or regulated tariffs from the Dutch Healthcare Authority[30]. The materials used and costs of materials were based on data from the eight selected hospitals. The material costs of MSL incorporated: the non-magnetic polymer surgical tools (Blunt retractor, sharp Weitlaner, retractor and a small or long forceps[15]), sterile cover for the probe, and magnetic probes. The costs were based on hospital data and expert interviews in the NKI (Table 1 and Supplement 3). For the costs of the polymer surgical tools, the average usage of the different tools was estimated (e.g. 50% for the blunt retractor). These values were multiplied by the prices of the tools (Internal cost information NKI-AVL). The magnetic seeds costs were based on list prices of two companies selling products for MSL and was included as a range between €100 and €500.

Although the equipment used in WGL and RSL were already bought and will be used for several procedures, we included the equipment costs to have a fair comparison to MSL. Equipment costs for RSL, WGL and MSL were based on actual acquisition costs from the participating hospitals. Since the gamma probe and contamination monitors, essential for using RSL, are also used in other procedures these costs were partly taken into account: 50% and 30% respectively. Finally, overhead was calculated over all costs except over the intervention and material costs to avoid double counting, using a general percentage of 44%[29].

The intervention costs included in the BIA model were: €2,617, €2,834, €2,662 (without magnetic seed) for WGL, RSL and MSL respectively. The additional costs per patient receiving NACT in WGL and MSL were €146, and the implementation costs for MSL and RSL were: €2,794 and €26,826 respectively. These costs and details on the analysis are presented in Table 1 and Appendices B and C.

Analysis

To perform the analysis, Microsoft Excel version 2010 was used. The BIA compares the total intervention costs of the localization technologies used per year plus the yearly implementation costs of the hospitals that are expected to transfer to a different technology for both SoC and SoC with MSL. To calculate the yearly treatment costs, the yearly BC incidence was multiplied

by the percentage of patients receiving localization during surgery and by the yearly uptake percentages of the localization technologies. These numbers were multiplied by the costs per localization technology including the additional costs for the proportion of patients receiving NACT. The future costs for 2018 and later were corrected using an average inflation rate based on the Dutch inflation rates of the previous five years[31].

Sensitivity analysis

The model structure and input parameters were based on several assumptions and therefore associated with a level of uncertainty. To evaluate the impact of our assumptions, deterministic sensitivity analyses (DSA) were conducted.

First, a one-way sensitivity analysis was conducted on the results of the cost analysis for RSL and MSL to identify the parameters with the highest influence. Upper and lower limits of 20% were used varying the for instance the number of patients per hospital, equipment costs and duration of placing the marker. Second, a one-way sensitivity analysis was conducted on the BIA results for the year 2022 with fixed magnetic seed costs of €200. Also upper and lower limits of 20% were used to check the influence of several input parameters. For example: implementation costs, treatment costs, and the percentage of patients receiving NACT. Finally, three alternative diffusion estimates were tested: a constant uptake of MSL, changing the adoption speed of MSL (slower, faster), and changing the initial uptake of RSL in 2017 to (1) 30% RSL and 70% WGL, and (2) 40% RSL and 60% WGL.

RESULTS

Budget impact analysis

The results of the BIA model are shown in Table 2 and Figure 3. Total costs for SoC with RSL and WGL use increased from €14.7m in 2017 to €16.9m in 2022 due to an increased number of BC patients and increased number of hospitals implementing RSL. When MSL is increasingly adopted and the magnetic seed would only cost €100, total costs increased from €14.7m to €16.7m resulting in a BI of -€0.2m in 2022. With a magnetic seed price of €500, total healthcare costs increased from €14.7m to €17.6m, resulting in a BI of €0.7m in 2022. At a price level of €178 for the magnetic seed, the BI in 2022 is neutral.

Figure 3 shows that there is a benefit to adopt MSL due to the lower implementation costs. However, when more hospitals are implementing MSL and the intervention costs of using MSL are higher than for RSL and/or the percentage of NACT patients is increasing, the use

of RSL and WGL is more cost-efficient for the Netherlands overall. For each hospital, which localization technology is most cost-efficient depends on the number of BC patients per year, proportion of patients receiving NACT and the current implemented localization technique. Supplement 4 contains an adjustable version of the BIA model to enable evaluation of the adoption of MSL for a different country or a hospital setting.

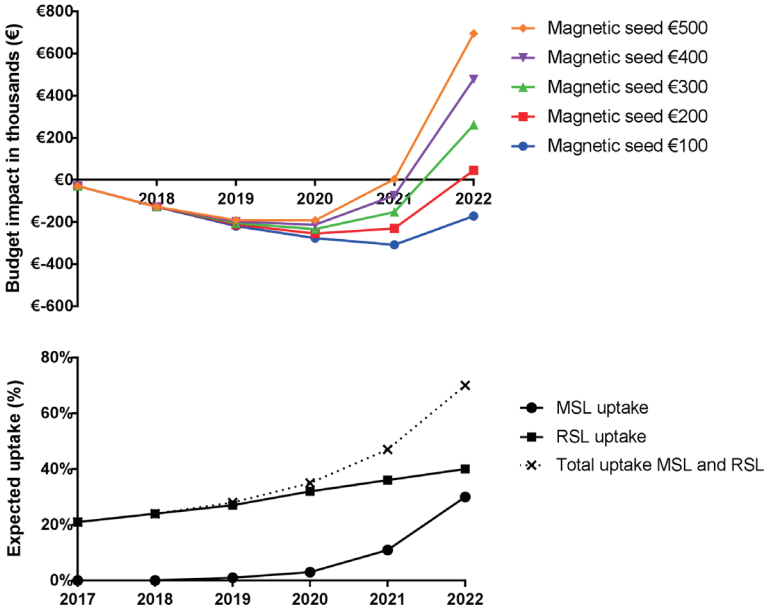


Figure 3. The total annual budget impact in respect of the uptake of RSL and MSL when MSL is adopted. The annual budget impact of a setting in which MSL is adopted compared to standard of care is visualized. In addition the expected uptake of RSL and MSL for the future situation is visualized as this explains the rise in budget impact. When a small percentage of hospitals is transferring to MSL instead of RSL e.g. year 2020 and 2021 and the cost of the magnetic seeds is ≤€200 a benefit is seen due to the smaller yearly implementation costs. This effect is overruled when more hospitals are transferring to MSL as the costs per patient are higher for MSL than for RSL. **Abbreviations:** WGL = wire-guided localization, RSL = radioactive seed localization, MSL = magnetic seed localization.

Table 2 - Base case results of the budget impact analysis of adopting magnetic seed localization (MSL) in breast conserving surgery

	2017	2018	2019	2020	2021	2022
BC patients that are being localized	5,472	5,504	5,536	5,568	5,600	5,633
Patients receiving NACT	821	826	830	835	840	845
Standard of care (without MSL)						
#hospitals using WGL	83	75	63	55	42	31
#hospitals using RSL	22	30	42	50	63	74
#hospitals using MSL	0	0	0	0	0	0
Patients localized with WGL (NACT)	4,323 (821)	3,908 (793)	3,322 (717)	2,895 (663)	2,240 (542)	1,690 (431)
Patients localized with RSL (NACT)	1,149 (218)	1,596 (324)	2,214 (478)	2,673 (612)	3,360 (813)	3,943 (1005)
Patients localized with MSL (NACT)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Healthcare costs (€)	14,689,668	15,031,576	15,412,565	15,769,062	16,177,124	16,571,516
Implementation costs (€)	46,647	111,532	159,268	231,702	292,765	309,402
Total (€)	14,736,316	5,143,108	15,571,833	16,000,765	16,469,889	16,880,918
Standard of care with adoption of MSL						
#hospitals using WGL	83	80	76	68	56	31
#hospitals using RSL	22	25	28	34	38	42
#hospitals using MSL	0	0	1	3	12	32
Patients localized with WGL (NACT)	4,323 (821)	4,183 (849)	3,986 (861)	3,619 (829)	2,968 (718)	1,690 (431)
Patients localized with RSL (NACT)	1,149 (218)	1,321 (268)	1,495 (323)	1,782 (408)	2,016 (488)	2,253 (575)
Patients localized with MSL (NACT)	0 (0)	0 (0)	55 (12)	167 (38)	616 (149)	1,690 (431)
Healthcare costs (€) when the magnetic seed costs:						
€ 100	14,689,668	14,987,005	15,298,487	15,645,802	16,046,702	16,564,325
€ 200	14,689,668	14,987,005	15,305,342	15,666,726	16,124,763	16,780,924
€ 300	14,689,668	14,987,005	15,312,197	15,687,651	16,202,824	16,997,522
€ 400	14,689,668	14,987,005	15,319,052	15,708,575	16,280,885	17,214,121
€ 500	14,689,668	14,987,005	15,325,907	15,729,499	16,358,945	17,430,720
Implementation costs (€)	17,493	35,803	66,881	96,055	132,563	162,039
Budget impact (€)	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
when the magnetic seed costs:						
€ 100	-29,155	-127,828	-219,411	-276,032	-308,493	-171,598
€ 200	-29,155	-127,828	-212,556	-255,107	-230,432	45,001
€ 300	-29,155	-127,828	-205,701	-234,183	-152,371	261,600
€ 400	-29,155	-127,828	-198,846	-213,259	-74,310	478,199
€ 500	-29,155	-127,828	-191,991	-192,334	3,750	694,798
Budget impact in healthcare costs only (€) when the magnetic seed costs:						
€ 100	0	-52,099	-127,024	-140,385	-148,292	-24,235
€ 200	0	-52,099	-120,169	-119,460	-70,231	192,364
€ 300	0	-52,099	-113,314	-98,536	7,830	408,962
€ 400	0	-52,099	-106,459	-77,611	85,891	625,561
€ 500	0	-52,099	-99,604	-56,687	163,952	842,160

Abbreviations: WGL = Wire-guided Localization, MSL = magnetic seed localization, RSL = Radio-active Seed Localization, BC = Breast Cancer, NACT = Neoadjuvant chemotherapy. All values are rounded.

Sensitivity analysis

The duration of the excision and seed placement, the costs of the magnetic seed and the overhead percentage drove the intervention costs of RSL and MSL the most (Figure 4a and 4b). Uncertainty in those parameters could have a great impact on the calculated costs per patient and thus on the results of the BIA. As Figure 4c demonstrates, the intervention costs had a substantial influence on the BIA results.

The different diffusion estimates incorporated in the DSA had a small impact on the budget impact (Figure 4c). A constant uptake of MSL showed an increased BI because the uptake of RSL in 2023 is then much higher than in the base case situation which results in higher implementation costs accounted in 2022 for the situation with MSL adoption. A steeper adoption curve of MSL showed an increased BI because the intervention costs of MSL are higher than those for RSL. These higher costs were not resolved by the lower implementation costs for MSL. The increased uptake of RSL in 2017 showed an increased BI, because the endpoint in 2022 for RSL was kept the same, and therefore less hospitals transferred to RSL per year in both scenarios resulting in lower total implementation costs especially for usual care.

DISCUSSION

The results of the BIA indicated that adoption of MSL in the Dutch healthcare system could be cost-saving due to the lower implementation costs for MSL (€2,974) compared to RSL (€26,826). However, to maintain this advantage after implementation phase: (1) the costs of using MSL per patient should not be substantially higher than those for RSL or (2) response monitoring with MRI should be enabled when using MSL in NACT patients or (3) the use of MSL should result in improved clinical outcomes compared to WGL and RSL.

To the best of our knowledge this is the first BIA on localization techniques in which the intervention costs of localization techniques have been evaluated in detail, including the additional activities related to using radioactive material. The results of the cost analysis could inform the decision to transfer from WGL to either MSL or RSL on a national or hospital level. The decision to adopt one of the technologies may be further supported by additional factors such as the improved resource allocation and impact of a localization method on logistics[3,4,6]. Besides, the MSL use could be used relatively easy outside the breast cancer indication, whereas the expansion of indications for RSL involves a time-consuming regulative route because of its radioactive nature[15,16]. These additional factors are important to take into account when deciding on the adoption of MSL but a detailed evaluation was out of the

scope of this study.

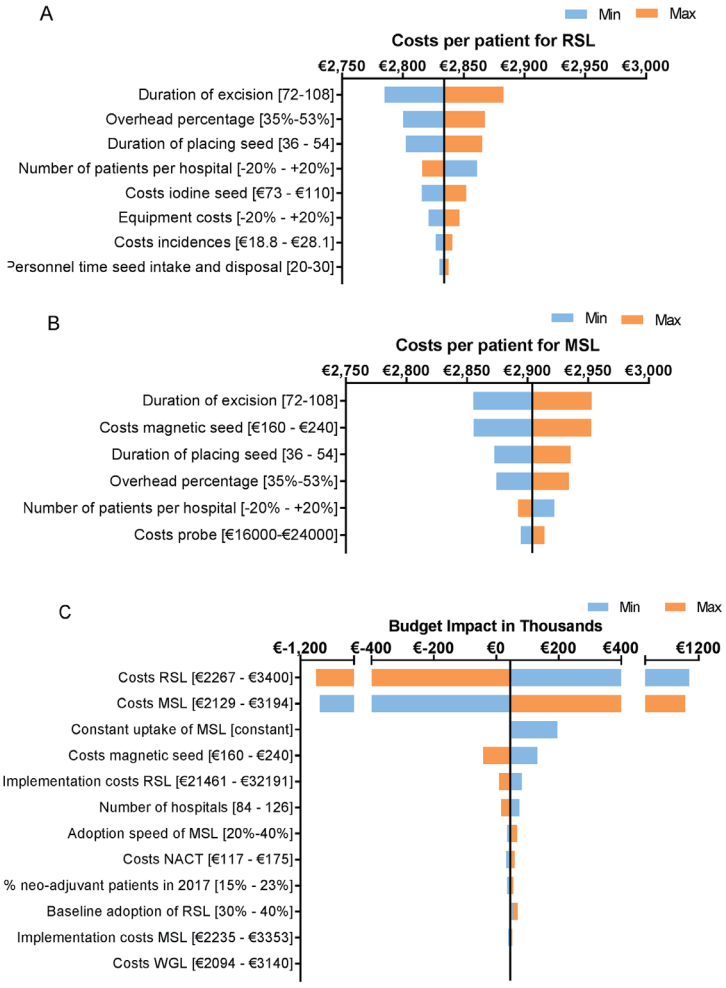


Figure 4. Results of the one-way sensitivity analyses. A. shows the results of the sensitivity analysis on the cost analysis results for RSL. The base case value is: €2,833.95. B. shows the results of the sensitivity analysis on the cost analysis results for MSL. The base case value is: €2,904.06 with a magnetic seed price of €200. C. shows the results of the sensitivity analysis on the results of the budget impact analysis in 2022, with a magnetic seed price of €200. The base case value is €45.00. Explanation regarding parameter "constant uptake of MSL": The yearly uptake of WGL was held constant and the uptake of RSL was linked to the uptake of MSL (RSL yearly uptake = 100%- %MSL- %WGL). **Abbreviations:** WGL = wire-guided localization, RSL = radioactive seed localization, MSL = magnetic seed localization.

The BIA results are mainly driven by the intervention costs (Figure 4). The costs used in our analysis are based on several Dutch hospitals. Since, country specific regulations related to safety of radioactive material can have an influence on the costs, the generalizability of the results from the cost-analysis to non-European countries is limited. Comparing our results to current evidence showed that the comparability of publications is limited because of variations in design choices, limited access to cost data, and differences in costs and materials used across countries[36–38]. Comparable studies included for instance re-excision rates, complication rates and cost savings related to logistics which resulted in overall savings for RSL compared to WGL[38–40]. Also, the presented costs are often relative differences instead of absolute numbers[38,39]. A study from a US perspective showed that RSL was also associated with higher material and personnel costs compared to WGL, but further validation of our results was limited as the results were presented as savings per patient ($115)[38]. In general for using RSL in the US, our results are expected to underestimate the costs and budget impact as the procedures related to radioactive material use are different. Especially, regarding the allowed duration of iodine seeds to remain in situ (max of 5-7 days) implying that in NACT treated patients an additional marker has to be placed for response monitoring[38,41]. As in our analysis the additional marker placement is the main disadvantage of MSL compared to RSL, this would have a significant impact on our results and conclusion (budget impact of €-21,900 in 2022 in favor of MSL (magnetic seed=€200).

The main strength of the current analysis is the detailed insight in the costs of all three localization techniques, based on data from 8 Dutch hospitals. Our results can be used and adjusted on a hospital and country level to guide the decision to adopt RSL or MSL using the general model (Supplement 4). As we were not able to include all available techniques in this field due to lack of detailed data, the model allows to include other promising alternatives to MSL such as radioguided occult lesion localization[39,42], radar technology (SAVI SCOUT)[43,44] or ultrasound[45,46] applications. We have not been able to compare the results from our analysis to alternatives such as SAVI SCOUT, that recently received $510k approval from the FDA. As a trial is still to be reported upon (NCT03015649), we advise to perform a comparable analysis once the technology proves to have equal or better clinical value compared to existing technologies. Another strength of our analysis is the inclusion of the implementation costs to clarify the relation between the acceptable higher treatment costs but a less labor intensive implementation process compared to RSL.

The main limitations in this study are the assumptions regarding uptake of various techniques for 2022, and the early stage of our analysis. As the present analysis evaluates a new technology still in development and subject of clinical trials, the results from the cost-analysis related to MSL (treatment and implementation costs) are uncertain and of potentially limited applicability. The impact of the implementation costs on the budget for 2017 to 2022 could

be underestimated, due to allocation of the implementation costs over 5 years and the majority of hospitals were assumed to adopt a new technology in the final two years. Other limitations were: (1) the selection of the hospitals, as this could have biased the cost-analysis results. Although we incorporated all types of hospitals (academic, general and specialized) and hospitals located in different areas of the country, this could limit the generalizability of the budget impact analysis. (2) Not being able to incorporate the logistical hurdles when using WGL and therefore the intervention costs of WGL were underestimated. This however would not have altered the conclusions as the benefits are similar for MSL and RSL compared to WGL. A final limitation (3) is the main assumption that the efficacy of MSL and RSL is similar to WGL. Future comparative studies should verify whether this is truly the case. If clinical benefit is expected these factors should be incorporated in this analysis or a cost-effectiveness analysis should be performed.

Conclusion

Our present analysis shows that MSL could be a new cost-efficient localization technology in guiding resections of non-palpable breast cancer tumors in the Netherlands. When the costs to use MSL are significantly higher than those for using RSL and WGL, the lower implementation costs for MSL will not outbalance these higher intervention costs. Manufactures should consider cost-effectiveness when determining retail price of the magnetic seed.

ACKNOWLEDGEMENTS

The authors want to thank all the experts of the eight Dutch hospitals willing to participate in the interviews concerning treatment process and obtaining the costs of using RSL, MSL and WGL.

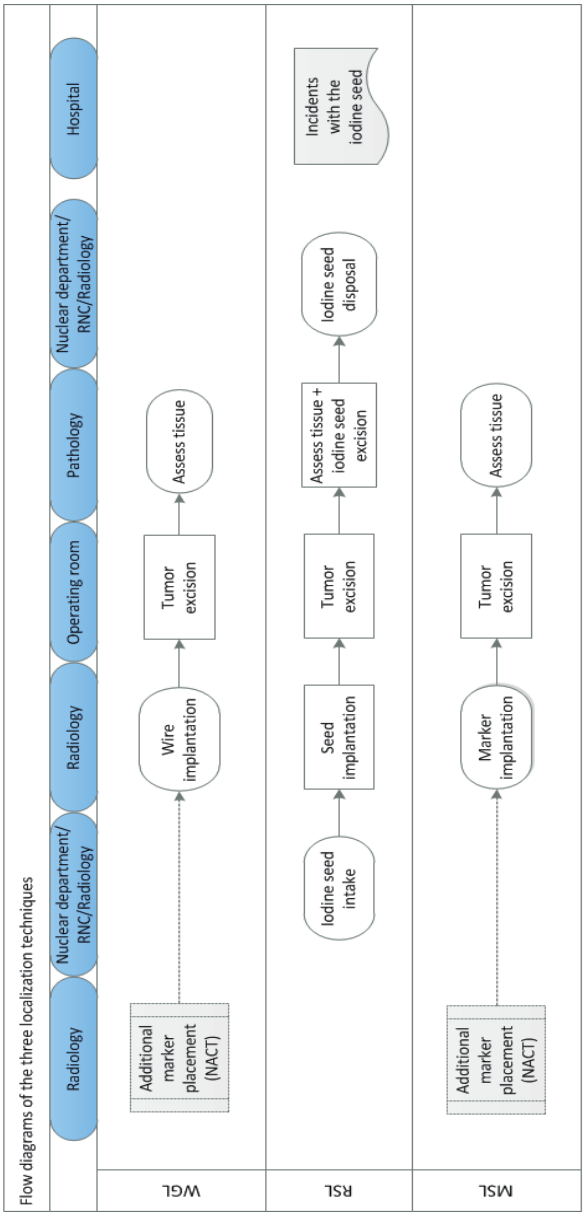
REFERENCES

1. Veronesi, U. *et al.* Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N. Engl. J. Med.* **347**, 1227–1232 (2002).
2. Noblin, A. M. Privacy policy analysis for health information networks and regional health information organizations. *Health Care Manag. (Frederick).* **26**, 331–340 (2007).
3. Jakub, J. W. *et al.* Current status of radioactive seed for localization of non palpable breast lesions. *Am. J. Surg.* **199**, 522–528 (2010).
4. Gray, R. J. *et al.* Randomized prospective evaluation of a novel technique for biopsy or lumpectomy of nonpalpable breast lesions: radioactive seed versus wire localization. *Ann. Surg. Oncol.* **8**, 711–715 (2001).
5. Ahmed, M. & Douek, M. Radioactive seed localisation (RSL) in the treatment of non-palpable breast cancers: Systematic review and meta-analysis. *Breast* **22**, 383–388 (2013).
6. Sharek, D. *et al.* Radioactive seed localization versus wire localization for lumpectomies: A comparison of outcomes. *Am. J. Roentgenol.* **204**, 872–877 (2015).
7. Gunn, J. & McLaughlin, S. Current Trends in Localization Techniques for Non-palpable Breast Lesions: Making the Invisible Visible. *Curr. Breast Cancer Rep.* **9**, 165–171 (2017).
8. Chan, B. K. Y., Wiseberg-Firtell, J. A., Jois, R. H. S., Jensen, K. & Audisio, R. A. Localization techniques for guided surgical excision of non-palpable breast lesions. *Cochrane database Syst. Rev.* **2015**, CD009206 (2015).
9. Parvez, E. *et al.* A cosmesis outcome substudy in a prospective, randomized trial comparing radioguided seed localization with standard wire localization for nonpalpable, invasive, and in situ breast carcinomas. *Am. J. Surg.* **208**, 711–718 (2014).
10. Hughes, J. H. *et al.* A multi-site validation trial of radioactive seed localization as an alternative to wire localization. *Breast J.* **14**, 153–157 (2008).
11. Bloomquist, E. V. *et al.* A randomized prospective comparison of patient-assessed satisfaction and clinical outcomes with radioactive seed localization versus wire localization. *Breast J.* **22**, 151–157 (2016).
12. Lovrics, P. J. *et al.* A multicentered, randomized, controlled trial comparing radioguided seed localization to standard wire localization for nonpalpable, invasive and in situ breast carcinomas. *Ann. Surg. Oncol.* **18**, 3407–3414 (2011).
13. Nuclear Regulatory Commission. Iodine-125 and Palladium-103 Low Dose Rate Brachytherapy Seeds Used for Localization of Non-Palpable Lesions. Available at: https://www.nrc.gov/materials/miau/med-use-toolkit/seed-localization.html. (Accessed: 7th August 2018)
14. THE COUNCIL OF THE EUROPEAN UNION. Council directive 2013/59/Eurotom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom a. (2014).
15. Schermers, B. et al. Feasibility of magnetic marker localisation for non-palpable breast cancer. Breast 33, 50–56 (2017).
16. Harvey, J. R. *et al.* Safety and feasibility of breast lesion localization using magnetic seeds (Magseed): a multi-centre, open-label cohort study. *Breast Cancer Res. Treat.* **169**, 531–536 (2018).
17. Zacharioudakis, K. *et al.* Is the future magnetic? Magseed localisation for non palpable breast cancer. A multi-centre non randomised control study. *Eur. J. Surg. Oncol.* **45**, 2016–2021 (2019).
18. Angarita, F. A. *et al.* Comparison of Radioactive Seed Localized Excision and Wire Localized Excision of Breast Lesions: A Community Hospital's Experience. *Clin. Breast Cancer* **19**, e364–e369 (2019).
19. Sullivan, S. D. *et al.* Budget impact analysis- Principles of good practice: Report of the ISPOR 2012 budget impact analysis good practice II task force. *Value Heal.* **17**, 5–14 (2014).
20. DICA. *Jaarrapportage NBCA 2014 (additional tables; available on request)*. (2014).
21. Dodd, G. D., Fry, K. & Delany, W. Pre-operative localization of occult carcinoma of the breast. *T.F. Nealon Manag. patient with cancer* 88–113 (1965).
22. Dua, S. M., Gray, R. J. & Keshtgar, M. Strategies for localisation of impalpable breast lesions. *Breast* **20**, 246–253 (2011).
23. Integraal Kankercentrum Nederland. *Cijfers over kanker: indicentie borst*.

24. Rijksinstituut voor Volksgezondheid en Milieu. Toekomstige stijging aantal nieuwe gevallen van borstkanker door alleen demografie. (2018). Available at: https://www.volksgezondheidenzorg.info/onderwerp/borstkanker/cijfers-context/trends#node-toekomstige-trend-borstkanker-door-demografische-ontwikkelingen. (Accessed: 8th August 2018)
25. NBCA. *NABON BREAST CANCER AUDIT (NBCA) YEAR REPORT 2017*. (2017).
26. Borstkankervereniging Nederland. B monitor; Vergelijk ziekenhuizen. Available at: https://monitorborstkankerzorg.nl/nl/monitor-borstkankerzorg/ziekenhuisnaam?status=1&ziekenhuisnaam=. (Accessed: 10th August 2018)
27. Rogers, E. M. The Diffusion of Innovations. *Diffus. Innov.* 576 (2003).
28. Nederlandse Vereniging van Ziekenhuizen (NVZ). Salaristabbelen 1 januari 2017. (2017). Available at: https://cao-ziekenhuizen.nl/salarisschalen-cao-ams. (Accessed: 10th August 2018)
29. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. *Manual for cost research: methods and standard cost prices for economic evaluations in health care*. (2015).
30. Dutch Healthcare Authority (NZa). DBC product finder for tariffs. (2019). Available at: http://dbc-zorgproducten-tarieven.nza.nl. (Accessed: 2nd February 2017)
31. StatLine. Jaarmutatie consumentenprijsindex; vanaf 1963. (2020). Available at: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?ts=1534249924434. (Accessed: 3rd February 2020)
32. Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **13**, 319 (1989).
33. Grimm, S. E., Stevens, J. W. & Dixon, S. Estimating Future Health Technology Diffusion Using Expert Beliefs Calibrated to an Established Diffusion Model. *Value Heal.* **21**, 944–950 (2018).
34. Lievens, Y., Van Den Bogaert, W. & Kesteloot, K. Activity-based costing: A practical model for cost calculation in radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **57**, 522–535 (2003).
35. Gray, R. J., Pockaj, B. A., Karstaedt, P. J. & Roarke, M. C. Radioactive seed localization of nonpalpable breast lesions is better than wire localization. *Am. J. Surg.* **188**, 377–380 (2004).
36. Rao, R. *et al.* Experience with seed localization for nonpalpable breast lesions in a public health care system. *Ann. Surg. Oncol.* **17**, 3241–3246 (2010).
37. Diego, E. J. *et al.* Localizing High-Risk Lesions for Excisional Breast Biopsy: A Comparison Between Radioactive Seed Localization and Wire Localization. *Ann. Surg. Oncol.* **21**, 3268–3272 (2014).
38. Loving, V. A. *et al.* Monte Carlo Simulation to Analyze the Cost-Benefit of Radioactive Seed Localization Versus Wire Localization for Breast-Conserving Surgery in Fee-for-Service Health Care Systems Compared With Accountable Care Organizations. *Am. J. Roentgenol.* **202**, 1383–1388 (2014).
39. Postma, E. L. *et al.* Cost-effectiveness of radioguided occult lesion localization (ROLL) versus wire-guided localization (WGL) in breast conserving surgery for nonpalpable breast cancer: Results from a randomized controlled multicenter trial. *Ann. Surg. Oncol.* **20**, 2219–2226 (2013).
40. Zhang, Y. *et al.* Radioactive Seed Localization Versus Wire-Guided Localization for Nonpalpable Breast Cancer: A Cost and Operating Room Efficiency Analysis. *Ann. Surg. Oncol.* **24**, 3567–3573 (2017).
41. Goudreau, S. H., Joseph, J. P. & Seiler, S. J. Preoperative Radioactive Seed Localization for Nonpalpable Breast Lesions: Technique, Pitfalls, and Solutions. *Radiographics* **35**, 1319–1334 (2015).
42. Hawkins, S. C. *et al.* Time to go wireless? A 15-year single institution experience of radioisotope occult lesion localisation (ROLL) for impalpable breast lesions. *Eur. J. Surg. Oncol.* **43**, 62–67 (2017).
43. Cox, C. E. *et al.* A Prospective, Single Arm, Multi-site, Clinical Evaluation of a Nonradioactive Surgical Guidance Technology for the Location of Nonpalpable Breast Lesions during Excision. *Ann. Surg. Oncol.* **23**, 3168–3174 (2016).
44. Falcon, S., Weinfurtner, R. J., Mooney, B. & Niell, B. L. SAVI SCOUT® localization of breast lesions as a practical alternative to wires: Outcomes and suggestions for trouble-shooting. *Clin. Imaging* **52**, 280–286 (2018).
45. Merrill, A. Y. *et al.* Cutting Healthcare Costs with Hematoma-Directed Ultrasound-Guided Breast Lumpectomy. *Ann. Surg. Oncol.* **25**, 3076–3081 (2018).
46. Arentz, C. *et al.* Ten-Year Experience with Hematoma-Directed Ultrasound-Guided (HUG) Breast Lumpectomy. *Ann. Surg. Oncol.* **17**, 378–383 (2010).

SUPPLEMENTARY MATERIAL

Supplement A - Flow diagrams of included steps in the ABC analyses of WGL, RSL and the MSL



Supplement B - Detailed description of intervention costs

Step	Cost category	WGL	RSL	MSL	Source of cost
Neo-adjuvant	Material	€102		€102	Interviews
	Overhead	€45		€45	[29]
Seed intake	Personnel		€8		[28,29]
	Overhead		€3		[29]
Implantation of seed, wire, marker	Personnel	€109	€109	€109	[28,29]
	Material	€43	€111	€0	Interviews
	Intervention	€179	€179	€179	[30]
	Equipment		€1		Interviews
	Overhead	€48	€48	€48	[29]
Tumor excision guided by seed or wire	Personnel	€170	€170	€170	[29]
	Material		€7	€12	Procurement NKI
	Intervention	€1,329	€1,329	€1,329	NKI
	Equipment		€29	€49	Interviews
	Overhead	€75	€90	€101	[29]
Assessment of tumor and seed excision	Personnel		€6		[28,29]
	Intervention	€666	€666	€666	[30]
	Equipment		€22		Interviews
	Overhead		€12		[29]
Seed disposal	Personnel		€5		[28,29]
	Overhead		€2		[29]
Incidents	Personnel		€23		[28,29]
	Equipment		€2		[29]
	Overhead		€ 11		[29]
Implementation: training	Personnel	N/A	€4,321	€1,189	[28,29]
Implementation: Protocol drafting	Personnel	N/A	€4,290	€751	[28,29]
Implementation: Obtaining license	Personnel	N/A	€2,224	N/A	[28,29]
Implementation: Risk analysis	Personnel	N/A	€2,972	N/A	[28,29]
Implementation: Internal procedures	Personnel	N/A	€6,430	N/A	[28,29]
Total implementation costs*	Personnel	N/A	€18,629	€1,940	[28,29]
Overhead costs	Overhead		€8,197	€854	[29]
Summary					
Average cost per patient		€2,617	€2,834	€2,662	
Variable extra cost per neo-adjuvant patient		€147	€0	€147	
Fixed implementation costs		N/A	€26,826	€2,794	

Material and equipment costs are including VAT. Implementation costs are shown as fixed costs. Other costs are shown as cost per patient. The average cost per patient does not include implementation costs and the additional costs for the neo-adjuvant setting. **Abbreviations:** MSL = magnetic seed localization, NKI = Netherlands cancer institute, NZA = Dutch healthcare authority, RSL = radioactive seed localization, WGL = wire-guided localization, N/A = not applicable. * the total costs are not corresponding with the sum of the averages per process step as the total average costs were calculated over the total implementation costs per hospital.

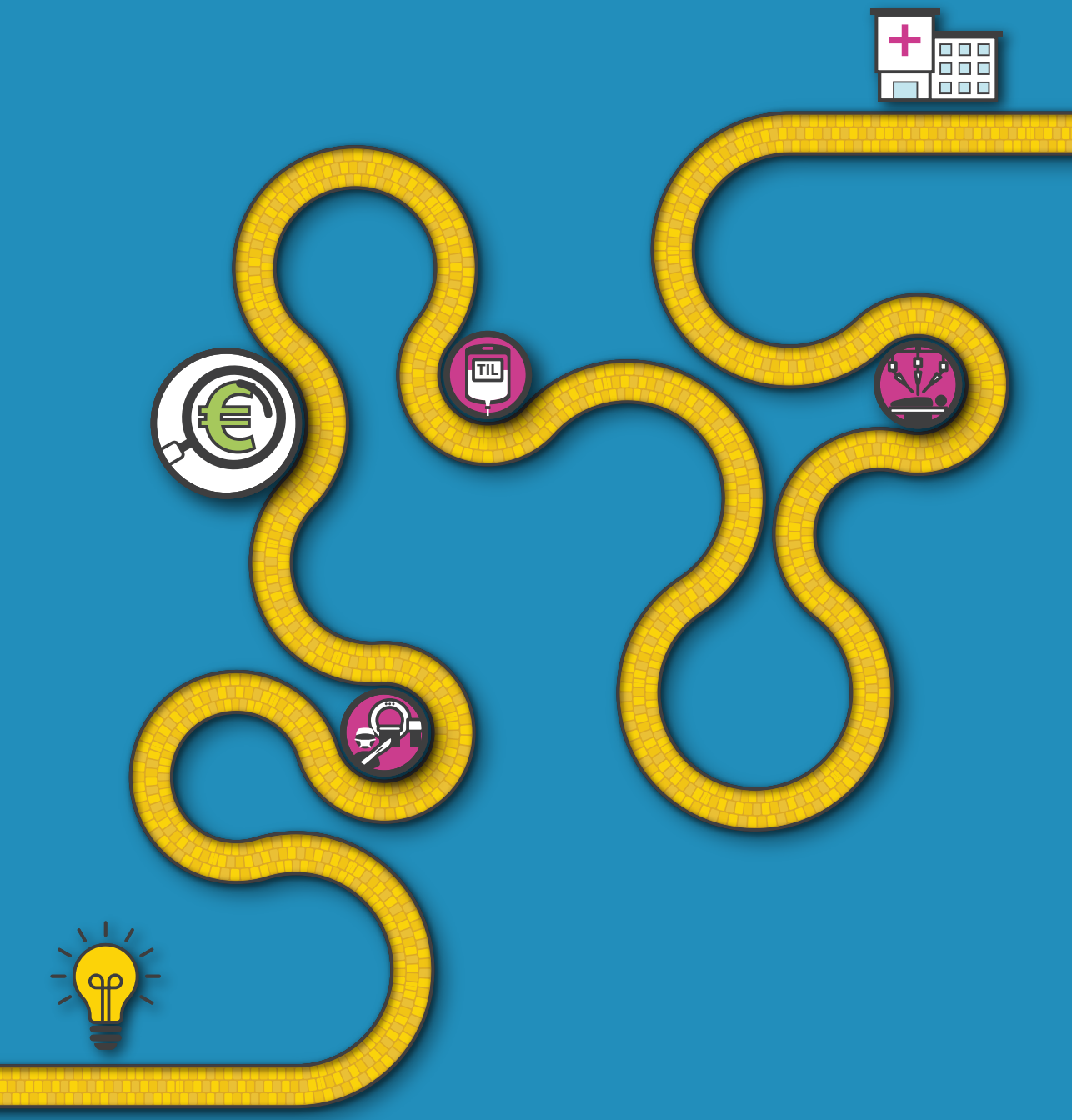
Supplement C - Overview of the included materials and activities

Step	Cost category	WGL	RSL	MSL
Neo-adjuvant	Material	Marker or clip	-	Marker or clip
Seed intake	Personnel	-	RNC assistant/radiation expert/nuclear medicine staff for 15 minutes	-
Seed/wire/ marker implantation	Personnel	Radio diagnostic laboratory worker and a radiologist for 45 minutes	Radio diagnostic laboratory worker and radiologist for 45 minutes	Radio diagnostic laboratory worker and radiologist for 45 minutes
	Material	Wire incl. needle Anesthesia	I-125 seed Needle Materials for preparing seed Anesthesia	Pre-loaded applicator with magnetic marker Anesthesia
	Intervention	Ultrasound* Mammography	Ultrasound* Mammography	Ultrasound* Mammography
	Equipment		Radiation detector	
Tumor excision guided by seed/ wire/marker	Personnel	Surgeon (assumed duration of surgery of 90 minutes)	Surgeon (assumed duration of surgery of 90 minutes)	Surgeon (assumed duration of surgery of 90 minutes)
	Material		Probe cover	Probe cover Polymer surgical instruments
	Intervention	Operation	Operation	Operation
	Equipment		Gamma probe	Magnetic probe
Assessment of tumor and seed excision	Personnel		Analyst for 12.5 minutes on average	
	Intervention	Complex resection	Complex resection	Complex resection
	Equipment		Radiation detector	
Seed disposal	Personnel	-	RNC assistant/radiation expert/nuclear medicine staff for 10 minutes	-

Step	Cost category	WGL	RSL	MSL
Incidents	Personnel incident		Radiation expert/ clinical physicist/ nuclear department for 16 hours	
	Personnel monitoring	-	Several people from the involved departments: radiology, surgery, pathology and nuclear department in total for an average of 42.5 hours per year	-
	Equipment		Contamination monitor	
Implementation	Personnel	-	Several people from the involved departments: radiology, surgery, pathology and nuclear department On average this took in total 322.75 hours	Several people from the involved departments: radiology and surgery Estimated to take 24 hours

*All seeds/wires/markers were assumed to be implanted by ultrasound guidance, although a small percentage of implantations is guided through stereotactic guidance. **Abbreviations:** MSL = magnetic marker localization, RSL = radioactive seed localization, WGL = wire-guided localization.

Supplement D - Adjustable model to calculate the budget impact of adopting MSL on a national or hospital level


TIL

5

Understanding the costs of surgery: A bottom-up cost analysis of both a hybrid operating room and conventional operating room

Sejal Patel*
Melanie Lindenberg*
Maroeska Rovers
Wim van Harten
Theo Ruers
Lieke Poot
Valesca Retèl
Janneke Grutters

*authors contributed equally

International Journal of Health Policy and Management (2020), Jul 27

ABSTRACT

Background

Over the past decade, many hospitals have adopted hybrid operating rooms (OR). As resources are limited, these ORs have to prove themselves in adding value. Current estimations on standard OR costs show great variety, while cost analyses of hybrid ORs are lacking. Therefore, this study aims to identify the cost drivers of a conventional and hybrid OR and take a first step in evaluating the added value of the hybrid OR.

Methods

A comprehensive bottom-up cost analysis was conducted in five Dutch hospitals taking into account: construction, inventory, personnel and overhead costs by means of interviews and hospital specific data. The costs per minute for both ORs were calculated using the utilization rates of the ORs. Cost drivers were identified by sensitivity analyses.

Results

The costs per minute for the conventional OR and the hybrid OR were €9.45 (€8.60- €10.23) and €19.88 (€16.10- €23.07), respectively. Total personnel and total inventory costs had most impact on the conventional OR costs. For the hybrid OR the costs were mostly driven by utilization rate, total inventory and construction costs. The results were incorporated in an open access calculation model to enable adjustment of the input parameters to a specific hospital or country setting.

Conclusions

This study estimated a cost of €9.45 (€8.60 - €10.23) and €19.88 (€16.10 - €23.07) for the conventional and hybrid OR, respectively. The main factors influencing the OR costs are: total inventory costs, total construction costs, utilization rate, and total personnel costs. Our analysis can be used as a basis for future research focusing on evaluating value for money of this promising innovative OR. Furthermore, our results can inform surgeons, and decision and policy makers in hospitals on the adoption and optimal utilization of new (hybrid) ORs.

INTRODUCTION

Over the past decade, many academic and teaching hospitals have adopted a hybrid operating room (OR), and many others are considering it. The compound annual growth rate for the coming 5 years (2019- 2023) of the hybrid OR market growth was estimated at 12.5%[1].

The hybrid OR claims to improve efficiency by means of reducing secondary procedures and improve surgical performance which results in improved clinical outcomes[2–5]. Hybrid ORs are currently mainly used for cardiovascular surgery[2–4], but for neurosurgery[6] and surgical oncology[7] interest is increasing. The first observational studies in cardiovascular surgery verified this promising nature by showing a reduced length of stay and reduced operation time[8,9]. Adoption of such an OR however is a large investment. Since surgical healthcare expenditures already account for a large part of the annual healthcare costs, these innovative ORs have to prove themselves in terms of value for money[10–12].

In evaluating the added value or cost-effectiveness of the hybrid OR, it is crucial to have insight in its incurred costs. So far, the costs related to the hybrid OR have not been studied. Furthermore, to put such costs into perspective, it is important to also gain insight into the incurred costs related to the conventional OR setting. Current estimations on the costs of a conventional OR report a great variety, ranging from $7 to over $100 per minute[13,14]. The variety can be explained by differences in study design such as the inclusion of different cost categories e.g. expensive implants, medical devices, robotics, and site differences as being a (non)teaching hospital, the number of available ORs, the occupancy rate and healthcare system (country-specific). These site characteristics can especially have an influence as OR costs are mainly evaluated top-down. As this method is known to provide little insight into cost drivers and often results in inaccurate cost estimates[15], a bottom-up cost analysis is proposed for further research. This method is used to provide more insight into the cost drivers and enable optimization of processes which could result in cost reductions[15–19].

The aim of this study was to perform a comprehensive bottom-up cost-analysis to inform clinicians and policy makers on the costs and cost drivers of an (hybrid) OR as a first step in evaluating the added value of the hybrid OR. The results are incorporated in a calculation model to enable usage and adjustment of the input parameters to a specific hospital or country setting (Supplement 1).

METHODS

The bottom-up cost-analysis was conducted following the Dutch guideline for costing analyses[20]. This guideline contains reference prices and formulas to estimate costs related to personnel, equipment, construction, and overhead. The analysis was conducted in five Dutch hospitals, all equipped with both a conventional and hybrid OR. Specific characteristics of the included hospitals are described in Supplement 2. Per hospital, we evaluated the following cost categories: construction costs (2.1), inventory costs including medical devices (2.2), personnel costs (2.3), and overhead (2.4) for a conventional and a hybrid OR. The construction, inventory, and personnel costs were evaluated by interviews and hospital-specific data such as invoices and utilization data. By using hospital-specific invoice data we could perform a bottom-up analysis, instead of using total annual expenses which is often done in a top-down analysis. Supplement 2 also reports the positions of the involved experts in these interviews.

In this study, the conventional OR was defined as an operating room where open procedures are performed to evaluate the costs of a basic OR environment. Endoscopic specific devices, to perform laparoscopic and robotic procedures, were thus left out of the analysis. The hybrid operating room was defined as an operating room in which an imaging technique- at least a fixed C-arm- is installed.

Construction costs

In constructing an (hybrid) OR, each hospital makes specific choices for its design because of, among other things, architectural preferences or limitations, preferences for ventilation systems, country-specific legislation, and budget constraints. To avoid such hospital-specific differences, we estimated total costs for constructing a square meter (m^2) of an OR based on Dutch key numbers presented by the Dutch advisory board on healthcare housing[21]. Following these key numbers, constructing a standard m^2 in a hospital costs €3,479 in 2018. For the OR department, these costs should be differentiated with 160%, resulting in a cost of €5,595 per m^2 ($C_{m2\ costs\ OR\ department}$).

This differentiated cost results in the costs of a m^2 which does not take into account the different functionalities available within the OR department (corridors, stockrooms, offices, holding and recovery department and the ORs). To estimate the specific costs for a m^2 of OR per hospital, we identified the total m^2 of each of these specific functionalities within the OR department based on floor plans (e.g. total m^2 of offices). The costs of these m^2 were calculated using their specific differentiation based on key numbers and expert opinions, such as 140% for holding and recovery, 75% for corridors and offices as described in Table 1.

The general m^2 price of an OR per hospital was calculated by:

$$1 \qquad \sum C_{M2\ not\ labeled\ as\ OR} = \sum M^2_{functionality\ i} * C_{differentiated\ i}$$

$$2 \qquad \frac{(\sum M^2_{OR\ department} * C_{M2\ costs\ OR\ department}) - \sum C_{M2\ not\ labeled\ as\ OR}}{\sum M^2_{OR\ within\ OR\ department}}$$

The "C" in this formula refers to costs and the "M^2" to the square meters. An example of the calculation can be found in Supplement 3.

To estimate the construction costs of both ORs, the average m^2 costs of an OR were multiplied with the mean surface of a conventional OR and a hybrid OR including the control room, based on data from the participating hospitals. Yearly costs of interest and amortization were calculated by using a life span of 25 years and an interest rate of 4.2%. A 5% maintenance cost over the construction costs was included[20].

Inventory costs including medical devices

For each hospital the inventory for the conventional and hybrid OR was identified, comprising all equipment standing and hanging in the OR such as operating table, operating lights, (computer) screens, chairs, instrument tables, step stools, and closets. To evaluate the actual (negotiated) costs, the equipment and inventory were linked to the actual acquisition costs paid by each hospital based on their recent invoices (including VAT and discounts). These costs were categorized as follows: general inventory, anesthesia equipment, OR lights, arm pendants, OR table, image routing system, X-ray radiation protection aprons, and fixed C-arm.

All costs made before 2018 were converted to 2018 Euro by using the consumer price index value for the Netherlands[22]. Per cost category, average costs were calculated to determine the average inventory cost for each OR. The yearly costs of interest and amortization were calculated using a depreciation period of 10 years, and interest rate of 4.2%[20]. Yearly maintenance costs of 5% of the average acquisition costs for general inventory and 8% for imaging equipment were included (Table 1). The percentage of maintenance costs for imaging equipment was based on expert opinion and only focusses on costs directly related to the inventory, as no formal estimate is known for any other additional cost such as personnel costs.

Table 1 - Input parameters.

Parameter	Input value	Input value Specific per OR setting		Source
		Conventional (range)	Hybrid (range)	
General				
Surface of OR (m^2)		48.5 (40.6-57.0)	85.4 (52.3-106.2)	Hospital data
OR availability per year (min)		122,400	122,400	Hospital data
Utilization rate of OR		92% (87-96)*	43% (14-55)*#	and available hours assumed to be the same for the hybrid OR
Construction				
Costs for a general m^2 in a hospital	€3,479			[21]
Differentiation rate over a general square meter in a hospital per category (%):				
OR department	160%			[21] and expert opinion
Holding	140%			
Recovery	140%			
Corridor and offices	75%			
Technical rooms	75%			
Sanitary/washing rooms	100%			
Annual interest (%)	4.2			[20]
Lifespan (years)	25			[20]
Maintenance (%)	5			[20]
Inventory				
Lifespan (years)	10			[20]
Maintenance for general inventory	5%			[20]
Maintenance for medical imaging devices	8%			Expert opinion
Personnel				
Costs per hour academic setting				
Medical specialist	€117.59			[20,23]
Medical assistant	€36.84			
Technician			€36.84	
Costs per hour general hospital				[20,23]
Medical specialist	€120.71			
Medical assistant	€38.06			
Technician			€38.06	
Annual loaded working hours				
Medical specialist	2,100			[20]
Medical assistant	1,558			
Technician			1,558	
Overhead				
Calculated over construction and personnel	38%			[20]

*based on only 4 hospitals # Utilization rate of hybrid OR has a slightly different definition than the utilization rate of the conventional OR: annual number of procedures with the use of the C-arm divided by the total annual procedures in the hybrid OR, multiplied with the utilization rate of a conventional OR.

Personnel costs

The personnel costs were based on the number of staff needed for a surgical procedure. We used this approach, and not the actual annual personnel spending of the different hospitals, to overcome the differences between the hospitals.

Per hospital, we identified the composition of the OR team that is available during a surgery in general in the conventional OR and the hybrid OR. The personnel costs per hour were calculated by dividing the total number of effective working hours per year of each function by the total annual loaded salary[20]. The total annual loaded salaries were retrieved from collective labor agreements for academic and general hospitals[23,24]. The costs of a medical specialist were obtained from the Dutch guideline[20]. To account for costs related to irregular working hours, holiday allowance and social security, the salaries were corrected by a percentage of 39% for general personnel and 35% for higher (medical) personnel. The total personnel costs were calculated by taking the mean of these hospital estimations.

Overhead

Overhead expenses, the costs that are not directly attributable to a particular resource but are essential in providing care e.g. electricity, water, cleaning service, and administrative tasks, were only calculated over the construction costs and personnel costs to avoid double counting. We used the general percentage for overhead on the direct costs for medical departments of 38%, as recommended by the Dutch guideline[20]. The expected higher overhead costs for the hybrid OR (e.g. larger demand in electricity) are incorporated by calculating the overhead over the average construction costs because the surface of a hybrid OR is larger than the surface of the conventional OR.

Analysis

The costs from the bottom-up cost analysis, except the costs for personnel, are expressed in yearly costs. Those are the average costs of the five consulted hospitals. To calculate the total costs per minute, it was needed to combine the average total yearly costs with the average utilization rate. Therefore, the available hours and occupied hours of a conventional OR were identified per hospital for 2018. This resultedin an average utilization rate of the conventional OR. As cross-functional use of the hybrid OR was encountered in several hospitals, we calculated the utilization rate only based on the procedures that were performed with the fixed C-arm, i.e. the hybrid OR was used as such and not as a conventional OR. The utilization rate was obtained by dividing the annual number of procedures with the fixed C-arm by the total annual procedures in the hybrid OR, multiplied with the utilization rate of a conventional

OR. The average yearly costs for both ORs were divided by the average occupied hours per year. The sum of these costs and the average hourly personnel costs resulted in a cost per hour for both ORs. Those costs were converted to costs per minute (Supplement 4).

To evaluate the cost drivers for the conventional and hybrid OR, a deterministic sensitivity analysis was performed. The main cost parameters were varied using the minimum and maximum values identified in the five hospitals. To evaluate the influence of personnel costs and overhead percentage, we used a 10% upper and lower limit.

In addition, as a second sensitivity analysis, the utilization rate was varied from 30% to 100% for both ORs to show the influence on the costs per minute. As a final sensitivity analysis for the hybrid OR, the influence of different utilization rates of the fixed C-arm, varying from 10% to 100%, on the costs per minute were evaluated. In this analysis, it was assumed that the hybrid OR was optimally used (utilization rate of 92%) since the hybrid OR can also be used for surgeries without the use of a fixed C-arm (cross-functional).

RESULTS

Base case cost analysis

Table 2 shows the costs per cost category, the total annual costs and costs per minute for both ORs. The average surface of a conventional and hybrid OR was 48.52m^2 and 85.36m^2, respectively.

For the conventional OR the estimated annual construction costs were €71,673 (range €47,141-€97,118), the annual inventory costs were €113,330 (€63,710-€153,379) and the hourly personnel costs were €328 (€324-€333). Over construction costs, the annual overhead was €27,236 (€17,914-€36,905) and over personnel, the hourly overhead costs were €125 (€123-€127).

For the hybrid OR the estimated annual construction costs were €126,092 (€60,726-€180,945), the annual inventory costs were €433,241 (€329,938-€519,947) and the hourly personnel costs were €366 (€361-€371). The overhead costs over the construction costs were €47,915 (€23,076-€68,759) and €139 (137-€141) over personnel per hour. The fixed C-arm accounted for 40% of the total fixed costs. Figure 1 shows the fixed annual costs related to both ORs.

The mean utilization rate for the conventional and hybrid OR were 92% (range 87%-96%) and

48% (range 14%-55%), respectively. These mean utilization rates were based on four hospitals because of a registration problem in one hospital. The detailed costs per cost category can be found in Table 2. For the conventional OR we found a total cost per minute of €9.45 (€8.60-€10.23), for which the personnel costs amounted to 58%. The total costs per minute for the hybrid OR were €19.88 (€16.10-€23.07). Since the costs for construction (12%) and inventory (41%) substantially increased compared to the conventional OR, the personnel costs amounted only to 31%. The costs per minute for both ORs are shown in Figure 2.

Table 2 - Base case results from the bottom-up cost analysis including VAT and presented in 2018 Euros.

Parameter	**Conventional OR**	**(Range)**	**Hybrid OR**	**(Range)**
Construction				
Costs per m^2 OR	€12,804	(€10,064- €14,768)	€12,804	(€10,064- €14,768)
Total construction costs	€621,231	(€408,598- €841,776)	€1,092,915	(€526,347- €1,568,362)
Total annual costs	€71,673	(€47,141- €97,118)	€126,092	(€60,726- €180,945)
Inventory				
General inventory	€230,421*	(€190,027- €255,000)	€244,421*	(€190,027- €296,500)
Anesthesia	€128,308	(€65,000- €177,052)	€128,308	(€65,000- €177,052)
OR lights	€38,592	(€20,030- €73,134)	€42,598	(€32,000- €73,134)
Arm pendants	€69,262	(€40,000- €96,450)	€94,789	(€56,000- €135,333
OR table	€105,891	(€50,000- €155,160)	€265,222	(€215,000- €307,131)
Image routing system	€76,812	(€0- €122,059)	€93,126	(€0- €150,000)
X-ray radiation protection aprons	-	-	€18,384 *	(€13,354- €23,682)
Fixed C-arm	-	-	€1,361,554	(€1,125,649- €1,550,000)
Total inventory costs	€649,375	(€365,057- €878,855)	€2,248,402	(€1,697,029- €2,712,832)
Total annual costs	€113,330	(€63,710- €153,379)	€433,241	(€329,938- €519,947)
Personnel				
Staff occupation per OR	Surgeon (1x)	€ 1.99	Surgeon (1x)	€ 1.99
	Anesthetist (0.5x)	€ 0.99	Anesthetist (0.5x)	€ 0.99
	Medical assistant (1x)	€ 0.62	Medical assistant (1x)	€ 0.62
	OR assistant (3x)	€ 1.87	OR assistant (3x)	€ 1.87
	Technician (0x)	€ 0	Technician (1x)	€ 0.62
Total personnel costs	€5.48	(€5.40- €5.56)	€6.10	(€6.01- €6.19)
Overhead				
Personnel costs (per minute)	€2.08	(€2.05- €2.11)	€2.32	(€2.28- €2.35)
Construction costs (annual)	€27,236	(€17,914- €36,905)	€47 915	(€23,076- €68,759)
(per minute)	€0.24	(€0.16- €0.33)	€0.90	(€0.44- €1.30)
Total overhead costs	€2.32	(€2.21- €2.41)	€3.22	(€2.72- €3.65)
Costs per minute (%)				
Construction costs	€0.64 (6.8%)	(€0.42- €0.87)	€2.38 (12.0%)	(€1.15- €3.42)
Inventory costs	€1.01 (10.7%)	(€0.57- €1.37)	€8.18 (41.1%)	(€6.23- €9.81)
Personnel costs	€5.48 (57.9%)	(€5.40- €5.56)	€6.10 (32.4%)	(€6.01- €6.19)
Overhead costs	€2.32 (24.6%)	(€2.21- €2.41)	€3.22 (16.2%)	(€2.72- €3.65)
Total	€9.45	(€8.60- €10.23)	€19.88	(€16.10- €23.07)

*based on only four hospitals. One hospital had in both conventional and hybrid OR no imaging routing system, therefore the minimal costs of the subcategory "image routing system" are zero.

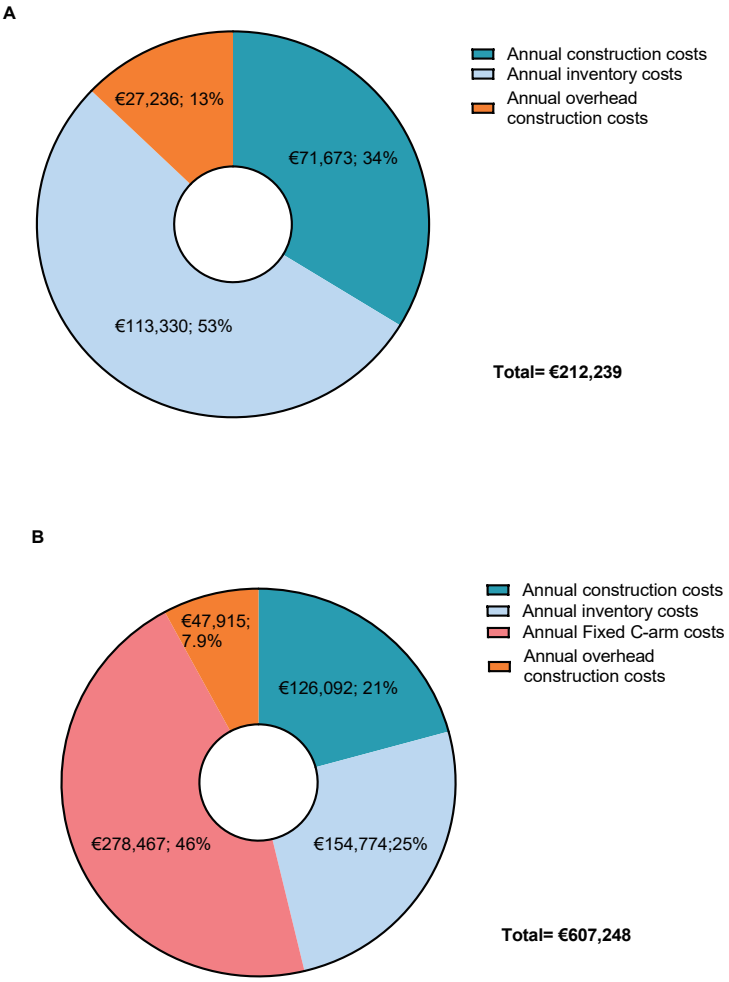


Figure 1. Total fixed costs of the conventional OR (A) and hybrid OR (B).

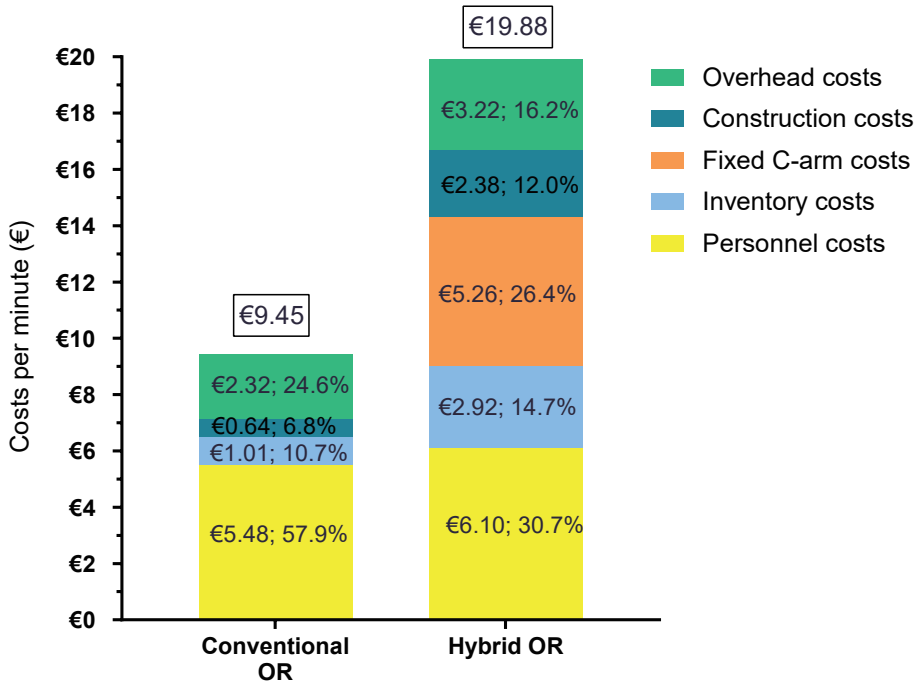


Figure 2. Total costs per minute for the conventional OR and hybrid OR. In this figure the costs for the fixed C-arm are separated from the inventory costs.

Cost drivers

Figure 3 shows the results from the deterministic sensitivity analysis. The total personnel costs had most influence on the total costs of the conventional OR, followed by total inventory and overhead costs. The inventory costs were mainly driven by the costs for imaging routing system, anesthesia and OR table costs. For the hybrid OR, the utilization rate had most influence on the total costs, followed by the inventory costs and construction costs. The inventory costs were mainly driven by the costs of the fixed C-arm. Figure 4a shows that less efficient utilization of the conventional OR, for instance, 50%, results in an increase of €1.73 per minute compared to a fully used OR (100%). For the hybrid OR a difference in cost per minute of €4.96 was seen, comparing a utilization rate of 50% to 100%. Figure 4b shows that using the fixed C-arm only half of the time, assuming a utilization rate of 92% for the hybrid OR, results in a total OR costs of €19.17 per minute.

Calculation model

As seen by Figure 3, the input parameters such as specific inventory costs can have a large influence on our calculated costs per minute. As those costs are different per hospital and country, a calculation model (Supplement 1) was attached to this article in which input parameters can be adjusted.

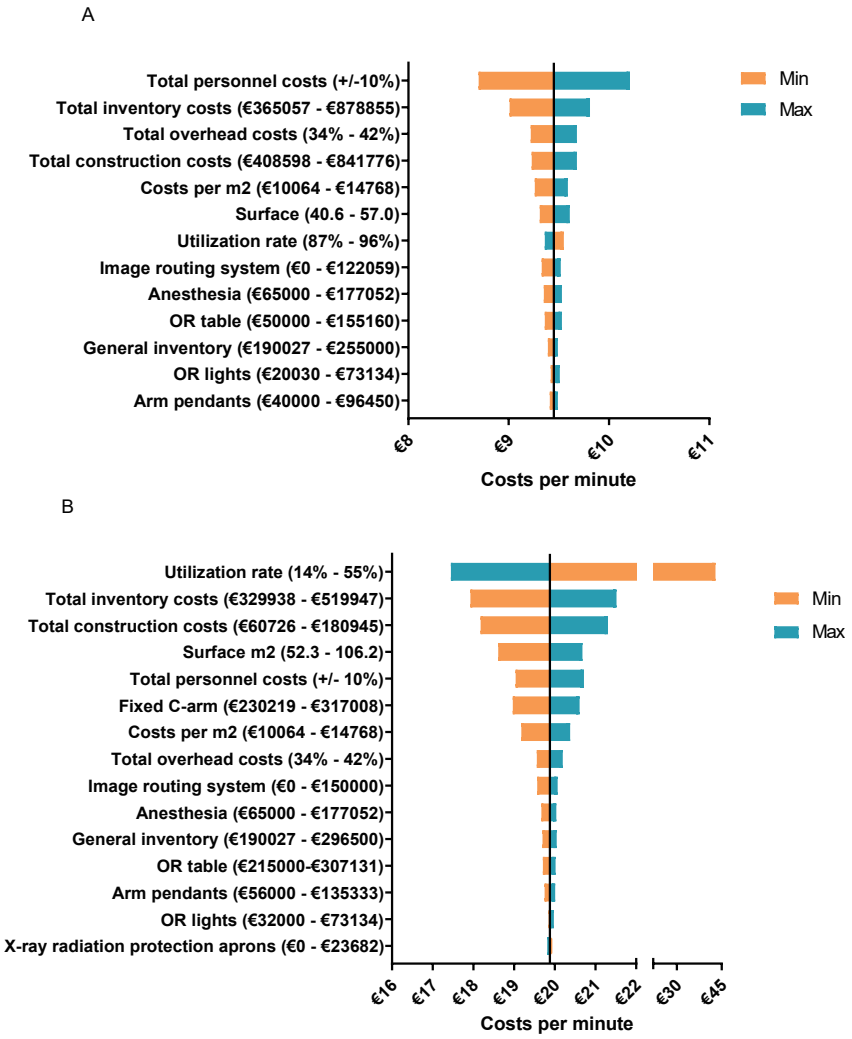


Figure 3. Results of the deterministic sensitivity analysis for both the conventional OR (A) and hybrid OR (B) in a tornado diagram. The values between brackets show the minimal and maximum values used in the sensitivity analysis.

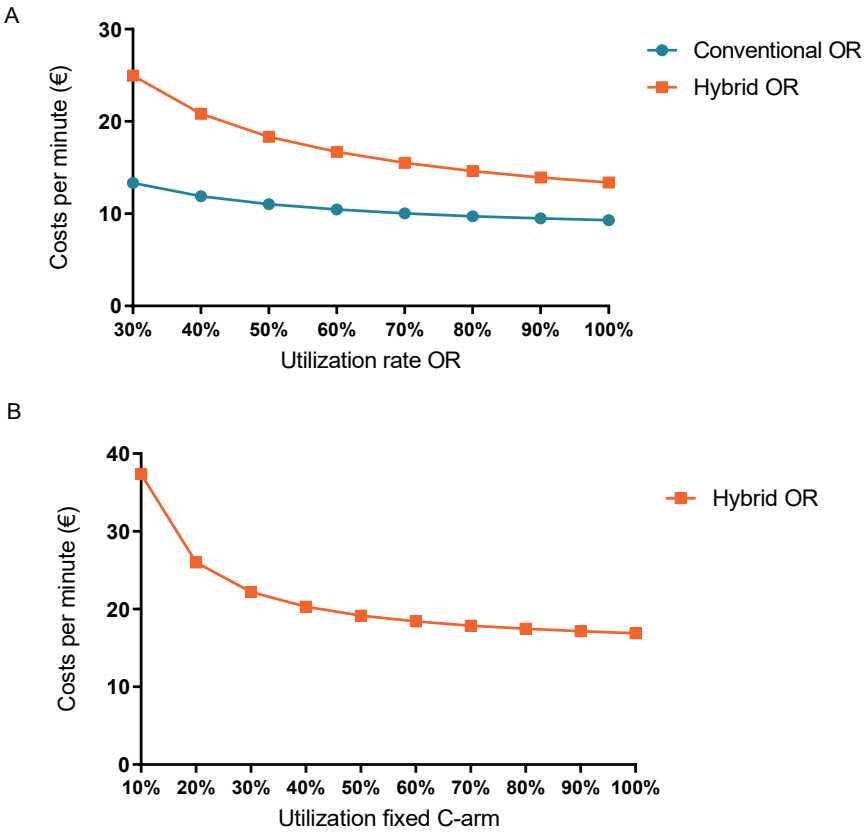


Figure 4. (A) Costs per minute of both conventional OR and hybrid OR by the utilization rate of the OR. It was assumed that the C-arm was used during all procedures performed in the hybrid OR. (B) Costs per minute of the hybrid OR by utilization rate of the fixed C-arm assuming an utilization rate of 92% for the hybrid OR.

DISCUSSION

The identified costs per minute for the conventional OR were €9.45 (€8.60-€10.23) and for the hybrid OR €19.88 (€16.10-€23.07), measured in a setting of five hospitals in the Netherlands. The difference between the two ORs (€10.43) is mainly explained by higher inventory costs for the hybrid OR namely, €7.17. Noteworthy is that although advanced imaging technology is expensive, personnel costs remain an important element in the costs of both ORs (58% and 31% of the costs for the conventional and hybrid OR, respectively). For the hybrid OR, the construction costs became an important element compared to the conventional OR (increase from 7% to 12% for the per-minute costs), which is explained by the larger room needed.

Comparing our results with available results in the literature is challenging because existing studies include different input parameters and adopt different perspectives. For instance, a mean cost of $37.45 per minute was found for a conventional OR in the USA[13]. This study incorporated costs for disposables and had a top-down perspective which presumably explains the difference to our results. Raft et al. found a cost per minute of €10.80 for using the OR and PACU by a top-down cost analysis. Since they also incorporated the costs of the PACU, medicines, and disposables used during operations the results are hard to compare[14]. As purchasing power parity differences are rather modest, this might reflect differences in norms for capital expenditure between US and European hospitals; it is however likely that the relative difference between conventional and hybrid ORs is comparable.

To date, no studies that solely evaluated the costs of using a hybrid OR were identified. However, some studies compared a specific hybrid surgery with a non-hybrid surgery in which intervention costs were taken into account[8,25–27]. Two studies showed that the hybrid approach reduced operating time, length of stay and resulted in less resource use for cardiovascular surgery. They however neglected the potential higher intervention cost for using the hybrid OR[8,26]. Another cardiovascular study evaluated the additional costs for the hybrid approach and reported higher costs, but took a reduction of procedure time for the hybrid approach into account in reporting the total costs. Therefore, it is not possible to deduce the additional costs only for the use of the hybrid OR[25].

To the best of our knowledge, this is the first bottom-up cost analysis that provides insight into the costs and cost drivers for both the conventional and the hybrid OR. The main strength of our study is that we evaluated the costs transparently by (1) performing a bottom-up cost-analysis, (2) specifically stating results for the different cost categories, (3) comprehensively describing the methods and calculations used in the analysis and (4) enclosing a calculation model in which input parameters can be changed to specific settings for instance, to incorporate the costs of a mobile C-arm in the conventional OR or include costs for endoscopic devices and disposables (Supplement 1). Another strength in this study was the comparison and anonymous discussion of the data of the five hospitals, academic and non-academic, to result in a general cost price for the Dutch setting.

The present study has several limitations. First, the base case analysis assumed that personnel had no idle time and a minimal team needed for running the (hybrid) OR was taken into account which underestimates the actual costs. Furthermore, the correction for irregular hours on personnel costs was assumed equal for the conventional and hybrid OR. This might underestimate the personnel costs for the hybrid OR, when this OR is often used during irregular hours (e.g. in trauma interventions). This underestimation could influence our results as personnel costs are an important factor driving the costs of both ORs (Figure

3). This topic should therefore be evaluated in the future. Second, the construction costs were evaluated based on square meters, to avoid hospital differences and choices in design. Especially for the hybrid OR, these may not reflect the actual costs since we did not evaluate the specific costs of the conditional adaptations, e.g. a larger room, lead lining in the walls and additional installations[28]. In addition, the size of the hybrid OR differed substantially in our analysis (range: 52.3-106.2) by using the average surface we may underestimate these costs as well. Third, as we only included Dutch hospitals, the generalizability of our results may be limited. Therefore, we contacted one additional hospital in a high-income country (Oslo University Hospital), which recently built innovative operating suites, to verify the construction and inventory costs. The costs of this hospital seemed comparable to the Dutch setting. Also, their construction costs, which were based on the actual construction costs, were in line with our estimates. Finally, overhead costs are very hard to obtain using a bottom-up methodology, therefore we chose a fixed percentage over the construction costs and personnel costs defined in the Dutch manual for cost calculations[20]. This could result in an over or underestimation of the indirect costs. However, ranging the overhead over 34% and 42% (base case 38%) had a relatively small influence on the total costs per minute. Also, the maintenance costs for imaging equipment were difficult to obtain, especially when considering that additional staff is necessary to support the maintenance activities within the hospitals. The estimated maintenance of 8%, which was obtained through expert consultation, might therefore be an underestimation. As in our sensitivity analysis the upper and lower values of the C-arm showed a relatively limited effect on the results, we expect that a different percentage for maintenance costs would not change our conclusions.

The following example gives an impression on the added value per patient needed to consider the use of the fixed C-arm (in the hybrid OR) cost-effective. From the estimated costs for the conventional and hybrid OR, we can calculate the base case incremental costs for the hybrid OR per minute (€10.43). Assuming a surgical procedure of approximately 2.5 hours, for example, an endovascular aneurysm repair (EVAR), results in an incremental cost of €1.565 per procedure. These incremental costs only incorporated the additional costs of using the hybrid OR, not additional materials that might be needed. When a cost-effectiveness threshold of €50,000 per QALY is assumed[29], a minimum gain of 0.031 QALY is required. This means, using this simplified calculation, that for performing an EVAR in the hybrid OR, at least 11 days in perfect health should be gained to be considered cost-effective compared to a conventional OR. This is based on the utilization rate of 43%. When calculating the incremental costs using a different procedure duration, an estimation can be made for other procedures as well. Based on this calculation it may seem difficult for the hybrid OR to become cost-effective, however by increasing the utilization rate or comparing the intervention to a different technique, e.g. laparoscopy which results in lower incremental costs, it becomes more likely that the hybrid OR is cost-effective. This suggestion is strengthened by a recent

evaluation of a navigation system that is used during surgery, which requires the hybrid OR. In removing locally recurrent rectal tumors the use of a navigation system in the hybrid OR has the potential to become cost-effective. However, the level of cost-effectiveness of the navigation system is also strongly dependent on the utilization rate of both the navigation system and the hybrid OR.[30]

Conclusion

This study identified that the main factors influencing the OR costs are: total inventory costs, total construction costs, utilization rate, and total personnel costs. Therefore, our results could inform surgeons, decision and policy makers in hospitals on the adoption and optimal utilization of new (hybrid) ORs. Although it seems that hospitals have to strive to use the fixed C-arm as often as possible, one should keep in mind that using the hybrid OR should add value to the patient. To evaluate the added value, the calculation model (Supplement 1) can be used to evaluate the costs and estimate the required added value for a specific setting and/or country. These estimations can be useful in identifying the most promising procedures performed in the hybrid OR to steer future research directions.

Future directions

As it is expected that the number of hybrid ORs will rise worldwide[1], and those are likely to become more advanced (e.g. including an MRI) it is important to evaluate its (cost-) effectiveness. Therefore, in the near future, prospective comparative studies have to be performed to evaluate the actual benefits of using these advanced ORs in terms of complication rates, efficiency, and survival. Those outcomes can be used to identify interventions that yield the most from the advances of the hybrid OR, as well as informing cost-effective usage of the hybrid OR in general. Finally, in order to assess the generalizability of our results to non-European or low and middle-income countries it would be valuable to validate our calculation model from a different perspective (e.g. US perspective).

ACKNOWLEDGEMENTS

The authors would like to thank all the participating hospitals that contributed to this study by delivering their hospital data.

REFERENCES

1. Market Research Future. *Global Hybrid Operating Room Market Insights, forecast to 2023*. (2019).
2. Kpodonu, J. Hybrid cardiovascular suite: The operating room of the future. *Journal of Cardiac Surgery* vol. 25 704–709 (2010).
3. Hertault, A. *et al.* What should we expect from the hybrid room? *J. Cardiovasc. Surg. (Torino).* **58**, 264–269 (2017).
4. Kaneko, T. & Davidson, M. J. Use of the hybrid operating room in cardiovascular medicine. *Circulation* **130**, 910–7 (2014).
5. Cabo, R., McIff, G. B., Abate, E. & Amato, S. F. PHP51 Review of the Clinical and Economic Value in the Use of Hybrid Operating Room. *Value Heal.* **15**, A22 (2012).
6. Smeltzer, H. G. *et al.* Collaboration between Interventional Neurosurgery and Vascular Surgery in the Hybrid Operating Room. *J. Radiol. Nurs.* (2014) doi:10.1016/j.jradnu.2014.02.002.
7. Singhal, S. The future of surgical oncology image-guided cancer surgery. *JAMA Surgery* vol. 151 184–185 (2016).
8. Babaliaros, V. *et al.* Comparison of Transfemoral Transcatheter Aortic Valve Replacement Performed in the Catheterization Laboratory (Minimalist Approach) Versus Hybrid Operating Room (Standard Approach). *JACC Cardiovasc. Interv.* **7**, 898–904 (2014).
9. Varu, V. N., Greenberg, J. I. & Lee, J. T. Improved Efficiency and Safety for EVAR with Utilization of a Hybrid Room. *Eur. J. Vasc. Endovasc. Surg.* **46**, 675–679 (2013).
10. Muñoz, E., Muñoz, W. & Wise, L. National and Surgical Health Care Expenditures, 2005–2025. *Ann. Surg.* **251**, 195–200 (2010).
11. Birkmeyer, J. D. *et al.* Medicare Payments for Common Inpatient Procedures: Implications for Episode-Based Payment Bundling. *Health Serv. Res.* **45**, 1783–1795 (2010).
12. Hirst, A. *et al.* No Surgical Innovation Without Evaluation: Evolution and Further Development of the IDEAL Framework and Recommendations. *Ann. Surg.* **269**, 211–220 (2019).
13. Childers, C. P. & Maggard-Gibbons, M. Understanding Costs of Care in the Operating Room. *JAMA Surg.* **153**, e176233 (2018).
14. Raft, J., Millet, F. & Meistelman, C. Example of cost calculations for an operating room and a post-anaesthesia care unit. *Anaesth. Crit. Care Pain Med.* **34**, 211–215 (2015).
15. Kaplan, R. S. & Porter, M. E. How to solve the cost crisis in health care. *Harv. Bus. Rev.* (2011).
16. Najjar, P. A., Strickland, M. & Kaplan, R. S. Time-Driven Activity-Based Costing for Surgical Episodes. *JAMA Surg.* **152**, 96 (2017).
17. Macario, A. What does one minute of operating room time cost? *J. Clin. Anesth.* **22**, 233–236 (2010).
18. Lievens, Y., van den Bogaert, W. & Kesteloot, K. Activity-based costing: a practical model for cost calculation in radiotherapy. *Int. J. Radiat. Oncol.* **57**, 522–535 (2003).
19. Ismail, I., Wolff, S., Gronfier, A., Mutter, D. & Swantröm, L. L. A cost evaluation methodology for surgical technologies. *Surg. Endosc.* **29**, 2423–2432 (2015).
20. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. *Manual for cost research: methods and standard cost prices for economic evaluations in health care*. (2015).
21. Dutch advisory board on healthcare housing for building costs in a hospital. *Handreiking kengetallen benchmark zorgvastgoed (Guidance benchmark indices healthcare property in Dutch)*. (2018).
22. StatLine. Dutch price index numbers. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?ts=1534249924434 (2020).
23. Nederlandse Vereniging van Ziekenhuizen (NVZ). *Salarisschalen Cao Ziekenhuizen (2018) (Dutch collective labor agreement)*. (2018).
24. Dutch Federation of Academic Medical Centers. *Collective labor agreement 2018-2020 for academic medical centers*. (2018).
25. Rimestad, J. M., Christiansen, E. H. & Modrau, I. S. One-year cost-effectiveness and safety of simultaneous hybrid coronary revascularization versus conventional coronary artery bypass grafting. *Interact. Cardiovasc. Thorac. Surg.* **29**, 217–223 (2019).

26. Attigah, N. *et al.* Hybrid operating rooms versus conventional operating rooms: Economic comparisons in vascular surgery using the example of endovascular aneurysm repair [Hybridoperationssaal vs. konventioneller Operationssaal: Betriebswirtschaftliche Gegenüberstellung in d. *Chirurg.* **88**, 587–594 (2017).
27. CADTH. *Operating room versus interventional suite for interventional radiation procedures: cost -effectiveness*. (2017).
28. Kant, S., Gupta, S. K. & Siddharth, V. Planning Premises and Design Considerations for Hybrid Operating Room. *Int. J. Res. Found. Hosp. Healthc. Adm.* **2**, 50–56 (2014).
29. Versteegh, M. M. *et al.* Severity-Adjusted Probability of Being Cost Effective. *Pharmacoeconomics* **37**, 1155–1163 (2019).
30. Lindenberg, M. *et al.* Image-guided navigation for locally advanced primary and locally recurrent rectal cancer: Evaluation of its early cost-effectiveness. in (J Clin Oncol 38: 2020 (suppl; abstr e16015), 2020).

SUPPLEMENTARY MATERIAL

Supplement 1 - Calculation model of the costs of a conventional and a hybrid OR

Supplement 2 - Characteristics of participating hospitals

Hospital	Number of beds	Number of ORs	Involved experts
A	212 (2019)	12	Anesthetist Department building management Financial department Head of the OR Staff involved in capacity measurements
B	766 (2019	22	Head of the OR Department building management Board of directors Staff involved in capacity measurements
C	593 (2018)	20	Head of the OR Coordinator OR technology Managing director OR department Department building management Financial department Staff involved in capacity measurements
D	776 (n.d.)	14	Clinical physicist Medical technician Staff involved in capacity measurements Project manager OK
E	630 (2017)	16	Clinical physicist Staff involved in capacity measurements Technical coordinator Coordinator surgery in the OR department

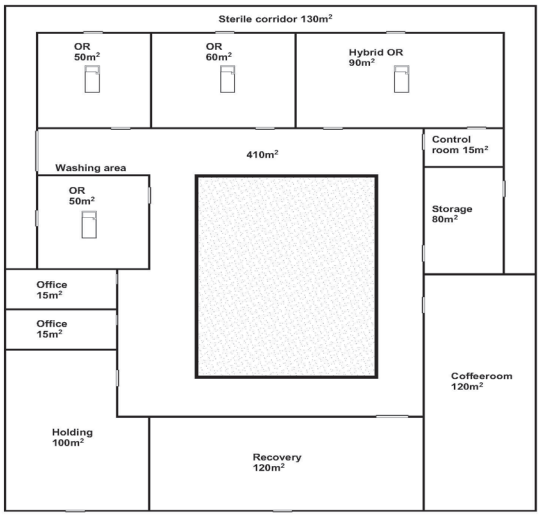
Supplement 3 - Example of the calculation of the construction costs

The following formulas were mentioned in the article:

$$1 \qquad \sum C_{M2\ not\ labeled\ as\ OR} = \sum M^2_{functionality\ i} * C_{differentiated\ i}$$

$$2 \qquad \frac{(\sum M^2_{OR\ department} * C_{M2\ costs\ OR\ department}) - \sum C_{M2\ not\ labeled\ as\ OR}}{\sum M^2_{OR\ within\ OR\ department}}$$

This supplement gives an example of the construction cost calculation. The "C" in the formula refers to costs and the "M^2"to square meters. Picture 1 shows a simplified OR department containing corridors, ORs, a storage room, offices, the holding, the recovery and a place to have lunch and coffee.



Picture 1. A simplified OR department

The total surface of this OR department is 1195m^2.
Costs for constructing a m^2 = €3,479.
Costs for constructing a m^2 in an OR department = €5,566 (differentiated with 160%).

The first part of the formula focuses on the surface of the OR department that is not belonging to the OR.
In this example that means: 30m² office, 100m² holding, 120m² recovery, 120m² coffee room, 80m² storage room, 410m² standard corridor which sums up to: 860m²
The sterile corridor and control room were included in the surface belonging to the OR = 395m².

$$1 \qquad \sum C_{M2\ not\ labeled\ as\ OR} = \sum M^2_{functionality\ i} * C_{differentiated\ i}$$

Filling in this formula gives the table below:

$M^2_{functionality\ i}$	Differentiated with	$C_{differentiated\ i} * M^2_{functionality\ i}$
Moffice = 30m²	100%	= (3,479 * 100%) * 30 = 104,370
Mholding = 100m²	140%	= (3,479 * 140%) * 100 = 487,060
Mrecovery = 120m²	140%	= (3,479 * 140%) * 120 = 584,472
Mcoffee room = 120m²	75%	= (3,479 * 75%) * 120 = 313,110
Mstorage = 80m²	75%	= (3,479 * 75%) * 80 = 208,740
Mcorridor = 410m²	75%	= (3,479 * 75%) * 410 = 1,069,792

$\sum C_{M2\ not\ labeled\ as\ OR}$	= € 2,767,544

This number is used in the next formula

$$2 \qquad \frac{(\sum M^2_{OR\ department} * C_{M2\ costs\ OR\ department}) - \sum C_{M2\ not\ labeled\ as\ OR}}{\sum M^2_{OR\ within\ OR\ department}}$$

Two of the four variables were given at the beginning:
M² OR department = 1,195m²
M² costs of OR department = €5,566

The final variable can be extracted from Picture 1:
m² of OR within OR department = 50 + 50 + 60 + 90 + 15 + 130 = 395 m²

We can fill in the formula:
= ((1,195*5,566) – 2,767,544) / 395
= € 9,832

In this example constructing a m² of an OR would cost €9,832.

Supplement 4 - Formulas to calculate the costs per OR

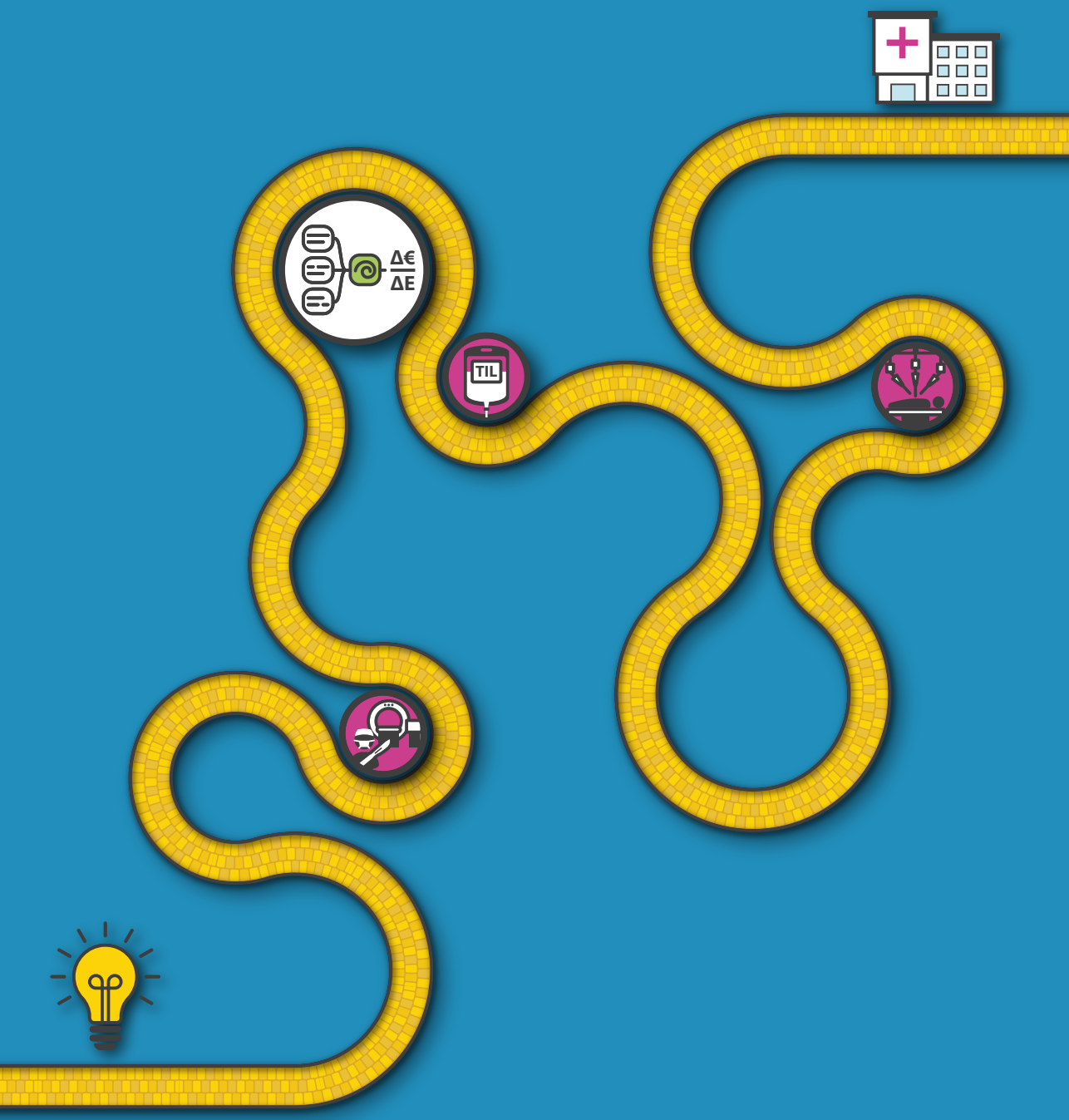
In this analysis the formulas used to calculate the OR costs for both the hybrid and conventional OR were as follows:

Costs of the OR per hour =

$$\frac{Annual\ construction\ costs + annual\ inventory\ costs + annual\ overhead\ costs}{Utilization\ rate * available\ production\ hours\ of\ OR} + hourly\ personnel\ costs$$

Costs of the OR per minute =

$$\frac{Costs\ of\ the\ OR\ per\ hour}{60}$$


ΔE
TIL

6

Image-guided navigation for locally advanced primary and locally recurrent rectal cancer: Evaluation of its early cost-effectiveness

Melanie Lindenberg*
Astrid Kramer*
Esther Kok
Valesca Retèl
Geerard Beets
Theo Ruers
Wim van Harten

*Contributed equally

Submitted

ABSTRACT

Background

Use of an image-guided navigation system showed improved resection margin rates in locally advanced (LARC) and locally recurrent rectal cancer (LRRC) patients. Incremental surgical innovation is often implemented without reimbursement consequences, health economic aspects should however also be taken into account. This study evaluates the early cost-effectiveness of navigated surgery compared to standard surgery in LARC and LRRC.

Methods

A Markov decision model was constructed to estimate the expected costs and outcomes for navigated and standard surgery. The input parameters were based on pilot data from a prospective (navigation cohort n=33) and retrospective (control group n=142) data. Utility values were measured in a comparable group (n=63) through the EQ5D-5L. Additionally, sensitivity and value of information analyses were performed.

Results

Based on this early evaluation, navigated surgery showed incremental costs of €3,141 and €2,896 in LARC and LRRC. In LARC, navigated surgery resulted in 2.05 quality-adjusted life years (QALYs) vs 2.02 QALYs for standard surgery. For LRRC, we found 1.73 vs 1.67 QALYs respectively. This showed an incremental cost-effectiveness ratio (ICER) of €136,604 for LARC and €52,510 for LRRC per QALY gained. In scenario analyses, optimal utilization rates of the navigation technology lowered the ICER to €61,817 and €21,334 for LARC and LRRC.

Conclusion

Adding navigation system use is expected to be cost-effective in LRRC and has the potential to become cost-effective in LARC. To increase the probability of being cost-effective, it is crucial to optimize efficient use of both the hybrid OR and the navigation system and identify subgroups where navigation is expected to show higher effectiveness.

INTRODUCTION

Rectal cancer is mainly treated by surgical resection, often complemented with pre- and/or postoperative (chemo)radiotherapy in stage II-IV tumors[1–3], showing a 5-year survival rate of ~45% for stage III and ~20% for stage IV tumors[4]. Surgical resection of both locally advanced (LARC) and locally recurrent rectal cancer (LRRC) requires special consideration because the disruption of normal anatomical planes and radiotherapy-induced fibrosis can lead to a higher risk of a tumor positive involved circumferential resection margin[1,5]. In 10%-15% of rectal cancer patients, positive surgical margins are found[6,7] which negatively affects the prognosis[8–10]. Local recurrence can cause debilitating symptoms, and often requires additional treatment, such as chemoradiotherapy and radiotherapy. Optimizing surgical practice and decreasing the risk of positive resection margins is therefore of great clinical and financial importance.

Multiple technologies have emerged to improve the quality of surgery and surgical outcomes[11]. The Netherlands Cancer Institute (NKI-AVL) has developed an image-guided navigation system to improve tumor localization during the operative procedure and prevent damage to surrounding vital structures (see Box 1)[12]. Recently, this navigation system has been evaluated in the first series of LARC and LRRC patients, showing substantially improved negative surgical margin rates compared to standard surgery in a historical control group[13]. Since the use of a navigation system is associated with extra costs (e.g. due to extra imaging, the navigation system, and personnel), and hospital budgets are limited, new surgical technologies have to prove themselves in terms of cost-effectiveness to have a chance of reimbursement.

To evaluate the potential value of this navigation system to inform policymakers and to guide subsequent decisions on for further research and development[14], early cost-effectiveness analyses can be performed. This study evaluates the early cost-effectiveness of the image-guided navigation system used during surgery for LARC and LRRC patients compared to standard surgery based on the first clinical data sampled in the Netherlands Cancer Institute[13].

Box 1 - Intervention of interest and usual care.

Intervention of interest: Navigated surgery

The addition of the navigation system for rectal surgery in patients with LARC or LRRC changed the regular workflow before and during surgery. One day before surgery, a multiphase contrast- enhanced CT-scan (with early arterial and excretion phase) was acquired. A 3D anatomical model was made, including the most important anatomical structures (blood vessels, ureters, bones and targets). Before surgery, in the hybrid operating room, three patient trackers were taped to the skin of the patient. Subsequently, a cone-beam CT-scan was performed. The acquired intraoperative images were registered to the preoperative CT-scan. During surgery, the patient lied on a specific imaging bed including an electromagnetic field generator. This enabled visualization of the patient trackers and a tracked pointer in the 3D anatomical model. During surgery, the tracked pointer was used by the surgeon to navigate towards the tumor. A more detailed description of the navigation system can be found in the article of Nijkamp et al., 2018[12].

Standard surgery

The rectal resection of LARC or LRRC is performed with an abdomioperineal resection (APR) or a low anterior resection (LAR), with or without resection of the surrounding organs (exenterative procedures, sacral bone etc. and intra-operative radiotherapy, depending on the patient and tumor characteristics (e.g. tumor location, previous surgeries, etc.). Procedures can be performed open or laparoscopically.

6

METHODS

Study design and model structure

The analysis was based on a combination of a decision tree resulting in a positive (R1) or negative resection margin (R0) after standard and navigated surgery[13] and a Markov model comprising the mutually exclusive health states: “disease-free”, “progression of disease” and the absorbing state “death” (Supplement 1). In the model, all patients start in “disease free” and could either remain in “disease free” or transfer to “progression of the disease” or “death”. Since the course of disease for LARC and LRRC is different, two separate models were constructed with a similar design. The time horizon was set at three years because most recurrences develop in the first three years after (curative) resection[15], and recent literature reported a median survival time of 37[16] and 30 months[17] for LARC and LRRC, respectively. A cycle time of 3 months was chosen according to guidelines for follow-up visits[18]. The analysis was performed from a Dutch healthcare perspective, using the Dutch guideline for health economic costing studies[19].

Population

The patient population of the study of Kok et al. was used[13]. They prospectively included 33 patients who received navigated surgery for either LARC (n=14) or LRRC (n=19) between

2016-2019 in the Netherlands Cancer Institute (NKI-AVL). As a control group, 142 patients having standard surgery for LARC (n=101) and LRRC (n=41) were evaluated by means of a retrospective cohort study. These patients had a similar indication and type of surgery at the NKI-AVL[13]. Supplement 2 shows the characteristics of this patient population[13]. The Institutional Review Board of the NKI-AVL approved data extraction for the included patients. Since no representative Quality of Life (QoL) data was available in literature[20], we used data from 63 patients of an ongoing prospective cohort study of patients with colorectal cancer who undergo standard and navigated surgery within the NKI-AVL with similar inclusion criteria. The clinical characteristics of these patients are presented in Supplement 3. Based on these clinical characteristics, the group was judged sufficiently comparable to the control group to be used in our analysis. Supplement 4 shows a schematic overview of the data sources used for the input parameters.

Input parameters

The input parameters are presented in Table 1.

Clinical effectiveness

Among LARC patients, 93% R0 resections were achieved after navigated- and 84% after standard surgery. Among LRRC patients, 79% had an R0 resection after navigated- and 49% after standard surgery[13]. These values were incorporated in the decision tree.

To calculate the transitions between the health states in the Markov model, progression of disease was evaluated in the retrospective control group. Based on literature, we assumed that progression of disease was affected by the resection margin status[5,21] and that death due to colorectal cancer (CRC) was affected by progression status. Information on progression of disease stratified by margin status and mortality data stratified by progression status were retrieved from medical records. Progression of disease was defined as “local recurrence or distant metastasis after surgery”, as the sample size was too small to stratify for local and distant recurrence. Among these patients, some had limited metastatic disease prior to surgery (e.g. liver metastasis). To prevent overestimating the risk of progression in the whole population, these patients were incorporated in the progression of disease state in the first cycle after surgery.

Table 1 - Input parameters for the decision tree and the Markov model.

The observed number of patients	**LARC**		**LRRC**			**Source**
Having a negative surgical margin (R0)						
After navigated surgery	13 (n=14)		15 (n=19)			[A]
After standard surgery	85 (n=101)		20 (n=41)			[B]
Having progression per year						
after R0 1st year	29 (n=85)		9 (n=20)†			[B]
after R0 2nd (for LRRC: and 3rd) year	9 (n=85)		4 (n=20)†			[B]
after R0 3rd year*	4*(n=85)		-			[B]
after R1 1st year	11 (n=16)		11 (n=20)			[B]
after R1 2nd and 3rd year	1 (n=16)		3 (n=20)			[B]
Died due to CRC after progression; over a timeframe of 3 years						
after progression in the 1st year	25 (n=41)‡		13 (n=20)			[B]
after progression in the 2nd and 3rd year	2 (n=13)‡		3 (n=7)			[B]
Parameters used in the decision model	**LARC**		**LRRC**		**Distribution**	**Source**
	Mean	SE	Mean	SE		
Negative surgical margin rate (R0)						
Navigated surgery	0.929	0.067	0.789	0.091	Beta	[A]
Standard surgery	0.842	0.036	0.488	0.077	Beta	[B]
Transition probability for Disease-free to progression after a negative surgical margin (R0)						
from DF to PD in the 1st year	0.103	0.033	0.159	0.080	Beta	[B]
from DF to PD in the 2nd year	0.047	0.023	0.100	0.066	Beta	[B]
from DF to PD in the 3rd year	0.013	0.012	0.100#	0.066	Beta	[B]
Transition probability for Disease-free to progression after a positive surgical margin (R1)						
from DF to PD in the 1st year	0.252	0.105	0.252	0.093	Beta	[B]
from DF to PD in the 2nd year	0.028	0.040	0.159	0.078	Beta	[B]
from DF to PD in the 3rd year	0.028#	0.040	0.159#	0.078	Beta	[B]
Transition probability for Progressive Disease to Death						
from PD in the 1st year to Death	0.090	0.067	0.135	0.075	Beta	[B]
from PD in the 2nd and 3rd year to Death	0.030	0.036	0.089	0.101	Beta	[B]
Transition probability for Disease-free to Death	0.003	-	0.004	-	-	[25] background mortality
State costs	**LARC**		**LRRC**			
Disease-free	€492	€63	€492	€63	Gamma	Expert
Transition from DF to PD	€14,883	€1,898	€13,107	€1,672	Gamma	Expert
Progressive disease	€585	€75	€585	€75	Gamma	Expert
Intervention costs	**Combined LARC & LRRC**			**SE**	**Distribution**	**Source**
Surgery	€10,970			€1,399	Gamma	[27]
Addition of navigation	€3,388			€432	Gamma	Expert; Box 1
Developing 3D model and preoperative CT-scan	€269					[29,30,46]

(continued on next page)

Table 1 (continued) - Input parameters for the decision tree and the Markov model.

Intervention costs (addition of navigation)	**Combined LARC & LRRC**	**SE**	**Distrub-tion**	**Source**
Additional personnel during OR	€197			[29,46]
Navigation system	€2,745			List prices; expert
Overhead	€177			[29]
Utilities	**Combined LARC & LRRC**	**SE**	**Distribu-tion**	**Source**
First cycle	0.70 (n=63)	0.029	Beta	[C] (1mo survey)
Disease free (subsequent cycles)	0.85 (n=44)	0.022	Beta	[C] (6mo survey)
Progressive disease (subsequent cycles)	0.77 (n=14)	0.050	Beta	[C] (6mo survey)
Additional parameters used in the scenario analysis				
Scenario 1: Including the costs of constructing a hybrid OR when no hybrid OR is available in the hospital				
Additional costs to use a hybrid OR and the C-arm CBCT in hybrid OR	€2,975			[47]; Supple-ment 4
Total costs of the addition of navigation	€6,363	€812	Gamma	
Scenario 2: Using the navigation system for 50% instead of 12%				
Navigation system	€1,027			Supple-ment 4
Addition of navigation costs per patient	€1,670	€213	Gamma	Expert; Box 1; supplement 4

Abbreviations: SE = standard error, DF = disease free, PD = progression of disease, CRC = colorectal cancer.
[A] Prospective data collection within the navigated group at the NKI-AVL[13]; [B] Retrospective data collection within the control group at the NKI-AVL[13]; [C] Prospective observational cohort study.
* Only in the LARC group, among patients showing a negative surgical margin we enough events were found in both the 2nd and 3rd year to calculate probabilities for both years. In the other groups, we found limited events and decided to calculate a combined probability for the 2nd and 3rd year.
# shows the transitions that were similar for the 2nd and 3rd years. This probability was based on the sum of events occurring in the 2nd and 3rd years.
† 1 of the LRRC patients received two surgeries and were both included in the analysis by Kok et al. For evaluating progression of disease this does not make sense, therefore this patient was excluded. Therefore the sum is 40 instead of 41.
‡ The total number of patients having progression in the 1st, and 2nd and 3rd year is different from the number presented between brackets in the lines for died due to progression. After R1 in the 1st year, all 12 events occurred in the 1st year and none in the 2nd and 3rd year. To incorporate uncertainty surrounding the chance on having progression in the 2nd and 3rd year we moved 1 event to the second year to calculate the transitions from disease-free to progression. Therefore, the number of patients progressed in the row for patients died due to progression shows one person more for the 1st year, and one person less for the 2nd and 3rd year.

Clinical effectiveness (continued)

The probabilities to experience events (progression or death) per 3 months were calculated linearly using the number of events and the total number of patients at risk with the following formula: 1-exp(-r*t). Where 'r' stands for the rate per 3 months calculated by -(ln (1-observed chance)/time of the observation), and 't' stands for the time[22]. To incorporate time or disease history in the model, two tunnel states were incorporated in the model: "1st-year progression of disease" and "2nd or 3rd-year progression of disease"[23,24]. The risk of dying due to progression within 3 years was evaluated separately for patients having progression in the first year and separately for patients having progression in the second and third year. The second and third year were combined because of limited number of cases. Mortality due to all causes was based on data from the Dutch central bureau for statistics, mirroring the average age of the two patient populations (LARC: 60 years and LRRC: 65 years)[13,25]. All observed events and transition probabilities incorporated in the Markov model are listed in Table 1.

Health-related quality of life

Utilities were measured using the EQ5D-5L[26] to incorporate quality adjusted life years (QALYs) in the analysis. Utilities are values between 0 and 1 where a higher value indicates a better health status. These values were based on measurements in the ongoing prospective cohort study (Supplement 3). Patients completed questionnaires before surgery and after 1, 3, 6, 12, and 24 months. The utility value for the first cycle was based on the first-month questionnaire, incorporating the impact of the surgery itself. The utility value of the subsequent cycles (progression of disease or disease-free) was based on the questionnaire completed at 6 months after surgery stratified for the health status at 6 months. Due to the limited number of observations per indication (LARC and LRRC), we did not stratify for LARC or LRRC, assuming similar QoL when having progression or being disease-free (Table 1). Furthermore, we did not stratify for navigation or standard surgery as we hypothesize that the effect of using navigation is found in the number of patients showing progression of disease as a consequence of a higher positive resection margin rate.

Intervention costs

For the costs of surgery, the formally average registered tariff (DRG) for open and laparoscopic low anterior resection (LAR) and abdominoperineal resection (APR) in the Netherlands were used. Open and laparoscopic APR and LRP showed the same tarrif[27]. The additional costs for using the image-guided navigation system were estimated using a bottom-up costing methodology, taking into account additional activities, additional required time, and personnel (Box 1). We assumed implementing navigation in an already existing hybrid OR and, for the base case, exclusive use of the navigation system for this indication (12%). This resulted in an additional cost of €3,388 per patient. Details on the calculation are described

in Supplement 5.

State and transition costs

The health state costs and transition costs were based on the care delivered per state and transition. Care consumed per health state was based on the Dutch guideline on follow-up care for colorectal cancer[28]. Expert elicitation was used to estimate a weighted average of care used in case of an event (local recurrence, distant metastasis) for both LARC and LRRC, such as radiotherapy and chemoradiation. To calculate the transition costs, the identified consumed care was linked to tariffs for DRGs, health activities, and medications[29–31] (Table 1). Details on these costs are listed in Supplement 6.

Model analysis and probabilistic sensitivity analysis

The models were programmed in Microsoft Excel 2010. Costs were discounted at a rate of 4% and effects at a rate of 1.5% according to Dutch guidelines[19]. The primary outcome of the models was the incremental cost-effectiveness ratio (ICER) which was calculated by dividing the incremental costs by the incremental QALYs. The involved experts of the NKI-AVL (TR, EK, GB) collaborated to validate the model, input parameters, and assumptions. Because this analysis evaluates an innovation early in its development process, the input parameters are subject to uncertainty. This uncertainty was evaluated using a probabilistic sensitivity analysis. Table 1 shows the distributions surrounding the parameter values used in the Monte Carlo simulations (2000 random samples) for this analysis. The results of the probabilistic analysis are shown in a cost-effectiveness (CE-)plane. Furthermore, cost-effectiveness acceptability curves (CEAC) were generated, indicating the probability that an intervention is cost-effective, given a certain Willingness To Pay (WTP). The informal WTP ratio for diseases with a high symptom burden is €80,000 per QALY[32] in the Netherlands.

Sensitivity analyses

A deterministic one-way sensitivity analysis was performed, evaluating the influence of the uncertainty surrounding each of the input parameters. All parameters were varied over their upper and lower limits. The outcomes were plotted in a tornado diagram. Besides, two scenarios were evaluated: 1) Inclusion of construction costs for a hybrid OR to use the navigation system (in case a hospital does not have this yet), 2) Utilization of the navigation system was set at 50%, as it is assumed that the system is valuable in other indications as well. The input parameters for these scenarios are incorporated in Table 1 and detailed information is listed in Supplement 7. Finally, since the costs of the navigation system are still uncertain, a threshold analysis was performed assuming a WTP of €80,000 per QALY to identify the maximum incremental costs per patient[33].

Value of Information Analysis

The expected value of perfect information (EVPI), indicating the maximum amount a decision-maker would be willing to spend on further research to obtain perfect information, was estimated[34]. These analyses are used to support decisions on further research in early stages of technology development. The EVPI was calculated by taking the difference between the expected net monetary benefit- obtained under perfect information- and the expected net monetary benefit obtained based on the current data. To evaluate the EVPI for the beneficial population, we used the incidence numbers of LARC (n=1,384) and LRRC (n=250) based on the Dutch situation[35,36].

RESULTS

Base case results

For LARC, we found 2.50 total life years (LYs) after standard surgery versus 2.53 LYs for navigated surgery. Total QALYs were 2.02 for standard and 2.05 for navigated surgery. Total utility gain after 3 years was 0.0004 for navigated surgery. Total costs for standard surgery were €23,238 compared to €26,379 for navigated surgery, resulting in an ICER of €136,604/QALY for LARC.
For LRRC, we found 2.11 LYs after standard surgery compared to 2.17 LYs after navigated surgery. Total QALYs were 1.67 and 1.73 for standard and navigated surgery, respectively. Total utility gain after 3 years was 0.0034 for navigated surgery per patient. Total costs of standard surgery were €25,862 and €28,719 for navigated surgery, resulting in an ICER of €51,802 per QALY gained (Table 2A).

Probabilistic sensitivity analysis

The CE-plane in Figure 1 shows that most observations for LARC (84%) indicated that navigated surgery resulted in better outcomes at higher costs. The cost-effectiveness acceptability curve (CEAC) shows that standard surgery in LARC has the highest probability of being cost-effective (78%) at a WTP of €80,000.

For LRRC, also most of the observations (79%) indicated improved outcomes at higher costs. The CE-plane shows more uncertainty compared to LARC which corresponds to the higher amount of uncertainty surrounding the data due to the smaller sample size. At a WTP of €80,000, navigated surgery has a probability of 52% to be cost-effective compared to standard surgery for LRRC.

Table 2 - A. Deterministic outcomes of the cost-utility analysis on navigated surgery compared to standard surgery. B. outcomes of the scenario analysis.

A. Base case results

	Treatment costs	QALYs	LYs	iCosts	iQALYs	ICER	Conclusion
Base case results LARC							
Navigated surgery	€26,379	2.05	2.53				
Standard surgery	€23,238	2.02	2.50				
				€3,141	0.02	€136,604	Navigated surgery is dominated at a threshold of €80.000/QALY (WTP)
Base case results LRRC							
Navigated surgery	€28,060	1.73	2.17				
Standard surgery	€25,164	1.67	2.11				
				€2896	0.06	€52,510	Navigated surgery dominates at a threshold of €80.000/QALY (WTP)

B. Results from the scenario analysis

Intervention	Intervention	ICER scenario	Conclusion scenario
Scenario 1: A hybrid OR has to be constructed before the navigation system can be used	LARC	€266,019	Dominated
	LRRC	€106,458	Dominated
Scenario 2: Increase in utilization of the navigation system to 50%	LARC	€61,817	Cost-effective at WTP threshold
	LRRC	€21,334	Cost-effective at WTP threshold
Combination of 1 and 2: increased use of the navigation system and including the costs of constructing a hybrid OR to use the navigation system#	LARC	€191,232	Dominated
	LRRC	€75,282	Cost-effective at WTP threshold

The WTP threshold used was €80,000. "Dominated" means that standard surgery is preferred in terms of cost-effectiveness compared to navigated surgery. Cost-effective at WTP threshold means that that navigated surgery is preferred in terms of cost-effectiveness over standard surgery at the WTP threshold of €80,000.

Abbreviations: QALYs = quality of life years, LYs = life years, iCosts = incremental costs, iQALYs = incremental quality of life years, ICER = incremental cost-effectiveness ratio; WTP = willingness to pay threshold.

= total costs for the use of navigation including the hybrid OR costs assuming a use of 50% = €4,644.28.

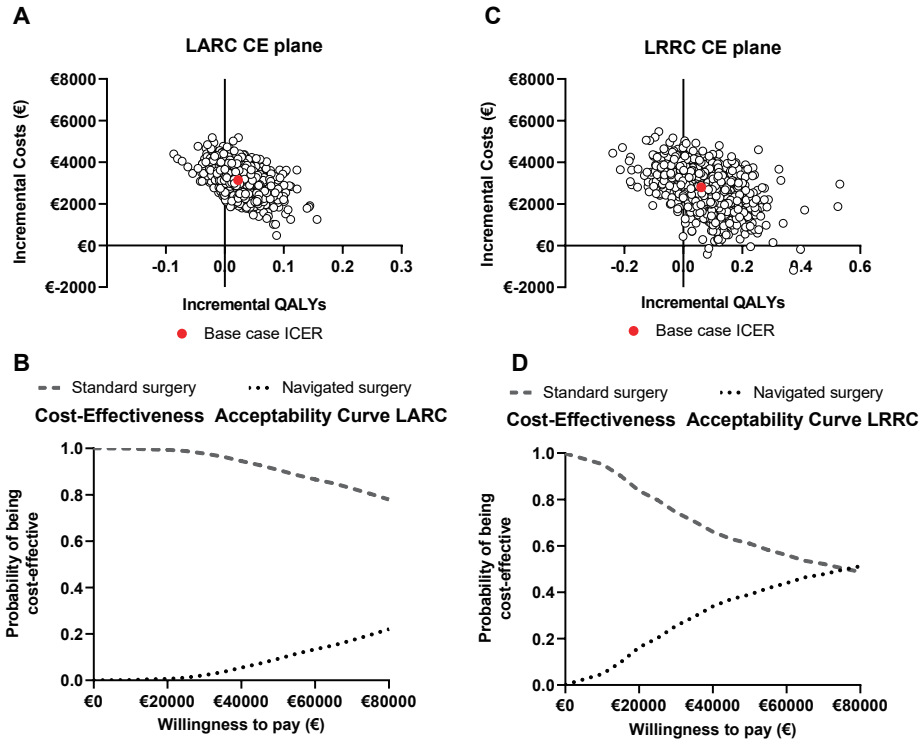


Figure 1. A and C show cost-effectiveness planes for LARC (A) and LRRC (C) showing the incremental quality adjusted life years (QALYs) per incremental costs for navigated surgery versus standard surgery. The scatterplots show the mean differences in costs and outcomes from the data using 2,000 random samples. In both indications, most of the observations are in the north-east quadrant which indicates improved outcomes at higher costs. B and D show cost-effectiveness acceptability curves for LARC (B) and LRRC (D) presenting the probability of the cost-effectiveness of navigated surgery and standard surgery for a range of willingness to pay thresholds.

Sensitivity analyses

Figure 2 shows that the results are mostly influenced by the uncertainty surrounding the transition probabilities for the first year, the surgical margin rate, and the costs of the navigation system in both groups. For example, when the maximum value for the transition from disease-free to progression after an R0 resection was used, showing similar or even worse progression than after R1 resection, LRRC and LARC show ICERs around €200,000. Contrary, when the maximum value from disease-free to progression after an R1 resection was used, the ICERs for LARC and LRRC decrease substantially. For LRRC, this resulted in a QALY difference of 0.11 compared to 0.06 in the base case.

Table 2B presents the results of the scenario analysis. When a hospital has to construct a hybrid OR before navigation can be used, navigated surgery is not cost-effective in LARC or LRRC (scenario 1). Increasing the utilization of the navigation system (scenario 2) results in

navigated surgery being cost-effective at a WTP threshold of €80,000 for LARC. For LRRC, navigated surgery is in this scenario cost-effective at most commonly used WTP thresholds. Since this is a realistic scenario, supplement 8 shows the probabilistic results for this scenario. Figure 3 shows the effect of various utilization ratios on the ICER for scenario 2 and the combination of Scenario 1 and 2.

Based on the threshold analysis, we found that the navigation system may have a maximum cost per patient of €1,839 in LARC and €4,412 in LRRC.

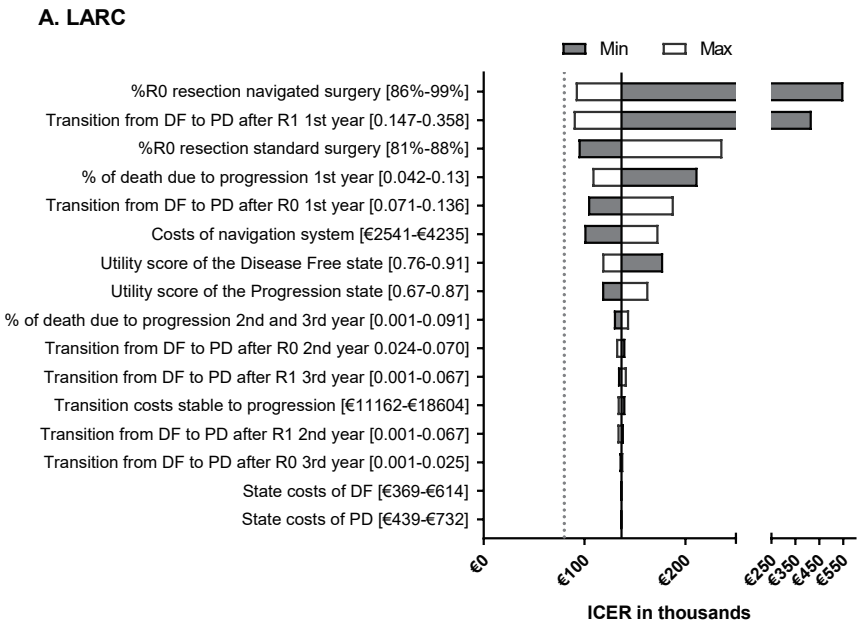


Figure 2 (continued on next page). Tornado diagram showing the results of the one-way sensitivity analysis. Figure A shows the results for the LARC group with a deterministic ICER of €136,604. A dotted line is placed at the willingness to pay threshold of €80,000 which is used in the Netherlands. **Abbreviations:** DF = disease free, PD = progression of disease, R0 = radical resection, R1 = a positive surgical margin.

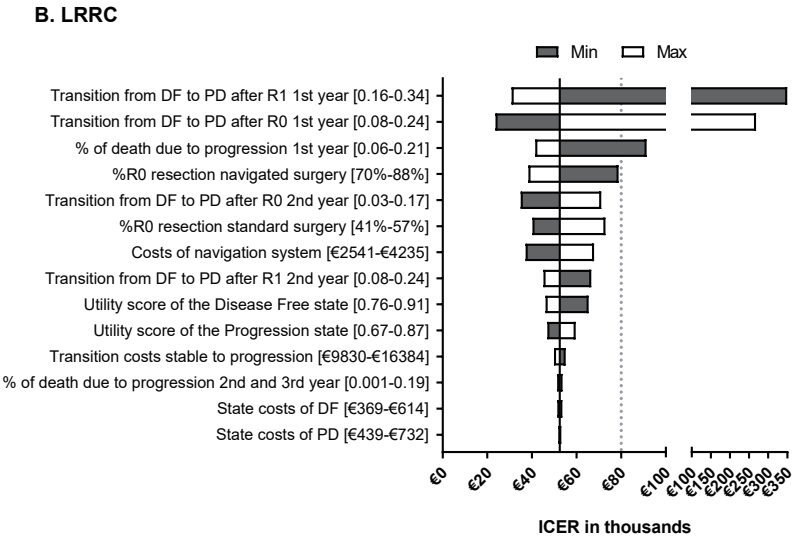


Figure 2 (continued). Tornado diagram showing the results of the one-way sensitivity analysis. B shows the results for the LRRC group with a deterministic ICER of €52,510. The scales of both figures are different and the gap on x-axis shows that a different scale is used after the gap. A dotted line is placed at the willingness to pay threshold of €80,000 which is used in the Netherlands. **Abbreviations:** DF = disease free, PD = progression of disease, R0 = radical resection, R1 = a positive surgical margin.

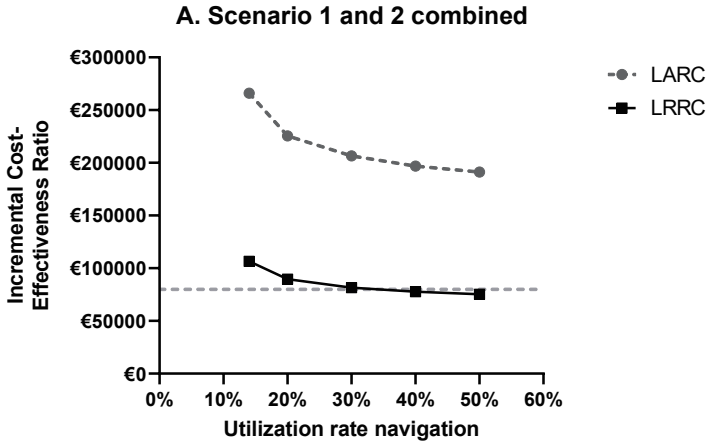


Figure 3 (continued on next page). Graphical illustration of the scenario analysis. Shows the impact of varying the utilization rate of the navigation technology on the ICER. 3A. shows the ICER for multiple utilization rates of navigation for the combination of scenario 1 and 2. Scenario 1 includes the construction costs for a hybrid OR when a hospital does not have this yet. In Scenario 2 the navigation system was used for 50% (in this figure thus varied over 12%- 50%).

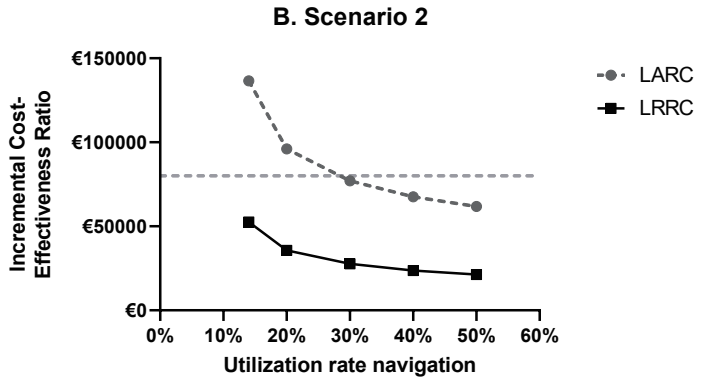


Figure 3 (continued). Graphical illustration of the scenario analysis. Shows the impact of varying the utilization rate of the navigation technology on the ICER. 3B. shows the ICER for multiple utilization rates of navigation of Scenario 2.

Value of information analysis

The EVPI was almost €3.7M in LRRC, which indicates the maximum required investment in further research to obtain perfect information on the cost-effectiveness of navigated surgery compared to standard surgery (Supplement 9). Since standard surgery was preferred at a WTP threshold of €80,000 in LARC, estimating the EVPI was not considered relevant for LARC.

DISCUSSION

Our results indicate that navigated surgery is expected to be cost-effective, especially in LRRC (ICER: €52,510), and has the potential to become cost-effective for LARC when the navigation system shows to be effective in other indications which could increase its use. These results are in line with the promising clinical results of Kok et al[13].

Based on Figure 2, a strong relationship between a R0 resection and a reduced risk of recurrence seems crucial for navigated surgery to become cost-effective. Although we concluded that navigation is cost-effective in LRRC, our data for LRRC showed no clear relationship between an R0 resection and a reduced risk of progression, e.g. reflected by the limited QALY gain found. Based on the significantly higher chance to achieve an R0 resection (79% vs 49% (p=0.047))[13] we expected a larger QALY difference. In a best-case situation, having a lower risk of progression with an R0 resection, the ICER could drop to €23,648 (Figure 2). Based on the current evidence base[17,37,38], it could be expected that a stronger relation between R0 and reduced risk on progression is found when the analysis is based on a larger dataset, and progression of disease is stratified in local recurrence and distant metastasis. This would

result in a higher chance for navigated surgery to become cost-effective. It should, however, be noted that resection margin status is also influenced by tumor biology.

Although a strong relationship between an R0 resection and a reduced risk of progression was found in LARC, navigated surgery was not cost-effective in the base case analysis, since the difference in having an R0 resection between standard and navigated surgery was small[13]. Identification of clinical subtypes that would benefit from navigated surgery would be of interest to become cost-effective in LARC.

The navigation system costs seem another crucial aspect. One could consider pricing the navigation system related to its cost-effectiveness, as in value-based pricing. The threshold analysis showed that the maximum per patient costs may be €1,839 in LARC and €4,412 in LRRC. When the navigation system is used in multiple indications, the value base in each of these indications should be taken into account.

A final important finding is that, when investing in a navigation system, users should aim for optimal capacity use. For instance by organizing centralization of care or identification of multiple indications where navigation could be of added value (Table 2). This is especially the case when a hospital has to invest in a hybrid OR. Currently, the use of navigation is piloted in multiple oncologic indications[12,39,40], which could facilitate the future adoption of navigated surgery.

This study presents results from the first cost-effectiveness analysis for navigated surgery based on the first clinical data available. The results could inform its further development and the start of subsequent clinical or pilot studies. Further strengths include, (i) the inclusion of tunnel states to incorporate time-to-event information, and (ii) the utility values that were based on prospective data from a relatively large (n=63) and similar patient group, that showed utility values that were in line with literature[41].

This study has several limitations that should be acknowledged. In general, early cost-effectiveness analyses are associated with uncertainty in the input parameters, for example, because of small sample sizes and suboptimal study designs. Therefore, the outcomes could be debatable. This is also shown by the EVPI of 3.7M. More specifically, our analysis is limited because we did not incorporate patient characteristics (e.g. tumor stage, tumor location) since treatment history could affect margin status and progression of disease[42]. Besides, by not stratifying for local recurrence and distant metastasis, evaluation of the relationship between R0 and a reduced risk of progression was challenging, because achieving an R0 resection has a limited influence on reducing the risk of metastases. Related to the utility values, the utility-scores for LARC and LRRC were assumed to be equal, although LRRC

patients are expected to receive multiple chemotherapy lines which is expected to result in lower utility values[43]. Another limitation is that we used the 6-month questionnaire to base our utility value on for the disease-free and progression state, which is likely to overestimate the utility-score since patients experience more complaints as the disease progresses. As the utility scores show a large impact on the cost-effectiveness results, utility values for 1 and 2 years after surgery should be incorporated. Furthermore, as curative surgery (R0 resection) is expected to decrease pain complaints, we expect to underestimate the QoL benefit of navigated surgery leading to rather negative ICER estimates. Another issue is related to the costs used in the analysis. The costs were calculated from a Dutch healthcare perspective while the costs of the navigation system were based on list prices and expert elicitation. Finally, since the navigation system is a new surgical tool, a learning curve may be present which potentially underestimates the performance of navigated surgery.

Although there is a tendency in surgery to not formally evaluate incremental improvements in technology, we recommend comparing navigated and standard surgery prospectively, preferably multi-center, in terms of resection margin rate (R0, R1, and R2), complication rate, QoL and utility values. Based on this data, the cost-effectiveness should be simulated using the large (inter)national studies presenting survival after R0 and R1/R2 margins[16,17,37,38] to inform adoption and reimbursement decisions. Furthermore, we suggest validating the mapping study of Wong et al. when QoL and utilities are measured at several moments in time since the EQ5D-5L seems not sensitive to capture CRC specific complaints[44]. Subsequently, to inform decision-makers with the best available evidence, also on potential unforeseen effects e.g. learning curve, the cost-effectiveness analysis should be updated (iterative approach[45]) when more robust survival data is available.

Conclusion

Navigated surgery is expected to be cost-effective in LRRC patients and has the potential to become cost-effective for LARC patients. Since the navigation system seems to be associated with high costs per patient, it is crucial to, when hospitals invest in such an innovative medical device, use it optimally (centralization of care) and seek other indications where it could be of additional value.

ACKNOWLEDGEMENTS

We want to thank the research group of Theo Ruers and Esther Kok for providing the clinical data for this analysis. Furthermore we want to thank all the patients that participated in the prospective cohort study evaluating quality of life after colorectal surgery.

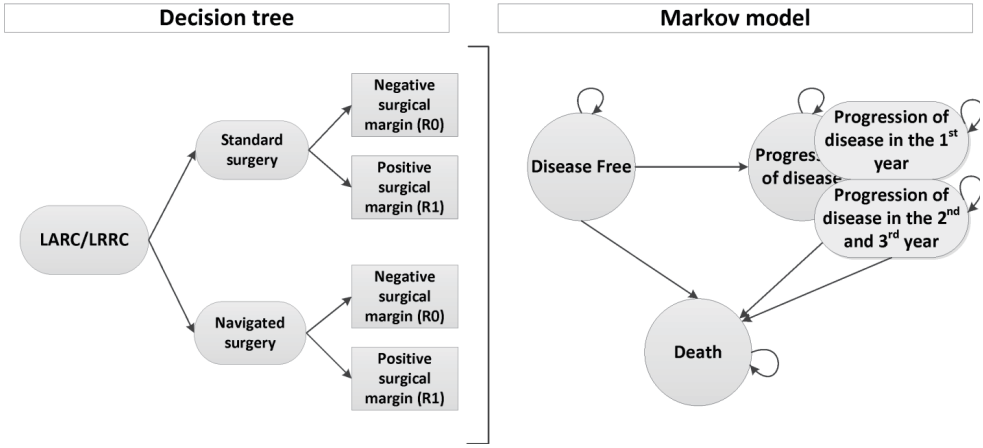
REFERENCES

1. Glynne-Jones, R. *et al.* Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol.* **28**, iv22–iv40 (2017).
2. Nederlands Tijdschrift voor Geneeskunde. *Behandeling van lokaal recidiverend rectumcarcinoom*. *Nederlands TIjdschrift voor Geneeskunde*
3. Integraal Kankercentrum Nederland. Darmkanker in Nederland: cijfers uit de Nederlandse Kankerregistratie. Available at: https://www.iknl.nl/nieuws/2019/darmkanker-in-nederland-cijfers-uit-de-nederlandse. (Accessed: 4th February 2020)
4. Lee, Y.-C., Lee, Y.-L., Chuang, J.-P. & Lee, J.-C. Differences in Survival between Colon and Rectal Cancer from SEER Data. *PLoS One* **8**, e78709 (2013).
5. Bhangu, A., Ali, S. M., Darzi, A., Brown, G. & Tekkis, P. Meta-analysis of survival based on resection margin status following surgery for recurrent rectal cancer. *Color. Dis.* **14**, 1457–1466 (2012).
6. Bonjer, H. J. *et al.* A Randomized Trial of Laparoscopic versus Open Surgery for Rectal Cancer. *N. Engl. J. Med.* **372**, 1324–1332 (2015).
7. Rickles, A. S. *et al.* High Rate of Positive Circumferential Resection Margins Following Rectal Cancer Surgery. *Ann. Surg.* **262**, 891–898 (2015).
8. Yang, H. Y., Park, S. C., Hyun, J. H., Seo, H. K. & Oh, J. H. Outcomes of pelvic exenteration for recurrent or primary locally advanced colorectal cancer. *Ann. Surg. Treat. Res.* **89**, 131 (2015).
9. Palmer, G., Martling, A., Cedermark, B. & Holm, T. A Population-Based Study on the Management and Outcome in Patients with Locally Recurrent Rectal Cancer. *Ann. Surg. Oncol.* **14**, 447–454 (2007).
10. Vermaas, M. *et al.* Total pelvic exenteration for primary locally advanced and locally recurrent rectal cancer. *Eur. J. Surg. Oncol.* **33**, 452–458 (2007).
11. Tejedor, P. & Khan, J. Surgical Trends in the Management of Rectal Cancer. *Clin Oncol* **3**, (2018).
12. Nijkamp, J. *et al.* Prospective study on image-guided navigation surgery for pelvic malignancies. *J. Surg. Oncol.* **119**, 510–517 (2019).
13. Kok, E. N. D. *et al.* Association of image-guided navigation with complete resection in patients with locally advanced primary and recurrent rectal cancer: a nonrandomized trial. *JAMA Netw. Open* **3**, e208522–e208522 (2020).
14. IJzerman, M. J. & Steuten, L. M. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Heal. Econ Heal. Policy* **9**, 331–347 (2011).
15. Vera, R. *et al.* Recommendations for follow-up of colorectal cancer survivors. *Clin. Transl. Oncol.* **21**, 1302–1311 (2019).
16. PelvEx Collaborative. Surgical and Survival Outcomes Following Pelvic Exenteration for Locally Advanced Primary Rectal Cancer. *Ann. Surg.* **269**, 315–321 (2019).
17. Kelly, M. E. *et al.* Factors affecting outcomes following pelvic exenteration for locally recurrent rectal cancer. *Br. J. Surg.* **105**, 650–657 (2018).
18. NCCN. *NCCN Guidelines for patients Colon cancer*. (2018).
19. Zorginstituut Nederland (Dutch institute of healthcare). *Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg*. (2016).
20. Jeong, K. & Cairns, J. Systematic review of health state utility values for economic evaluation of colorectal cancer. *Health Econ. Rev.* **6**, 36 (2016).
21. Renehan, A. G. Techniques and Outcome of Surgery for Locally Advanced and Local Recurrent Rectal Cancer. *Clin. Oncol.* **28**, 103–115 (2016).
22. Briggs, A., Claxton, K. & Sculpher, M. *Decision Modelling for Health Economic Evaluation*. (Oxford university press, 2006).
23. Hawkins, N., Sculpher, M. & Epstein, D. Cost-Effectiveness Analysis of Treatments for Chronic Disease: Using R to Incorporate Time Dependency of Treatment Response. *Med. Decis. Mak.* **25**, 511–519 (2005).
24. Sonnenberg, F. A. & Beck, J. R. Markov Models in Medical Decision Making. *Med. Decis. Mak.* **13**, 322–338 (1993).
25. Centraal Bureau van Statistiek. Prognose periode-levensverwachting; geslacht en leeftijd, 2019.

26. Ramos-Goni, J. M. & Rivero-Arias, O. Eq5d: A command to calculate index values for the EQ-5D quality-of-life instrument. *Stata J.* **11**, 120–125 (2011).
27. Dutch Healthcare Authority (NZa). DBC-zorgproduct 029199032 (DRG open tariff). (2018). Available at: https://www.opendisdata.nl/msz/zorgproduct/029199032. (Accessed: 2nd February 2020)
28. Integraal Kankercentrum Nederland. Colorectaal carcinoom. (2014).
29. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. *Manual for cost research: methods and standard cost prices for economic evaluations in health care*. (2015).
30. Dutch Healthcare Authority (NZa). DBC product finder for tariffs. (2019). Available at: http://dbc-zorgproducten-tarieven.nza.nl. (Accessed: 2nd February 2020)
31. Zorginstituut Nederland (Dutch institute of healthcare). Medicijnkosten (costs of pharmaceuticals). (2019). Available at: http://www.medicijnkosten.nl/. (Accessed: 2nd April 2020)
32. Versteegh, M. M. *et al.* Severity-Adjusted Probability of Being Cost Effective. *Pharmacoeconomics* **37**, 1155–1163 (2019).
33. Girling, A., Lilford, R., Cole, A. & Young, T. Headroom approach to device development: current and future directions. *Int. J. Technol. Assess. Health Care* **31**, 331–338 (2015).
34. Willan, A. R. & Pinto, E. M. The value of information and optimal clinical trial design. *Stat. Med.* **24**, 1791–1806 (2005).
35. Klaver, C. E. L. *et al.* Locally Advanced Colon Cancer: Evaluation of Current Clinical Practice and Treatment Outcomes at the Population Level. *J. Natl. Compr. Cancer Netw.* **15**, 181–190 (2017).
36. IKNL. Oncoline guideline on colorectal cancer. Available at: https://www.oncoline.nl/colorectaalcarcinoom. (Accessed: 20th December 2019)
37. Westberg, K. *et al.* Management and prognosis of locally recurrent rectal cancer – A national population-based study. *Eur. J. Surg. Oncol.* **44**, 100–107 (2018).
38. Harris, C. A. *et al.* The Outcomes and Patterns of Treatment Failure After Surgery for Locally Recurrent Rectal Cancer. *Ann. Surg.* **264**, 323–329 (2016).
39. Ivashchenko, O. *et al.* Intraoperative electromagnetic navigation towards liver tumors. *Int. J. Comput. Assist. Radiol. Surg.* **13**, 1–273 (2018).
40. Janssen, N. *et al.* The use of real-time tumor tracking to perform navigation guided breast conserving surgery: Feasibility of the Calypso systemTM in breast phantoms. *Int. J. Comput. Assist. Radiol. Surg.* **12**, 1–286 (2017).
41. Ness, R. M., Holmes, A. M., Klein, R. & Dittus, R. Utility valuations for outcome states of colorectal cancer. *Am. J. Gastroenterol.* **94**, 1650–1657 (1999).
42. Bhangu, A., Mohammed Ali, S., Brown, G., Nicholls, R. J. & Tekkis, P. Indications and outcome of pelvic exenteration for locally advanced primary and recurrent rectal cancer. *Ann. Surg.* **259**, 315–322 (2014).
43. Mayrbäurl, B. *et al.* Quality of life across chemotherapy lines in patients with advanced colorectal cancer: a prospective single-center observational study. *Support. Care Cancer* **24**, 667–674 (2016).
44. Wong, C. K. H., Lam, C. L. K., Wan, Y. F. & Rowen, D. Predicting SF-6D from the European organization for treatment and research of cancer quality of life questionnaire scores in patients with colorectal cancer. *Value Heal.* (2013). doi:10.1016/j.jval.2012.12.004
45. 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* **24**, 459–464 (2008).
46. Dutch Federation of Academic Medical Centers. *Collective labor agreement 2018-2020 for academic medical centers*. (2018).
47. Lindenberg, M. A. *et al.* Understanding the costs of surgery: a bottom-up cost analysis of both a hybrid and conventional operating room (manuscript has been accepted beginning of July 2020). *Int. J. Heal. Policy Manag.* (2020).

SUPPLEMENTARY MATERIAL

Supplement 1 - Overview model



On the left, the decision tree is visualized in which the margin status after navigated and standard surgery is incorporated. On the right, the Markov model is shown which is used to model the costs and effects after having a negative or positive surgical margin. It also shows the tunnel states used to incorporate time effects on the transition from progression to death due to progression.

Supplement 2 - Patient characteristics of patients included in the study of Kok et al. 2020

	Navigation group (A)	Control group (B)	P-value
Locally advanced primary rectal cancer	**N = 14**	**N = 101**	
Sex			0.351
Male	12 (85.7)	73 (72.3)	
Female	2 (14.3)	28 (27.7)	
Age at start treatment (median, years)	58.0 (35-71)	61.0 (25-82)	0.486
Clinical tumor and nodal stage primary tumor			0.562
T3N0-2 MRF+	7 (50)	62 (61.4)	
T4N0-2 MRF+	7 (50)	39 (38.6)	
Distant metastases			1.00
Present	3 (21.4)	24 (23.8)	
Absent	11 (78.6)	77 (78.6)	
Primary tumors location from anorectal verge			0.461
Low (0-5 cm)	10 (71.4)	52 (51.5)	
Middle (5-10 cm)	3 (21.4)	31 (30.7)	
High (10-15 cm)	1 (7.1)	18 (17.8)	

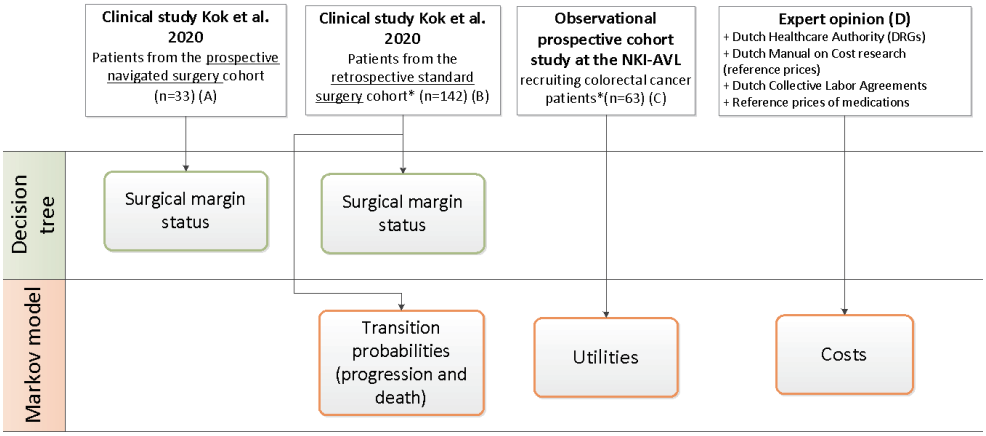
	Navigation group (A)	Control group (B)	P-value
Locally advanced primary rectal cancer	**N = 14**	**N = 101**	
Neoadjuvant treatment primary tumor			0.130
None	0 (0)	1 (1.0)	
Short course radiotherapy (5 x 5 Gy)	0 (0)	3 (3.0)	
Chemoradiation	8 (57.2)	80 (79.2)	
5x5 Gy + chemotherapy	5 (35.7)	16 (15.8)	
Chemoradiation + chemotherapy	1 (7.1)	1 (1.0)	
Type of surgery			0.103
Open APR	2 (14.3)	38 (37.6)	
Lap. APR	0 (0)	3 (3.0)	
Open LAR	12 (78.6)	35 (34.7)	
Lap. LAR	1 (7.1)	10 (9.9)	
Exenteration	0 (0)	15 (14.9)	
Pathological outcome			0.390
ypT0N0	2 (14.3)	5 (5.0)	
ypT2N0-2	2 (14.3)	8 (7.9)	
ypT3N0	2 (14.3)	35 (34.7)	
ypT3N1	3 (21.4)	14 (13.9)	
ypT3N2	4 (28.6)	28 (27.7)	
ypT3N04	0 (0)	7 (6.9)	
ypT4N1	1 (7.1)	1 (3.0)	
ypT4N2	0 (0)	1 (1.0)	
Recurrent rectal cancer	**N = 19**	**N = 41**	
Sex			0.788
Male	11 (57.9)	22 (53.7)	
Female	8 (42.1)	19 (46.3)	
Age at start treatment (median, years)	61.5 (52 – 78)	67.0 (41 – 82)	0.079
Recurrent tumor location			0.560
Pelvic wall / presacral	14 (73.7)	26 (63.9)	
Staple line recurrence	5 (26.3)	15 (36.6)	
Neoadjuvant treatment recurrent rectal cancer			0.061
None	0 (0)	5 (12.2)	
Short course radiotherapy	0 (0)	2 (4.9)	
Chemotherapy	1 (5.3)	2 (4.9)	
Chemoradiation	11 (57.9)	29 (70.7)	
Chemoradiation + chemotherapy	7 (36.8)	3 (7.3)	
Type of surgery			0.100
Open APR	3 (15.8)	19 (46.3)	
Open LAR	4 (21.1)	6 (14.6)	
Exenteration	8 (42.1)	14 (34.2)	
Local resection	4 (20.0)	2 (4.9)	

Abbreviations: APR = Abdominal Perineal Resection, LAR = Lower Anterior Resection. Similar to Table 1 of Kok et al. 2020 in JAMA network open[13].

Supplement 3 - Characteristics of patients included in the Quality of Life input

	Respondents completed the first-month questionnaire	Respondents completed the 6-month questionnaire	
		Disease-free at 6 months	Progressive disease at 6 months
	N = 63	N = 44	N =14
Sex			
Male	47 (74.6%)	28 (63.6%)	10 (71.4%)
Female	16 (25.4%)	16 (36.4%)	4 (28.6%)
Age at start treatment (median, range) (years)	62.0 (35-82)	64.50 (47-79)	67.50 (35-82)
Clinical tumor and nodal stage primary tumor			
Recurrence	18 (28.6%)	12 (27.3%)	5 (35.7%)
T2N0-2	6 (9.5%)	6 (13.6%)	0
T3N0-2	30 (47.6%)	21 (47.7%)	5 (34.7%)
T4N0-2	9 (14.3%)	5 (11.4%)	4 (28.6%)
Type of tumor			
Primary rectal cancer (LARC)	48 (76.2%)	32 (72.7%)	9 (64.3%)
Recurrent rectal cancer (LRRC)	18 (28.6%)	12 (27.3%)	5 (35.7%)
Distant metastases			
Present	15 (23.8%)	4 (9.1%)	8 (57.1%)
Absent	48 (76.2%)	40 (90.9%)	6 (42.9%)
Primary tumors location from anorectal verge			
Low (0- 5 cm)	38 (60.3%)	26 (59.1%)	5 (35.7%)
Middle (5- 10 cm)	22 (34.9%)	16 (36.4%)	7 (50.0%)
High (10- 15 cm)	3 (4.8%)	2 (4.5%)	2 (14.3%)
Neoadjuvant treatment primary tumor			
None	6 (9.5%)	3 (6.8%)	1 (7.1%)
Short course radiotherapy (5 x 5 Gy)	3 (4.8%)	3 (6.8%)	0
Chemoradiation	36 (57.1%)	31 (70.5%)	4 (28.6%)
5x5 Gy + chemotherapy	11 (17.5%)	3 (6.8%)	6 (42.9%)
Chemoradiation + chemotherapy	6 (9.5%)	4 (9.1%)	2 (14.3%)
Chemotherapy	1 (1.6%)	0	1 (7.1%)
Type of surgery			
Open APR	15 (23.8%)	9 (20.5%)	3 (21.4%)
Lap. APR	13 (20.6%)	10 (22.7%)	1 (7.1%)
Open LAR	14 (22.2%)	8 (18.2%)	5 (35.7%)
Lap. LAR	11 (17.5%)	8 (18.2%)	2 (14.3%)
Exenteration	7 (11.1%)	6 (13.6%)	3 (21.4%)
Local resection	2 (3.2%)	2 (4.5%)	0
Hipec	1 (1.6%)	1 (2.3%)	0
Navigation			
Yes	36 (51.7%)	18 (40.9%)	6 (42.9%)
No	27 (42.9%)	26 (59.1%)	8 (57.1%)

Supplement 4 - Overview of sources for input of the model



Schematic overview of studies used for input of the model; * same inclusion criteria.

Supplement 5 - Details on the additional costs for using the navigation system during surgery

Prior to navigated surgery, a CT scan is performed to build the 3D anatomical model. The intervention and personnel costs for building the 3D anatomical model result in €269[1-3]. For the attendance of a technician during the surgery €197 was charged, calculated by taking the average duration of the procedure (5.73h) multiplied with the annual loaded salary retrieved from Collective Labor Agreements[1,2]. The costs of the navigation system, as still in development are yet unknown. Therefore, an average was based on list prices of comparable systems recently launched or planned to launch for example by Intuitive and Auris Health. Assuming a lifetime of 10 years, annual interest percentage of 4.2%[4], the utilization rate of the navigation system at the NKI-AVL (12%), service costs (10%) and disposable costs (€500), resulted in an expected costs of €2,745 per patient. Overhead (38%) was calculated over the costs of the additional steps and personnel costs (€177)[1]. We did not include the additional preparation time of less than 20 minutes as it is assumed to be balanced out by the reduced duration to find the tumor[5]. In total, the use of the navigation system during surgery resulted in €3,388 per patient (Table 3). In this estimation, we assume that a hospital has a hybrid OR with a Cone-beam CT scanner system.

The references used can be found at the end of this chapter.

Supplement 6 - Details of state costs LARC and LRRC

	Times per year*	Costs per activity	Costs per year	Source
Disease free state				
CEA test	4	€7.54	€30.16	6
face-to-face consultation	2	€168.61	€337.21	4
CT abdomen + CT rectum	2	€375.04	€750.08	6
Consultation over phone	2	€17.69	€35.38	4
MRI-scan	0,5	€316.95	€158.47	6
PET/CT-scan	0,5	€1,069.76	€534.88	6
Colonoscopy	0,5	€240.24	€120.12	6
Follow-up care per year			€1,966.31	
Follow-up care per cycle (3 months)			€491.58	
Progression of disease state				
CEA	4	€7.54	€30.16	6
face-to-face consultation	2	€168.61	€337.21	4
CT abdomen + CT rectum	3	€375.04	€1,125.12	6
Consultation over phone	2	€17.69	€35.38	4
MRI-scan	0.5	€316.95	€158.48	6
PET/CT-scan	0.5	€1,069.76	€534.88	6
Colonoscopy	0.5	€240.24	€120.12	6
Follow-up care per year			€2,341.35	
Follow-up care per cycle (3 months)			€585.34	
Costs of progression	**% of patients**	**Costs per activity**	**Costs * %**	**Source**
Local recurrence (LARC)				
Surgery	50%	€10,040.00	€5,020.00	7
Radiotherapy (3D)	5%	€6,865.00	€343.25	7
Chemoradiotherapy (CRT: 25 x 2 Gy met capecitabine)	90%	€7,592.50	€6,833.25	7-9
Chemotherapy (6 cycles of capox)	20%	€7,590.36	€1,518.07	8,9
palliative chemotherapy	25%	€7,590.36	€1,897.59	9
Total treatment costs			€15,612.16	
Local recurrence (LRRC)				
Surgery	25%	€10,040.00	€2,510.00	7
Radiotherapy (extern)	5%	€6,865.00	€343.25	7
Chemoradiotherapy (same as LARC)	20%	€7,547.50	€1,509.50	7-9
Chemotherapy (same as LARC)	20%	€7,590.36	€1,518.07	8,9
Palliative chemotherapy	80%	€7590.36	€6,072.29	9
Total treatment costs			€11,953.11	
Distant metastasis				
Radiotherapy (extern)	10%	€7,485.00	€748.50	7
Chemoradiotherapy	10%	€8,000.00	€800.00	7-9

Chemotherapy	90%	€6,813.24	€6,131.92	[7-9]
Surgery (liver/lung/lymph)	30%	€15,360.00	€4,608.00	[7]
RFA/MWA	10%	€15,360.00	€1,536.00	[7]
Immunotherapy (nivolumab)	3%	€11,006.58	€330.20	[9]
HIPEC	5%	€10,930.00	€546.50	[7]
Total treatment costs			€14,701.11	
	% of patients with LR	% of patients with DM	Total treatment costs	
Transitions costs of progression LARC	20%	80%	€14,883.32	[B]
Transitions costs of progression LRRC	42%	58%	€13,107.27	[B]

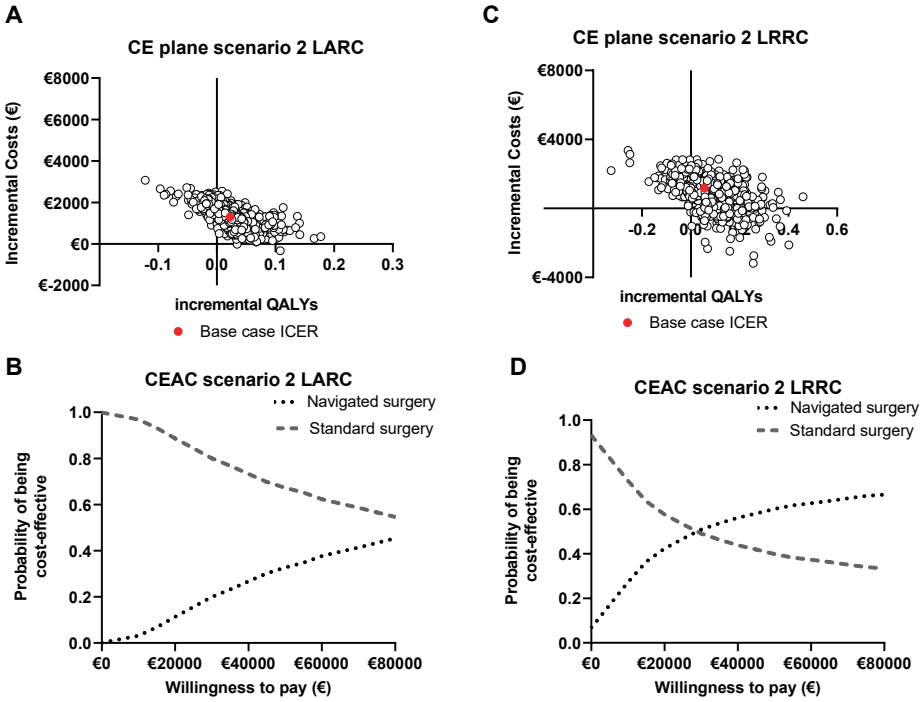
*= for 5 years. The references used can be found at the end of this chapter. **Abreviations:** LR = local recurrence, DM = distant metastasis; [B] Retrospective data collection within the control group at the NKI-AVL[13].

Supplement 7 - Detailed information on the scenario input parameters

For scenario 1, we evaluated the costs when a hybrid OR needs to be constructed and a fixed C-arm CBCT is used. Based on a Dutch study evaluating the additional costs of the hybrid OR compared to a conventional OR, 3.43 euros per minute are accounted for general additional inventory and higher construction costs. The C-arm CBCT adds on average 5.22 euros per minute. For the average duration of the surgery per patient (5,73h (range of 3,08h-10,4h)), the use of the hybrid operating room including a C-arm costs €2,975 based on a recent evaluation of the costs of conventional and hybrid operating rooms[10].

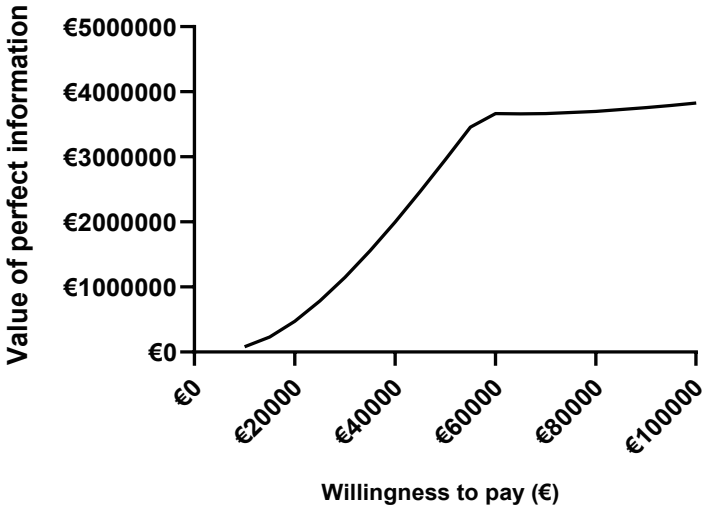
For scenario 2, we evaluated the costs of the navigation system when its use increases from 12% to 50%. This results in a decreased cost of €1,027 to use the navigation system solely instead of €2,745 (Table 1). Resulting in a total cost of €1,670 including the additional steps to use the navigation during a procedure.

Supplement 8 - Probabilistic results for LARC and LRRC when Scenario 2 is present



This figure shows the probabilistic sensitivity analysis results when the navigation system is used more often (Scenario 2). A and C show cost-effectiveness planes for LARC (A) and LRRC (C) with the decreased costs of the navigation system showing the incremental quality adjusted life years (QALYs) per incremental costs for navigated surgery versus standard surgery. The scatterplots show the mean differences in costs and outcomes from the data using 2000 random samples. B and D show cost-effectiveness acceptability curves for LARC (B) and LRRC (D) presenting the probability of the cost-effectiveness of navigated surgery and standard surgery for a range of willingness to pay thresholds.

Supplement 9 - Graphical visualization of Expected Value of Perfect Information



The line presents the expected value of perfect information. The highest value surrounding the willingness to pay threshold of €60,000 is 3.7 milion indicating that 3.7 million is needed to obtain perfect information on whether navigated surgery is cost-effective at a threshold of €60,000.

References used in the supplementary material

1. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. Manual for cost research: methods and standard cost prices for economic evaluations in health care. (2015).
2. Nederlandse Vereniging van Ziekenhuizen (NVZ). Salarisschalen Cao Ziekenhuizen (2018) (Dutch collective labor agreement). (2018).
3. Dutch Healthcare Authority (NZa). DBC product finder for tariffs. http://dbc-zorgproducten-tarieven.nza.nl (2019).
4. Zorginstituut Nederland (Dutch institute of healthcare). Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. https://www.zorginstituutnederland.nl/over-ons/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg (2016).
5. Nijkamp, J. et al. Prospective study on image-guided navigation surgery for pelvic malignancies. J. Surg. Oncol. 119, 510–517 (2019).
6. Nederlandse Zorgautoriteit. NZa zorgproductapplicatie.
7. Nederlandse Zorgautoriteit. Open data van de Nederlandse Zorgautoriteit. Nederlandse Zorgautoriteit [Internet] https://www.opendisdata.nl/ (2019).
8. Zorginstituut Nederland. Farmaceutisch Kompas. https://www.farmacotherapeutischkompas.nl/.
9. Zorginstituut Nederland. Medicijnkosten.nl. https://www.medicijnkosten.nl/ (2019).
10. Lindenberg, M. A. et al. Understanding the costs of surgery: a bottom-up cost analysis of both a hybrid and conventional operating room (manuscript has been accepted beginning of July 2020). Int. J. Heal. Policy Manag. (2020).

PART III

Early HTA: both phase I/II studies


TIL

7

Treatment with tumor-infiltrating lymphocytes in advanced melanoma: Evaluation of early clinical implementation of an advanced therapy medicinal product

Melanie Lindenberg
Valesca Retèl
Joost van den Berg
Marnix Geukes-Foppen
John Haanen
Wim van Harten

Journal of Immunotherapy 2018 41 (9) 413-425

ABSTRACT

Background

Tumor-infiltrating lymphocytes (TIL)-therapy in advanced melanoma is an advanced therapy medicinal product (ATMP) which, despite promising results, has not been implemented widely. In a European setting, TIL-therapy has been in use since 2011 and is currently being evaluated in a randomized controlled trial (RCT). As clinical implementation of ATMPs is challenging, this study aims to evaluate early application of TIL-therapy, through the application of a constructive technology assessment (CTA).

Methods

Literature on ATMP barriers and facilitators in clinical translation was summarized. Subsequently, application of TIL-therapy was evaluated through semi-structured interviews with 26 stakeholders according to six CTA domains: clinical, economic, patient-related, organizational, technical and future. Additionally, treatment costs were estimated.

Results

A number of barriers to clinical translation were identified in literature, including: inadequate financial support, lack of regulatory knowledge, risks in using live tissues, and the complex path to market approval. Innovative reimbursement procedures could particularly facilitate translation. The CTA survey of TIL-therapy acknowledged these barriers, and revealed the following facilitators: the expected effectiveness resulting in institutional support for an internal pilot, the results of which led to the inclusion of TIL-therapy in a national coverage with evidence development program, the availability of an in-house pharmacist, quality assurance expertise and a TIL-skilled technician.

Conclusion

Institutional and national implementation of TIL-therapy remains complex. The promising clinical effectiveness is expected to facilitate adoption of TIL-therapy, especially when validated through an RCT. Innovative and conditional reimbursement procedures, together with organization of knowledge transfer, could support and improve clinical translation of TIL and ATMPs.

INTRODUCTION

Advanced therapy medicinal products (ATMPs) are currently one of the most promising, personalized strategies for cancer treatment[1]. These products are "medicines for human use that are based on genes, tissues or cells"[2]. CAR-T cell treatment for leukemia is an example of such a product. Despite their promising nature, it remains a challenge to implement ATMPs into clinical practice.

In 2007, the European medicines agency (EMA) established a regulation concerning the path to market approval (MA) of ATMPs, namely No. 1394/2007[3]. It mandates that ATMP production requires compliance with the good manufacturing practices (GMP) guideline (2003/94/EC)[4]. This translates into a requirement for a solid quality system, suitable investments, and effective logistical preparation. Partly due to these regulations and necessary preparations, the clinical adoption of ATMPs has been limited[5–8]. This may be explained by the few number of ATMPs (9 of nearly 300 submitted ATMPs) that have achieved MA thus far in Europe[9]. In light of this, ongoing research has sought to identify potential solutions for translation of ATMPs into the clinic. Examples of this include gatekeeping flexibilities e.g. conditional coverage, simplification of ATMP regulations, and simplification of product development[10–13]. Beyond this, the lack of evidence surrounding clinical benefit is likely an important factor hampering the wide clinical adoption of ATMPs.

At the Netherlands Cancer Institute (NKI), TIL-therapy has been offered as an experimental treatment in patients with advanced melanoma since 2011 and is currently being evaluated in an international phase III randomized controlled trial (RCT) (Box 1). In this treatment, TILs residing within tumor material are isolated and expanded *ex vivo* to approximately one billion cells, and are then infused into the patient. Results from phase II studies reveal 1-year and 3-year survival rates following TIL-therapy of 55%–72% and 32%–55% respectively[14,15] and complete responses in 10%–25% of highly advanced melanoma patients who were not responsive to previous treatments[16]. Therefore, the treatment with TILs, which exists already for over two decades[17], appears to be a promising and complementary treatment option for advanced melanoma. Current standard treatment options in advanced melanoma, involve the utilization of antibodies which block checkpoint molecules such as CTLA-4 and PD-1. Despite, the promising results, TIL has not yet been adopted widely[18], which can be explained by a lack of robust clinical evidence.

Health technology assessment (HTA) can play an important role in supporting new technologies from 'bench to bedside'. HTA systematically evaluates social, clinical, economic and ethical & legal aspects of new interventions to inform reimbursement and coverage

decisions[19]. Generally, these methods are introduced in mature technologies that have proved their efficacy and safety. However, when introduced alongside the basic, translational and clinical research process it can steer technical development and even guide implementation by identifying barriers and facilitators while interacting with them. This process is described as "early HTA".[20–23] One of the early HTA methods is constructive technology assessment (CTA), which has its origin in industry to inform technological development before and during introduction of the technology[24,25].

This study aims to evaluate early application of TIL-therapy in the Netherlands by means of a CTA including the six CTA domains - clinical, economic, patient-related, organizational, technical and future in order to identify potential barriers and facilitators. Secondly, this study aims to review recent literature on ATMP implementation to compare findings from the TIL-therapy case to previously identified barriers and facilitators.

METHODS

The methodology in this analysis is two-fold: a literature overview on ATMP barriers and facilitators (2.1) and a CTA on the early application of TIL-therapy in the Netherlands (2.2).

Literature review of ATMP barriers and facilitators

Literature published between 2012 and 2017 was screened for barriers and facilitators in implementing ATMPs into the clinic, using the search terms: "advanced therapy medicinal products" and "implementation or regulation or translation". In addition, "snowballing" was used to identify other relevant articles that had been missed using this search strategy. In this analysis, snowballing entailed: (i) screening of the reference lists of the included articles,(i) using suggestions from journal websites after reading an included article and (iii) screening reference lists in governmental documents (grey literature) related to ATMP implementation. We included barriers and facilitators on the translational pathway from a basic research concept which demonstrated promise for clinical use until MA; hence we left out fundamental barriers commonly related to basic research such as unsuitable mouse models in pre-clinical testing. Supplement 1 describes the search strategy and reasons for in and exclusion.

Constructive Technology Assessment (CTA) framework

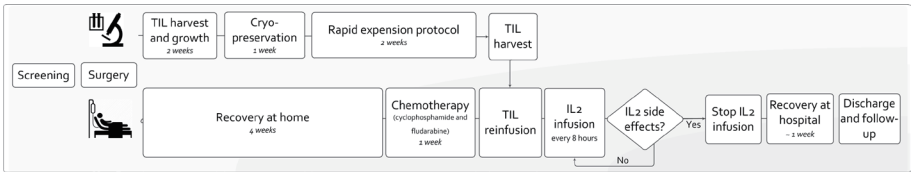
The exact methodology used within CTA depends on the nature of the technology in question, but consists mainly of accepted methods from social sciences and from health services research. In this analysis, a TIL-therapy specific CTA framework was designed comprising the

four domains proposed by Douma et al.: (1) clinical, (2) economic, (3) patient-related, and (4) organizational.[25] Two parameters relevant for the technology were added: (5) technological, referring to the production of TILs, and (6) the future perspective, since the treatment is implemented early in a treatment setting which evolves quickly. Evaluating all these domains enabled identification of a comprehensive range of barriers. (see Table 1)

Box 1 - The setting in which TIL-therapy was applied in the Netherlands Cancer Institute – Antoni van Leeuwenhoek hospital, and the clinical process as followed in the randomized controlled trial.

Since 2011, TIL-therapy has been in use at the Netherlands Cancer Institute – Antoni van Leeuwenhoek hospital (NKI-AvL) and has been included in a coverage with evidence development (CED) program since 2015[26,27]. In this CED program, the cost-effectiveness of TIL-therapy compared to ipilimumab in stage IIIC and IV melanoma is being evaluated, while the treatment is conditionally reimbursed by the government (NCT02278887). This study is conducted in collaboration with the Herlev hospital in Denmark. It was approved by the national medical ethical committee (centrale commissie mensgebonden onderzoek, CCMO) and the Ministry of Health Welfare and Sport, and is the first RCT comparing TIL-therapy to another immunotherapy (ipilimumab). In the Netherlands, fourteen hospitals (Werkgroep Immunotherapie Nederland voor oncologie (WIN-O) centers) are authorized to provide melanoma care in the advanced setting. Since, TIL-therapy is only implemented in the NKI-AvL – one of these 14 hospitals-, the patients treated in one of the other 13 hospitals must be referred by their clinician to the NKI-AvL in order to receive TIL-therapy.

Process of TIL-therapy as provided in the RCT
After screening, surgical resection of lesion(s) is scheduled. Following this, the resected tissue is transported to the production facility where the growing process (approximately five weeks) will begin (Figure 2). In the growing process, the cells are cryopreserved for approximately 1 week to ensure that the patient is recovered from surgery and to streamline logistics, before the growing process continues. When the growth is sufficient, admission for chemotherapy is scheduled, using a regimen of cyclophosphamide (2 days) and fludarabine (5 days). The first day following chemotherapy, TIL infusion will start. On the infusion day, the TIL product is harvested and formulated in the final infusion bag. After quality control, the product is directly released by the qualified person (QP) before administration to the patient. From this moment, one-to-one nursing care starts to monitor and deal with potential severe side effects (e.g. high fever, chills, oliguria, hypotension, weight gain due to fluid accumulation and hypoxia[28]). After four hours, the first bolus IL2 infusion is delivered, the following infusions are given after every eight hours allowing one infusion to be skipped, if this is clinically preferable. Whether a next IL2 dosage should be given depends on the health status of the patient for which the responsible clinician is contacted before every administration (including overnight). If a patient experiences one (or more) treatment-related side effects and this does not decrease before the next administration, IL2 treatment is stopped. One day after discontinuation of IL2, one-on-one nursing care is stopped. On average, patients recover and are discharged 7 days following the last IL2 infusion.



Abbreviations: CED = Coverage with Evidence Development, QP = Qualified Person, NKI-AvL = Netherlands Cancer Institute – Antoni van Leeuwenhoek, IL2 = Interleukin-2.

Semi-structured interviews

Each CTA domain was evaluated by using semi-structured face-to-face interviews with all stakeholders of the TIL-therapy process during the pilot phase (2012) and 1,5 years after the start of the phase III trial (2016). Stakeholders from the hospital included: two clinicians, a nurse-practitioner, four nurses, six patients, the head of the nursing department and one clinical research associate. From the specialized production sites (BioTherapeutics Unit [BTU] and Sanquin) the following stakeholders were included: a head of production facility, two project managers, seven technicians and a labworker involved in leukapheresis. For each interview, a tailored interview protocol was constructed to reflect the areas of interest specific to the role of the stakeholder. The domains and related parameters discussed with each stakeholder are listed in Table 1. In the national implementation phase, interviewees were also asked about: age, experience with TILs (years) and respondents' judgement on their level of innovativeness. Innovativeness was characterized according the theory of Rogers[29].

All interviews were recorded, fully transcribed and labeled using NVivo[30] according to the CTA framework. This labeling step was verified by a second researcher (VR), independently labelling the transcribed interviews of key stakeholders. In cases of discrepancy, labels were discussed (VR, ML). Labelled information was summarized and first discussed with the second researcher (VR) and afterwards discussed with the stakeholders to check whether the most relevant elements were extracted from the interviews.

Clinical domain

Interviews in this domain focused on (i) the clinical process, (ii) trainings required, and (iii) safety priorities. When comparing the interviews during the pilot phase to those in the national implementation phase (during RCT), changes in work routines were identified.

Patient-related domain

There were three measures used in this domain. First, patient interviews were held during the recovery phase of TIL-therapy focusing on waiting times experienced, information provision, treatment experience in each treatment phase, and reasoning behind participation in the TIL trial. Second, the clinical team was asked to explain their experiences related to social and patient impact and their role in providing supportive activities.

Third, a web-based questionnaire was developed to analyze factors influencing TIL trial participation. This was based on (i) aspects described in the accept/decline questionnaire of Penman et al (1984) and the adapted version by Jenkins et al (2013), (ii) the Attitudes on Randomized Trials Questionnaire[31], and (iii) factors described by Kaur et al. (2012)[32–34]. The questionnaire (Supplement 2) was distributed via the Dutch patient association, "stichting

Melanoom", and aimed at advanced melanoma patients. Additionally, during the ongoing RCT, quality of life is measured by means of the QLQ-C15, EQ5D-3L and the Impact of Event Scale, the results are however not included in this research article as the clinical trial is still ongoing.

Table 1 - CTA framework used in the semi-structured interviews.

Domain	Parameter	In interview with stakeholder
General	Age Experience melanoma Experience TIL Innovativeness	All stakeholders excluding patients included in national implementation interviews (12)
Clinical	Clinical process Changed work routine Safety issues Required trainings	Clinical team, clinical research associate, project managers
Patient-related	Patient and social impact Waiting times Treatment experience Information provision Ethical issues Acceptability Adoption	Patients, clinical team, head of nursing department, clinical research associate
Organizational	Implementation Adoption Capacity Communication Logistics and planning Procurement	All stakeholders, excluding patients
Technological	Implementation Quality and safety Technical process Changed work routine Required trainings Logistics	Project managers, technicians, lab technician, and clinicians
Economical	Funding Costs Process Implementation	Project leaders, technicians, clinical team, head of nursing department
Future	Institutional level National level International level Future scenarios	All stakeholders, excluding patients

Describing which domain and parameters were discussed with which stakeholders. In total 26 individuals were included in the interviews to identify faced barriers and facilitators in early application of TIL-therapy.

Organizational domain

This domain focused on the clinical implementation, and includes: logistical alignment, communication, procurement, and planning and capacity.

Technological domain

Interviews within the production facilities evaluated the technical implementation process including required training, and quality and safety regulations. Furthermore, capacity and the production process itself were discussed. When comparing the interviews during the pilot phase to those during national implementation phase, changes in the process were identified.

Economical domain

In this domain, the process of obtaining funding for research and achieving coverage under health insurance schemes were discussed. Furthermore, a bottom-up costing approach was used to estimate the economic burden per patient from screening until the first follow-up appointment. In this approach, resource use is identified per patient resulting in patient-specific unit cost[35]. Therefore, the process of TIL treatment was observed in real-time and verified in the semi-structured interviews during the phase III trial. TIL treatment consists of: (1) screening; (2) isolation of TILs, and (3) hospital admission. For each step, activities were described such as the duration of hospital admission, diagnostic activities, medicine use, laboratory tests, blood products, surgery, and consultations. This information was retrieved from the medical record for the first ten patients treated in the phase III trial, and was linked to unit prices obtained from recent Dutch reference prices to calculate the costs of every process step[36–38]. The production costs were estimated before the start of the RCT by the production facilities.

Future perspectives

In this domain, estimates of the expected uptake, implementation and potential process changes at the institutional, national and international levels were discussed. For various aspects, possible but feasible scenarios for the coming five years were identified. These aspects, such as the degree of effectiveness, the emergence of competing therapies, and the attitudes of clinicians towards the technology, can be used to describe their possible influence on adoption and diffusion[39].

RESULTS

Literature overview: ATMP barriers and facilitators

Of the 65 identified articles, 12 articles were selected[5–7,10–12,40–45]. The two main barriers identified were: (1) inadequate financial support for both the required investments for GMP manufacturing and for setting up first pilot series and clinical trials (described in 8/12 articles) and (2) obtaining the required efficacy results and demonstrating long-term effectiveness data, towards market access (MA) and implementation in clinical practice. This was hampered by for example: a lack of harmonization in the hospital exemption clause, and difficulties in setting-up and receiving approval for clinical trials with ATMPs (8/12). Other barriers described were: compliance with GMP regulation which requires specific standard operating procedures (SOPs), and specific documentation (7/12), potential therapeutic and technical risks in using live tissues as a basis for a treatment strategy (5/12), and a lack of regulatory knowledge to build a full product dossier for obtaining MA (2/12). The main facilitators or suggested solutions were (described in 11/12 articles): using adaptive licensing approaches such as coverage with evidence development (CED) programs, applying risk sharing principles, or the use of accelerated assessment (5/11 articles), the organization of a (national) knowledge platform for information on GMP compliance and route to MA (2/11), securing engagement of HTA organizations alongside ATMP development to estimate the added value of a new ATMP in a certain field (e.g. headroom analysis, cost-effectiveness analysis) (2/11), stimulating harmonization of ATMP and hospital exemption definitions and their procedures across Europe (2/11), and the use of a clinical implementation model in which the trained personnel is responsible for integrating a new therapy into routine clinical practice (2/11). The complete overview of barriers and facilitators, categorized according the CTA domains, is listed in Supplement 3.

CTA on TIL-therapy: Results of semi structured interviews

The barriers and facilitators revealed by the CTA are summarized in Table 3. The following subsection is structured according the six CTA domains.

Characteristics of participants

In total, 26 stakeholders participated in the semi-structured interviews during the pilot study phase and the national implementation phase. We included two medical oncologists, six patients, four nurses, one nurse practitioner, the head of the nursing department and one clinical research associate. From the two production facilities three project managers, seven

technicians and one lab scientist took part. The average age was 47 (range: 32–59) years, having on average 6 (range: 3–9) years of experience with aspects of TIL-therapy, and for the clinically involved stakeholders (7/12) the average experience with melanoma was 18 (range: 5–31) years. The participants showed different levels of innovativeness: five judged themselves as "innovator" / "early adopter", five as "early majority", one as "late majority" and one as a "laggard"[29], supporting the presence of critical respondents.

Clinical domain

A facilitator for the implementation of TIL-therapy in the clinic was the clinical training of a clinician from the NKI-AVL at an expert center (Surgery Branch of the National Cancer Institute). This clinician trained the other oncologists during the pilot study in the NKI-AVL (skills training). Subsequently, the nurses were regularly trained and informed by a clinician or nurse-practitioner on the treatment itself and the treatment effects that could be expected. One of the barriers was the observed high toxicity (e.g. cold shivers, high fever [Box 1]), during the pilot phase after TIL and IL2 infusion. To decrease this toxicity, the clinical process was adapted before entering the RCT phase through the addition of further inclusion criteria: WHO status (≤ 1), a less stringent number of Interleukin-2 (IL2) doses, and providing supportive treatment (e.g. pethidine) at an earlier stage. In addition the apheresis process step, which was used to harvest feeder cells for the production process, was left out as a specific blood product (allogenic buffycoats) could be used instead, showing similar results in the growth of TILs.

These changes and the experience of the clinical team by treating the first 10 patients in the pilot, resulted in an improved acceptance by patients. A decrease in the average length of stay (from 22 to 19 days), and less frequent ICU admissions (a reduction of 40% to 10% in the first 10 patients enrolled in the RCT) were demonstrated. The ability to change the protocol based on experiences from the pilot study was a facilitator for further clinical implementation. From a nursing perspective the adjustment of the number of IL2 infusions was seen as a significant improvement regarding treatment intensity. The adjusted clinical process was felt to be more patient-centered. Clinical results of the current RCT study are expected in 2020 (NCT02278887) which will give more insight into the effect of these adaptations.

Patient-related domain

To anticipate the expected toxicity of TIL-therapy and IL2 infusions (facilitator in implementing TIL-therapy), a psychiatry nurse and social worker were included in the clinical pathway to support patients and their families. Patients reported the complete TIL-therapy process to be physically and mentally burdensome, though acceptable. As a result of the intensified nursing care, closely involved physicians, and the possibility for family members to stay overnight during the TIL and IL2 treatment, patients felt safe. The interviewed nurses considered the

intensity of the treatment to be acceptable, and emphasized the importance of the increased contact with health professionals. This was the case especially during the nights in the period of the TIL reinfusion and IL2 infusion to ensure adequate treatment of potential side effects that may arise. As the severity of the period after the IL2 infusions was underestimated by patients, the clinical team adapted the information provision to better prepare the patients for the potential adverse events that can occur after the IL2 infusions.

Study enrollment of the clinical phase III trial was slow and therefore evaluated as a barrier. The web-based questionnaire analyzing factors related to trial participation was completed by 11 stage IV melanoma patients. Figure 1 shows the results of this questionnaire. General RCT aspects e.g. receiving additional investigations or treatments, travel distance, and switching to another hospital showed limited or no influence in a patient's decision to participate. Expected side effects appeared to impede participation, whereas recommendation by a clinician and the expected improved results demonstrated a positive influence on participation. The statements revealed that the idea of randomization, the probability of not receiving TIL-therapy but ipilimumab, and the fear of becoming more sick over time, have a negative influence on trial participation. Conversely, the idea that participating would be beneficial to other patients and clinicians shows a positive influence on trial participation (Figure 1). The majority of respondents (6/11) were not informed about the TIL-therapy trial. Of the informed patients, one patient participated in the TIL-trial, four patients reported to consider participation but still had other effective standard treatment options, and the last patient suffered from an autoimmune disease and was therefore not eligible for the trial.

Organizational domain

Before the first application of the TIL-therapy, the following logistics were necessary: (i) agreements with surgical planning to align various steps with TIL production; (ii) during TIL and IL2 phase: arrangement of one-to-one nursing care, 24h availability of a trained oncologist (on-site or via telephone); and (iii) preventive ICU bed reservation after the TIL infusion. These arrangements were in the NKI-AvL case not evaluated as a barrier due to being a Comprehensive Cancer Center (CCC) in which research (TIL production) and clinic are already well integrated. The clinic is already well accustomed to adapting clinical processes to research projects.

A barrier that was seen in this domain was the slow adoption of TIL-therapy in the Netherlands, resulting from the low number of referrals to our study. To improve this, the 13 specialized melanoma hospitals in the Netherlands (Box 1) were visited to provide more information on the treatment and on the trial. In addition, a dedicated webpage was created for both physicians and patients, and finally social media platforms from the hospital and patient association were used for promotion. This resulted in twice as many referrals over the

following months (May- November 2017).

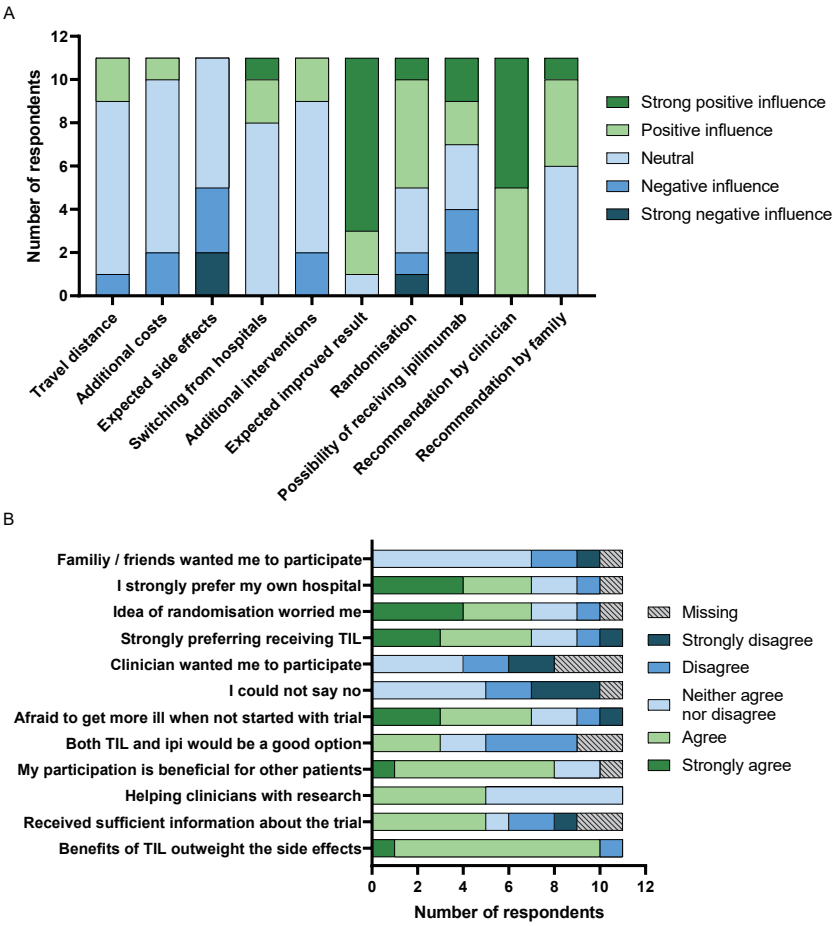


Figure 1. Results of web-based questionnaire. This questionnaire was aimed at patients with advanced melanoma distributed via the patient association (n=11). (A) Shows the level of influence of each aspect (positive or negative) in deciding to take part of the TIL trial. (B) shows the level of agreement with several statements. The majority of questions does not sum up to 11, this is because of missing values.

Technical domain

Figure 2 shows the implementation timeline of TIL-therapy in our institute (NKI-AvL). This paragraph explains some of these processes. Before TILs may be produced, a manufacturing license is required which demonstrates compliance with ATMP and GMP guidelines[3,4]. The process of obtaining approval for this specific product by the Dutch healthcare inspectorate (despite the production facility already holding a GMP permit for other products) took approximately two years. Approval by the Dutch healthcare inspectorate also allows for the acceptance to the entire European market, provided that EU regulations are followed. Two factors in our case facilitated this process: (1) the availability of an in-house pharmacy with

regulatory knowledge for advice; and (2) existing quality management system into which TIL production could be integrated. Examples of additions to the quality system were: generation of product-specific production runs and quality control protocols, general T-cell related SOPs, validation plans and reports, assessing suppliers and their materials for GMP use, and creating the investigators' medicinal product dossier (IMPD) required for MA. Unless these facilitators, the technical preparation was very time consuming and is therefore categorized as a barrier for clinical implementation of TIL-therapy.

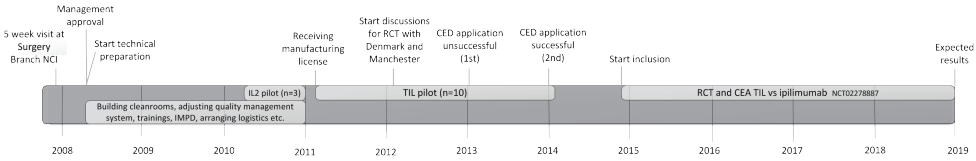


Figure 2. Implementation timeline of TIL-therapy in the Netherlands Cancer Institute – Antoni van Leeuwenhoek hospital. **Abbreviations:** CEA = cost-effectiveness analysis, CED = Coverage with Evidence Development, NCI = National Cancer Institute, IL2 = interleukin-2, IMPD = investigational medicinal product dossier.

Training and acquiring suitable staff for the production process was another barrier as the TIL process deviates from both a research and a standard production setting. A technician familiar with producing TILs (trained in an expert center i.e. National Cancer Institute) was involved in the implementation process, facilitating the training of new employees. The learning curve for TIL production is strongly dependent on the frequency of TIL patients presenting and the frequency of similar research projects. Training of a new technician would still take at least one year to be able to work independently. The challenging nature of the training and staff acquisition is mainly due to compliance with GMP guidelines which require: (i) regular quality assurance checks (e.g. on sterility, growth and viability of the TILs); (ii) creating an auditable process (e.g. all critical process steps performed by two operators) and; (iii) writing a report on the proceedings and potential deviations per TIL product. All the results from testing an individual batch are transcribed in a patient specific Batch Record which is reviewed by the manufacturing department or quality assurance (dependent on local procedures) and qualified person (QP) after which the QP can release the product. Figure 3 shows the TIL growth process schematically, which is described in more detail by Donia et al.[46].

For the alignment of the technical process to the clinical process, a cryopreservation step of approximately 1 week was included to control the start of chemotherapy and, thus, the day of TIL infusion. This is to ensure the availability of a complete medical team when TIL and IL2 side-effects are expected. This required a strict planning and regular communication between the various disciplines. A final barrier identified in this domain is that TIL production is time-consuming, especially on certain production days, such as the initial TIL isolation, initiation of the rapid expansion protocol (REP) and the harvesting day. Any deviation in the growing process will affect the schedule and work routine, potentially resulting in irregular

working days or hours. Although adequate, in a small academic production facility, with only a few trained technicians (which is often the case in such specific ATMPs) this could cause problems related to availability of staff and logistics in the production cycle.

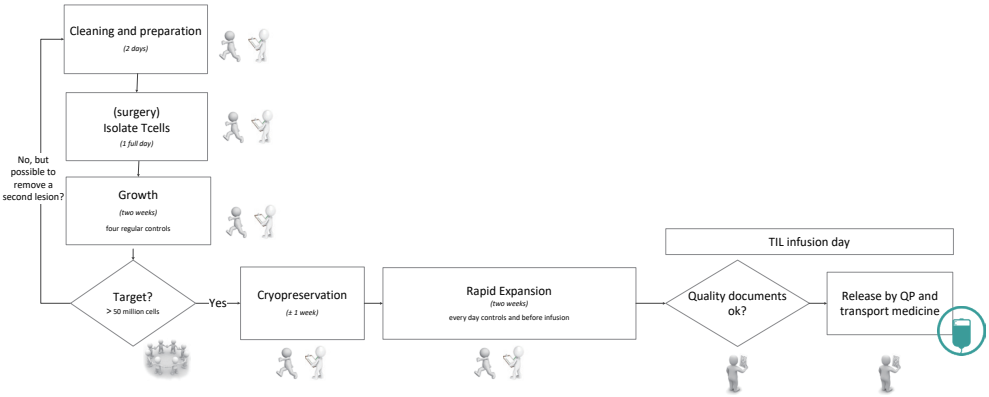


Figure 3. Visualization of technical process of generating TILs in production facility. This visualization includes all quality controls and the duration of steps. The icons reflect the technicians involved, in all process steps: one executes and one monitors. **Abbreviations:** QP = qualified person.

Economical domain

The clinicians reported that gaining financial support was seen as a hurdle by hospitals and laboratories around the world aiming to implement TIL-therapy. In our case, implementation and the pilot study were financially supported by charitable and institutional funding because of the therapy's promising effectiveness results in phase I/II trials. On our pilot results, the national Dutch CED program was granted which enabled a phase III trial and thus facilitated clinical implementation.

Treatment costs

The bottom-up costing approach resulted in a comprehensive insight into the financial impact of TIL-therapy from a hospital perspective. The TIL process was divided into four steps: screening, isolation of TILs, TIL production and admission costs. For all monitored activities, costs, and their respective averages are listed in Table 2. Screening included several physical scans, blood tests and consultations, resulting in a total average cost of €2,837. Isolation of TILs consisted of surgery, admission day(s) and consultations, resulting in a total average cost of €3,665. The TIL production costs consisted of personnel, materials, Quality Control and cleanroom use costs, resulting in a total cost between €35,500 and €50,000, depending on the number of productions per year (10 or 5 patients). These costs were estimated before the start of the CED program, assuming a nonprofit production base and should therefore be interpreted with caution. Admission costs included hospital admission, medication, one-on-one nursing care, blood tests, imaging, consultations and complication-control, which resulted in a total average cost of €27,743. Summing up those steps gives a total estimated

costs per patient between €69,745 and €84,245 depending on the number of patients that should receive TIL-therapy.

Table 2 - Bottom-up costing for the costs of TIL-therapy based on 10 patients treated in CED phase.

Activity	Unit	Unit costs	Quantity (average)	Costs (rounded)	Source
Screening					
Physical examination and tests	PET/CT	€ 695.00	1		[36,38] [B]
	CT chest and abdomen	€ 364.58	2		
	Lung photo	€ 41.24	1		
	MRI brain	€ 233.84	1		
	Ejection fraction	€ 279.47	1		
	Lung function test	€ 61.58	1		
	ECG	€ 46.22	1		
				€ 2,205	
Lab tests	Hematology set	€ 12.13	2		[38] [B]
	Chemical set	€ 47.73	2		
	Microbiologically and oncologic markers set	€ 163.59[α]	1,4[α]		
	Application costs	€ 10.94	2		
				€ 351	
Consultations	Face to face	€ 132	2		[36][B]
	Telephone consult	€ 17	1		
				€ 281	
			Total	€ 2,837	
Isolation of TILs					
Surgery	Surgical removal of tumor	€ 2,134*	1		[A][B]
				€ 2,134*	
Admission days	Nursing day on oncologic department	€ 636	2.2		[36][B]
				€ 1,399	
Consultation	Face to face	€ 132	1		[36][B]
				€ 132	
			Total	€ 3,665	
TIL production					
Situation A: 10 TIL productions per year					
Personnel (2 technicians + 1 production manager)		€ 18,000			[D]
Material and Quality control		€ 10,000			[D]
Cleanroom and equipment		€ 7,500			[D]
				€ 35,500	
Situation B. 5 TIL productions per year					
Personnel (1 technician + 1 production manager)		€ 25,000			[D]
Material and Quality control		€ 10,000			[D]
Cleanroom and equipment		€ 15,000			[D]
				€ 50,000	
		Total range	€ 35,500- € 50,000		

Continued on next page

Table 2 (continued) - Bottom-up costing for the costs of TIL-therapy based on 10 patients treated in CED phase.

Activity	Unit	Unit costs	Quantity (average)	Costs (rounded)	Source
Hospital admission					
Admission	Nursing day on oncologic department	€ 636	13.2		[36][B]
	ICU admission day	€ 2,015	2.2		
				€ 12,828	
Admission with One-on-one nursing	*Day with one on one nursing care on oncologic department*	€ 1,430#	3.8		[36][B]
				€ 5,242	
Medication	Chemotherapy:				
	• Cyclophosphamide	€ 92.14	2 days	€ 184	[37][B]
	• Fludarabine	€ 129.48	5 days	€ 647	
	IL2	€ 355.04	4 times	€ 1,549	
	Filgrastim	€ 123.45	7 times	€ 864	
	Supportive medicines	n.a.	n.a.	€ 36	
	Antibiotics:				
	• Ceftizidim	€ 7.42	15 times	€ 111	
	• Vancomycine	€ 15.93	10 times	€ 159	
				€ 3,553	
Laboratory tests	Hematology set	€ 12.13	16		[38][B]
	Chemical set	€ 47.73	16		
	microbiologically and oncologic markers	€ 113.6Δ	16Δ		
	application costs	€ 12.14	16		
				€ 2,380	
Blood products	Erythrocyte (radiated) application	€ 252.3	2		[38][B][C]
	Thrombocyte (radiated) application	€ 558.2	2		
				€ 1,621	
Consultations	Face to face	€ 132	2		[38][B]
	Telephone consult	€ 17	1		
				€ 281	
Other activities	X thorax	€ 41.24	1		[36,38][A]
	Central venous catheter	€ 980.17	1		[B]
	CT chest and abdomen	€ 729.16	1		
	ECG	€ 46.22	1		
				€ 1,838	
			Total	€ 27,743	
	Total cost range		€ 69,745- € 84,245		

Costs are rounded. α =Costs for microbiologically set is based on the first and most complete laboratory set. Therefore costs here are higher than during admission, the second set contains only 0,4 of the costs of the first set, therefore we describe it as 1,4 times. * = Average based on the different interventions (n=10) as tumors are located differently. # = Based on the reference cost for a oncologic nursing day (€ 636) without the share for the nurse (€ 245) but including 24 hours with an additional 39% for employers costs. Δ = Some microbiologically markers are not measured each of the 16 times, the amount of € 113,60 include all the tests. Therefore 16 times €113,60 will result in an overestimation of the costs. Sources: [A]: NKI-AVL; [B]:Medical Record; [C]:Sanquin; [D] BTU **Abbreviations:** CT = computed tomography, ECG = electrocardiogram, ICU = intensive care unit, IL2 = interleukin-2, PET = positron emission tomography.

Future domain

From a technical perspective, the production process could be simplified in the future for example by means of a more automated process or the use of a closed bioreactor. For the clinical process, it remains unclear whether lymphodepletion and additional IL2 treatment are necessary for its effectiveness. Decreasing the intensity of the non-myeoloablative chemotherapy or changing to a lower dose schedule of IL2 or even removal of these steps could ease clinical adoption (clinical application and clinicians attitude). For national clinical implementation, the stakeholders recommend to first start in specialized melanoma centers due to their experience in treating patients with high dose chemotherapy. Furthermore, other hospitals could receive training from clinicians and nurses at the NKI-AvL. Finally, centralization of the production of TILs seems advisable for the years to come as technical implementation is both highly challenging and time consuming.

Table 3 - Barriers and facilitators in early application of TIL in the Netherlands Cancer Institute.

CTA domain	Barriers (TIL case)	Facilitators / strategies used (TIL case)
1. Clinical	Toxicity of treatment	Being able to change the protocol after pilot phase Gained experience (pilot to phase III research phase) Having experience with high-dose chemotherapy Preventive arrangements (organizational domain)
	Clinical trainings	One clinician trained at Surgery Branch, other clinicians were trained during pilot phase. Nurses received short courses by clinician or nurse practitioner
2. Patient-related	Toxicity of treatment	Availability of a private room large enough for a second bed Involvement of psychiatry nurse and social worker in clinical pathway
	Slow study enrolment Possibility of receiving control arm Randomization No information about trial	Improving information provision for patients: TIL website, information in web-based questionnaire
3. Organizational	Upfront arrangements Aligning process Availability of a dedicated oncologist at all time Preventive ICU bed reservation after TIL infusion Monitor placement One-on-one nursing care (private room)	Being a CCC Short communication lines Small physical distance between production (research) and clinic Innovative institute Taking over monitor from other department Availability of personnel and private rooms

(Continued on next page)

Table 3 (continued) - Barriers and facilitators in early application of TIL in the Netherlands Cancer Institute.

CTA domain	Barriers (TIL case)	Facilitators / strategies used (TIL case)
3. Organizational **(continued)**	Slow study enrolment	Informative visits at other Dutch specialized melanoma centers (n=13) Informative newsletters to national melanoma experts Involvement in multidisciplinary meetings in other hospitals
4. Technological	Compliance with GMP and ATMP regulations	In-house pharmacy with knowledge on ethical and quality matters Existing quality system which could be adjusted QP already available Gained experience in one production facility (BTU) before implementation in another facility (Sanquin)
	Training	TIL skilled technician to provide trainings High educated staff, familiar with research activities with T-cells
	Alignment to clinical process	Cryopreservation Being a CCC (see organizational domain)
	Intensive process	Dedicated TIL team Accommodate work hours, days and holidays within team Two production facilities available (BTU and Sanquin)
5. Economical	Funding for implementation, upfront costs, and clinical research (phase I,II,III)	Financial support from institute and institutional charity program (preparing facility (GMP) and pilot study) Included in CED program since 2014 (phase III)
6. Future	Current referral rates are low indicating a troublesome national adoption	Future national implementation: Centralize production Start in specialized melanoma centers Educate external clinicians and nurses in expert center (NKI) To anticipate on future adoption: Simplification of production process (cost reduction) Reduce complexity/toxicity of treatment (IL2 and/or lymphodepletion may not be necessary) Biomarker to improve patient selection for TIL, and PD-1 inhibitors

Barriers and facilitators dentified by semi-structured interviews with 26 stakeholders in the constructive technology assessment using the following CTA domains: clinical, patient-related, organizational, technical, economical and future. **Abbreviations:** CCC = comprehensive cancer center (CCC), CED = coverage with evidence development program, GMP = good manufacturing products, IL2 = interleukin-2, NKI = Netherlands Cancer Institute, QP = qualified person.

DISCUSSION

To the best of our knowledge, this is the first study that comprehensively evaluates local and national implementation of a specific ATMP; in our case, early application of TIL-therapy. Our analyses showed that clinical implementation remains complex- mostly explained by the general ATMP barriers identified in literature. Despite these barriers, clinical implementation may become feasible when financial support, regulatory knowledge (GMP and route to market access), and both clinical and technical experience is available.

The CTA survey identified the following barriers: toxicity of TIL-therapy (clinical and patient-related), the need for trainings (clinical, technical), limited patient accrual (patient-related, organizational, future), upfront clinical arrangements (organizational), compliance with GMP and ATMP regulations (technical), and funding for upfront investments and following clinical studies (economic). Nonetheless, TIL-therapy was implemented in the NKI as part of the immunotherapy research program as a result of certain facilitators. The main facilitator was gaining financial support. First, from the institute to create the production facility and to start a pilot study to evaluate the feasibility of providing TIL-therapy. Second from the Dutch CED program which was received in 2014 based on the promising results in the pilot study *(n=10)*[26]. Furthermore, implementation was facilitated by: (i) the availability of an in-house pharmacist with regulatory knowledge and quality assurance department for advice on GMP regulations and quality assurance, (ii) having a TIL skilled technician (from an expert center), (iii) upfront clinical training in an expert center, and (iiii) being a CCC. For future adoption, the costs of TIL-therapy is likely to be a facilitator as it is predicted to be lower compared to ipilimumab and other standard treatments in advanced melanoma. A first model which assessed the cost-effectiveness of TIL-therapy compared to ipilimumab demonstrated that TIL-therapy is dominant over ipilimumab, hence showing higher quality adjusted life years (QALYs) correlated to lower treatment costs.[47]

Limited patient accrual in the RCT remained one of the barriers which may be explained by recent developments in treating advanced melanoma[48]. Since these newly developed treatments show at least similar response rates over the study period compared to TIL-therapy, but are easier to apply (not personalized, available off the shelf)[49]. This could negatively influence the attitude of clinicians towards the potential of TIL-therapy, as speed of adoption is related to the complexity, relative advantage, visibility, trialability, and compatibility[29] of a new treatment strategy which might affect the choice of treatment / participation in the trial by the patient.

Of all the identified barriers in the literature (discussed in Supplement 3) on translating ATMPs, gaining financial support, the route to MA, and compliance to GMP regulations were seen as the main hurdles. Gaining financial support is especially challenging as it is small and medium academic facilities which face the biggest financial risks related to ATMP development and demonstrating treatment efficacy (e.g. high upfront and manufacture costs, to show treatments' efficacy)[11,50]. It has been recognized that in translating innovative, and personalized technologies into the clinic, existing generic regulatory assessments may be unsuitable[44,51]. Therefore, the suitability of the NICE (the national institute for health and care excellence) appraisal methodology for regenerative and cell-based therapies was investigated[44]. They recognized that evidence for the efficacy of regenerative medicines can be associated with high uncertainty levels around long-term costs and benefits. Using existing methods to estimate the implications of this uncertainty - such as calculating cost-effectiveness acceptability curves (CEAC), expected value of perfect information (EVPI) and expected opportunity losses - were considered as sufficient. Yet, the NICE appraisal mentioned recommendations to gather the required data for this regulatory assessment. For instance, (i) use of surrogate endpoints, which should first be validated by systematic reviews, (ii) use of alternative trial designs in rare diseases e.g. single arm trials or responsive-adaptive randomization, and (iii) innovative reimbursement programs aiming to find a balance between shorter approval times and ensuring a flow to gain efficacy and safety data for promising medicines in patient categories with high unmet needs. This final aspect of risk sharing seems to be one of the key recommendations as this also enabled application of TIL-therapy in the NKI-AvL case. The ADAPT SMART project, funded by the EU innovative medicines initiative, seeks for solutions to develop such medicines adaptive pathways to patients (MAPPs)[44,52,53].

Limitations

In our analysis, several limitations emerged. All semi-structured interviews were conducted by one researcher to create uniformity in the several interviews, however this potentially resulted in interview bias. Therefore, we discussed our results with the stakeholders and a second researcher verifying the labels given to the interviews. Secondly, to optimally use the CTA methodology for steering development before application, we should have started the CTA survey before any clinical implementation. However, our first series of interviews were held during the pilot study. Furthermore, in terms of content, we were unable to effectively address the patient impact of TIL-therapy by means of the Impact of Event scale and EQ5D since we included a limited number of patients in the interviews *(n=6)*. In addition, the analysis would have been strengthened, especially towards the patient and clinical domain, by including the first clinical results of the TIL trial. Unfortunately, as the trial is still ongoing, the results could not be analyzed and published. Moreover, the external validity of the cost estimation is limited as Dutch reference prices were used and only a limited number of patients

were included as a basis for our analysis *(n=10)*[54]. However, Table 2 and the specification of the activities can be used by other hospitals to estimate the costs for their situation using country specific unit prices. To have a sense of the treatment costs for providing TIL-therapy (non-commercially) in other countries such as Canada, the United Kingdom and the US we translated our results by using a recent article that compared the healthcare costs across countries and using the GDP per capita of 2015 for these countries[55]. The average conversion rates of 2015 were used to generate a country specific range of the costs. For example, based on the results of Papanicolas et al. total healthcare costs in the US seem to be approximately 80% higher than in the Netherlands[55]. The treatment costs were increased with 80% which resulted in a cost range of $139,448–$168,440 per patient. Using the difference in GDP per capita between the US and the Netherlands the costs were multiplied by 1.25 which resulted in a range of $97,600-$117,891. Thus, the estimated cost range for providing TIL treatment in the US is $97,600-$168,440 per patient. For Canada and the UK a similar translation resulted in cost ranges of: C$89,072-C$116,295 and £32,945-£60,608, respectively. We acknowledge that using these ratios results in a rough estimation. It is however a more accurate indication than only using a conversion rate which neglects the differences in healthcare costs per country at all. In addition, the costs for the production of TILs in our analysis could be an underestimation as the initial costs were estimated on a higher throughput of patients (10 to 20 patients per year). Therefore the costs for TIL production with a smaller number of patients was estimated and included in our cost analysis. A cost-effectiveness analysis could be an informative approach to estimate the minimal throughput of patients per year required to result in a cost-effective alternative to standard of care. Finally, we may not have included all relevant literature because our literature overview was not a systematic literature search. However, we feel that by using snowballing methods we included the most important documents as we found several similarities between the articles regarding the identified barriers/facilitators, and these similarities were endorsed by the technical staff of the specialized production sites.

Future perspective

Implementation of TIL-therapy would have even been more challenging if the NKI-AvL had not enjoyed the financial support of the CED program. This shows that applying financial risk-mitigating principles have a big influence on patient access. Especially for ATMPs, it would be valuable to develop financing strategies with government, industry, research institutes and/or insurance companies to share risks and facilitate uptake of ATMPs. In these strategies, involvement of HTA bodies at an early stage of development would be beneficial as these methods (e.g. headroom analysis, multi decision criteria analysis [MCDA] and value of information analysis [VOI])can help to estimate the potential of the proposed ATMP[11] and give an estimate of the uncertainty surrounding outcome measures. Future research

should focus on the effect of using these risk-sharing principles for ATMPs on patient access. Beyond this, identifying the most valuable early HTA method per translational phase would be worthwhile as these not only assist in identifying critical aspects in the early development process, but also help in gaining financial support and continue on the route to MA by defining and accordingly reducing uncertainty. For TIL-therapy specifically, further research should focus on identifying aspects - such as clinician's attitudes and perceptions on (side) effects - influencing future adoption of TIL-therapy as this was evaluated as a hurdle in the early application of TIL-therapy.

Conclusion

Based on this comprehensive evaluation of early application of TIL-therapy we conclude that implementation is complex and - in at least the preparatory steps - is expensive, but feasible under optimal circumstances (i.e. sufficient financial support, and experience with TIL and GMP). Since TIL-therapy seems to be a potentially cost-effective alternative to ipilimumab[47] other institutes may want to consider TIL-therapy as a standard treatment option. For the TIL case, implementation in an already GMP-certified production facility would facilitate effective technical implementation and minimize some of the financial risks. Additionally, for clinical application, a medical team should be thoroughly trained at a specialist cancer center. Finally, for national implementation it was suggested to start first in the specialized, high volume melanoma centers and to centralize production at the specialized production sites. When these strategies are used, the likelihood of implementing TIL-therapy in clinical practice will increase, and thus provide increased patient access to this promising treatment.

ACKNOWLEDGEMENTS

We want to thank all the stakeholders that participated in the semi-structured interviews. Besides, we thank all patients that participated in the TIL-trial, the ones that completed the web-based questionnaire on factors related to participation, and especially we thank the patients that participated in the semi-structured interviews.

REFERENCES

1. Department for Business Innovation & Skills. Eight great technologies. (2013). Available at: https://www.gov.uk/government/speeches/eight-great-technologies. (Accessed: 11th July 2017)
2. European Medicines Agency. Advanced Therapy Medicinal Products. Available at: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000294.jsp&mid=WC0b01ac05800241e0. (Accessed: 30th May 2018)
3. The European Parliament and the Council of the European Union. *Regulation (EC) No 1394/2007 of the European Parliament and of the Council of 13 November 2007 on Advanced Therapy Medicinal Products and Amending Directive 2001/83/EC and Regulation (EC) No 726/2004*. (2007).
4. EudraLex. Volume 4- Good manufacturing practice (GMP) Guidelines. *European Commission, Public Health.* (2016). Available at: https://ec.europa.eu/health/documents/eudralex/vol-4_en. (Accessed: 24th July 2017)
5. de Wilde, S. *et al.* Hurdles in clinical implementation of academic advanced therapy medicinal products: A national evaluation. *Cytotherapy* **18**, 797–805 (2016).
6. Hartmann-Fritsch, F., Marino, D. & Reichmann, E. About ATMPs, SOPs and GMP: The Hurdles to Produce Novel Skin Grafts for Clinical Use. *Transfus. Med. Hemother.* **43**, 344–352 (2016).
7. Pearce, K. F. *et al.* Regulation of advanced therapy medicinal products in Europe and the role of academia. *Cytotherapy* **16**, 289–297 (2014).
8. Belardelli, F. *et al.* Translational research on advanced therapies. *Ann. Ist. Super. Sanita* **47**, 72–78 (2011).
9. Committee for Advanced therapies. *CAT monthly report of application procedures, guidelines and related documents on advanced therapies.* (2017).
10. Faulkner, A. Opening the gateways to market and adoption of regenerative medicine? The UK case in context. *Regen. Med.* **11**, 321–330 (2016).
11. Bubela, T. *et al.* Bringing regenerative medicines to the clinic: the future for regulation and reimbursement. *Regen. Med.* **10**, 897–911 (2015).
12. Heathman, T. R. J. *et al.* The translation of cell-based therapies: clinical landscape and manufacturing challenges. *Regen. Med.* **10**, 49–64 (2015).
13. EMA. *Advanced therapy medicines: exploring solutions to foster development and expand patient access in Europe*. (2016).
14. Rosenberg, S. A. *et al.* Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clin. Cancer Res.* **17**, 4550–4557 (2011).
15. Andersen, R. *et al.* Long-Lasting complete responses in patients with metastatic melanoma after adoptive cell therapy with tumor-infiltrating lymphocytes and an attenuated il2 regimen. *Clin. Cancer Res.* **22**, 3734–3745 (2016).
16. Merhavi-Shoham, E., Itzhaki, O., Markel, G., Schachter, J. & Besser, M. J. Adoptive Cell Therapy for Metastatic Melanoma. *Cancer J.* **23**, 48–53 (2017).
17. Rosenberg, S. A. *et al.* Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N. Engl. J. Med.* **319**, 1676–1680 (1988).
18. Svane, I. M. & Verdegaal, E. M. Achievements and challenges of adoptive T cell therapy with tumor-infiltrating or blood-derived lymphocytes for metastatic melanoma: what is needed to achieve standard of care? *Cancer Immunology, Immunotherapy* **63**, 1081–1091 (2014).
19. World Health Organization. Health Technology Assessment. Available at: http://www.who.int/medical_devices/assessment/en/. (Accessed: 12th May 2020)
20. IJzerman, M. J. & Steuten, L. M. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Heal. Econ Heal. Policy* **9**, 331–347 (2011).
21. IJzerman, M. J., Koffijberg, H., Fenwick, E. & Krahn, M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *Pharmacoeconomics* **35**, 727–740 (2017).
22. Husereau, D., Henshall, C., Sampietro-Colom, L. & Thomas, S. CHANGING HEALTH TECHNOLOGY ASSESSMENT PARADIGMS? *Int. J. Technol. Assess. Health Care* **32**, 191–199 (2016).
23. Miquel-Cases, A. *et al.* (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. *Cancer Treat. Rev.* **52**, 117–127 (2017).

24. Schot, J. & Rip, A. The past and future of constructive technology assessment. *Technol. Forecast. Soc. Change* **54**, 251–268 (1997).
25. Douma, K. F. L., Karsenberg, K., Hummel, M. J. M., Bueno-de-Mesquita, J. M. & van Harten, W. H. Methodology of constructive technology assessment in health care. *Int. J. Technol. Assess. Health Care* **23**, 162–168 (2007).
26. van Harten, W. H. & Retèl, V. P. Innovations that reach the patient: early health technology assessment and improving the chances of coverage and implementation. *Ecancermedicalscience* **10**, 683 (2016).
27. Zorginstituut Nederland. *Voorwaardelijke toelating tot het basispakket Voortgangsrapportage 2017 (Dutch report)*. (2017).
28. Geukes Foppen, M. H., Donia, M., Svane, I. M. & Haanen, J. B. A. G. Tumor-infiltrating lymphocytes for the treatment of metastatic cancer. *Mol. Oncol.* **9**, 1918–1935 (2015).
29. Rogers, E. M. The Diffusion of Innovations. *Diffus. Innov.* 576 (2003).
30. NVivo qualitative data analysis Software. QSR International Pty LTd. (2012).
31. Igwe, E. *et al.* Patient perceptions and willingness to participate in clinical trials. *Gynecol. Oncol.* **142**, 520–524 (2016).
32. Penman, D. T. *et al.* Informed consent for investigational chemotherapy: patients' and physicians' perceptions. *J. Clin. Oncol.* **2**, 849–55 (1984).
33. Jenkins, V. *et al.* Drivers and barriers to patient participation in RCTs. *Br. J. Cancer* **108**, 1402–1407 (2013).
34. Kaur Rosalind L.; Williamson, Paula, G. S. Developing a survey of barriers and facilitators to recruitment in randomized controlled trials. *Trials* **13**, 218 (2012).
35. Chapko, M. K. *et al.* Equivalence of two healthcare costing methods: Bottom-up and top-down. *Health Econ.* **18**, 1188–1201 (2009).
36. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. *Manual for cost research: methods and standard cost prices for economic evaluations in health care*. (2015).
37. Zorginstituut Nederland (Dutch institute of healthcare). Medicijnkosten (costs of pharmaceuticals). (2019). Available at: http://www.medicijnkosten.nl/. (Accessed: 2nd April 2020)
38. Dutch Healthcare Authority (NZa). DBC product finder for tariffs. (2019). Available at: http://dbc-zorgproducten-tarieven.nza.nl. (Accessed: 2nd February 2020)
39. Retèl, V. P., Joore, M. A., Linn, S. C., Rutgers, E. J. T. & van Harten, W. H. Scenario drafting to anticipate future developments in technology assessment. *BMC Res. Notes* **5**, 442 (2012).
40. Galli, M. C. ATMPs for Cancer Immunotherapy: A Regulatory Overview. in *Methods in Molecular Biology* **1393**, 1–9 (2016).
41. Abou-El-Enein, M., Elsanhoury, A. & Reinke, P. Overcoming Challenges Facing Advanced Therapies in the EU Market. *Cell Stem Cell* **19**, 293–297 (2016).
42. Gardner, J., Faulkner, A., Mahalatchimy, A. & Webster, A. Are there specific translational challenges in regenerative medicine? Lessons from other fields. *Regen. Med.* **10**, 885–895 (2015).
43. Ali, R., Hollander, A., Kemp, P., Webster, A. & Wilkins, M. Regulating cell-based regenerative medicine: the challenges ahead. *Regen. med.* **9**, 81–87 (2014).
44. Corbett, M. S., Webster, A., Hawkins, R. & Woolacott, N. Innovative regenerative medicines in the EU: a better future in evidence? *BMC medicine* **15**, 49 (2017).
45. Boráň, T. *et al.* Clinical Development and Commercialization of Advanced Therapy Medicinal Products in the European Union: How Are the Product Pipeline and Regulatory Framework Evolving? *Hum. Gene Ther. Clin. Dev.* **28**, 126–135 (2017).
46. Donia, M., Larsen, S. M., Met, Ö. & Svane, I. M. Simplified protocol for clinical-grade tumor-infiltrating lymphocyte manufacturing with use of the Wave bioreactor. *Cytotherapy* **16**, 1117–1120 (2014).
47. Retel, V. P., Steuten, L. M. G., Mewes, J. C. & van Harten, W. H. Early Cost-Effectiveness Modeling for Tumor Infiltrating Lymphocytes (TIL)-Treatment Versus Ipilimumab in Metastatic Melanoma Patients. *Value Heal. J. Int. Soc. Pharmacoeconomics Outcomes Res.* **17**, A640 (2014).
48. Sznol, M. Challenges in Conducting Clinical Research on Patients With Advanced Melanoma. *Cancer J.* **23**, 75–78 (2017).
49. Larkin, J. *et al.* Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N. Engl. J. Med.* **373**, 23–34 (2015).

50. House of Lords Science and Technology Committee. Regenerative Medicine Report. (2017). Available at: https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/275/27502.htm. (Accessed: 24th July 2017)
51. Faulkner, E. *et al.* Challenges in the development and reimbursement of personalized medicine-payer and manufacturer perspectives and implications for health economics and outcomes research: a report of the ISPOR personalized medicine special interest group. *Value Heal. J. Int. Soc. Pharmacoeconomics Outcomes Res.* **15**, 1162–1171 (2012).
52. ADAPT SMART-. *Accelerated Development of Appropriate Patient Therapies* Available at: http://adaptsmart.eu/. (Accessed: 8th August 2017)
53. Hettle, R. *et al.* The assessment and appraisal of regenerative medicines and cell therapy products: an exploration of methods for review, economic evaluation and appraisal. *Health Technol. Assess.* **21**, 1–204 (2017).
54. Welte, R., Feenstra, T., Jager, H. & Leidl, R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. *Pharmacoeconomics* **22**, 857–876 (2004).
55. Papanicolas, I., Woskie, L. R. & Jha, A. K. Health Care Spending in the United States and Other High-Income Countries. *JAMA* (2018). doi:10.1001/jama.2018.1150
56. European Commission. *Guidelines on Good Manufacturing Practice specific to Advanced Therapy Medicinal Products.* (2017).
57. Scientific Committees of the European Commission. Risk assessment. Available at: http://ec.europa.eu/assets/sante/health/scientific_committees/risk_assessment/index_en.htm. (Accessed: 24th July 2017)
58. Flory, E. & Reinhardt, J. European regulatory tools for advanced therapy medicinal products. *Transfusion Medicine and Hemotherapy* **40**, 409–412 (2013).
59. Salmikangas, P. *et al.* Marketing regulatory oversight of advanced therapy medicinal products (ATMPs) in europe: The EMA/CAT perspective. in *Advances in Experimental Medicine and Biology* **871**, 103–130 (2015).
60. Committee for Medicinal Products for Human Use. *Guideline on the requirements to the chemical and pharmaceutical quality documentation concerning investigational medicinal product in clinical trials.* (2006).

SUPPLEMENTARY MATERIAL

Supplement 1 - Search strategy literature overview ATMP barriers and facilitators

Results literature search Searches in July 2017 via Pubmed using filter: Past 5 years	
Terms	Results
advanced-therapy medicinal products AND regulation	41
advanced therapy medicinal products AND implementation	17 (after removing duplicates)
advanced therapy medicinal products AND translation	7 (after removing duplicates)
65 unique articles identified	
57 articles were excluded because: • The article was not about ATMPs (10) • The article was not available in English (9) • The article described ATMP regulation issues outside of Europe (2) • The article described matters related to a specific ATMP product but not the implementation barriers/facilitators of this specific product (3) • The article was related to ATMP issues but did not describe implementation barriers/facilitators (11) • The article focused on the regulation issues related to ATMPs only (e.g. providing regulatory frameworks) (22) 8 articles were included because of a description of either barriers or facilitators on ATMP implementation in the clinic.	
Articles included: [45] [10] [11] [41] [5] [7] [6] [40]	
Snowball	
From	Identified and included
Faulkner 2016 [10]	[42]
Suggestions from journal: Regenerative medicine (future medicine) after reading articles in their journal.	[43]
"people who read this article, also read:"	[12]
Regulatory website describing that authors from House of Lords Science and Technology Committee wrote a paper on this issue [22]	[44]

The reference numbers are similar to the ones used in the manuscript.

Supplement 2 - Web-based questionnaire aimed at advanced melanoma patients

Introduction

Welcome to this questionnaire evaluating factors related to TIL trial participation. This survey is aimed at stage 3 and 4 melanoma patients.

Tumor-infiltrating lymphocytes therapy is an immunotherapy that aims at strengthening immune response. Recent literature shows approximately 50% chance on response resulting in stable disease or complete remission of tumors. Based on these promising results, TIL-therapy is conditionally included in the insurance package of the Netherlands despite it is still under investigation. TIL treatment is thus only given in a research setting in the Antoni van Leeuwenhoek hospital in which the (cost-)effectiveness of TIL is being evaluated, compared to ipilimumab (another type of immunotherapy).

If you are not familiar with the TIL treatment, below some extra information is listed.

By means of a surgery, a tumor lesion is removed from which immune cells are isolated, the tumor-infiltrating lymphocytes. These are grown in approximately five weeks to a billion of cells. One week before TIL infusion, chemotherapy is given to create space for the big amount of cells that will be infused later on. Then the TILs are infused whereafter treatment with a growth factor will start to create an optimal environment for these cells. This treatment can result in severe side effects that are comparable with high fever.
After approximately three weeks of admission in the hospital (chemotherapy, TIL infusion and recovery) treatment is completed and the patient can recover from this therapy at home. TIL-therapy is thus a one-time, but intensive treatment.

Via the following links more information can be gathered about the study. (Dutch)
https://www.avl.nl/topmenu/over-avl/nieuws/nieuwe-behandeling-voor-uitgezaaide-melanoom-voorlopig-toegelaten-tot-verzekerde-zorg/
https://www.win-o.nl/klinische-studies/melanoom-trials/til
This questionnaire aims to identify factors related to the decision whether or not to participate with the TIL study. It is a project initiated by the Antoni van Leeuwenhoek hospital in collaboration with Stichting Melanoom. The questionnaire is aimed at stage 3 and 4 melanoma patients and will take approximately 15 minutes. Answers will be confidential and handled anonymously.

Thank you for your time and effort.

General domain

1. Age: _____ year
2. Gender: M / V
3. Education
   - ◯ Lower or primary school
   - ◯ Secundary school
   - ◯ Secondary vocational education
   - ◯ Upper secondary vocational education
   - ◯ Higher and academic education
   - ◯ Other, namely……

4. Diagnosis
   - ◯ Stage 1 melanoma — submit form (end questionnaire)
   - ◯ Stage 2 melanoma — submit form (end questionnaire)
   - ◯ Stage 3 / 4 advanced melanoma or non resectable melanoma
   - ◯ I don't know exactly but advanced melanoma
   - ◯ I don't know exactly but metastasized melanoma
   - ◯ I don't know

5. If you have been previously treated for melanoma, what was this treatment? (you can choose mutiple options)
   - ◯ Surgical removal
   - ◯ Chemotherapy, namely: _____(e.g. dacarbazin)
   - ◯ Immunotherapy, namely:_____(e.g. ipilimumab (yervoy), nivolumab (opdivo), pembrolizumab (keytruda), combination therapy, TIL)
   - ◯ Personalized treatment, namely: _____(e.g. trametinib (mekinist), dabrafenib (tafinlar), vemurafenib (zelboraf))
   - ◯ N.A. / I don't remember
   - ◯ Other, namely……

6. In which hospital have you been treated? (not mandatory)

7. Were you informed by your clinician about the TIL study/treatment? Or were you informed in another way?
   - ◯ No
   - ◯ Yes, it was discussed with me
   - ◯ Yes, I received information
   - ◯ Yes, it was discussed and I received information

   - ◯ If yes, was the TIL study at that moment a treatment option?
   - ◯ No
   - ◯ Yes
   - ◯ N.A.

8. In addition, have you become familiar with the TIL treatment / study via one of the following channels?
   - ◯ No, I have not obtained information about the TIL study in another way
   - ◯ Via media: television and / or (newspapers) articles
   - ◯ Came to my attention in fora/ Facebook / other online media
   - ◯ I found information on the internet
   - ◯ Was brought to my attention during an event (e.g. congress, peers, event from patient association) where I was present
   - ◯ Other, namely: ……

9. Did you consider, participation with the TIL trial? (Ipilimumab versus TIL)
   - ◯ No, I was not informed about the trial — Go to question 12
   - ◯ No, I was not eligible for this treatment/trial — Go to question 10
   - ◯ Yes, but I chose not to participate — Go to question 11
   - ◯ Yes, and I participated — Go to question 12

10. What was the reason that you were not eligible for the trial? (open)

11. What was the main reason for you to not participate with the TIL trial? (open)

A page with information about TIL trial, explaining both treatments, expected effectiveness and potential side effects. Containing videos and references to scientific articles on both TIL as well as ipilimumab.

12. Below you will find several factors that may influence the decision to participate with a trial. In the case of TIL, could you please describe whether these aspects have or would have an influence on the decision to participate. If you have any additional factors that influenced the decision there will be room for this later in the questionnaire.

Factor	Strong negative influence	Negative influence	Neutral	Positive influence	Strong positive influence
Travel distance and / or travel costs (TIL-therapy is only given in Amsterdam)	1	2	3	4	5
Expected additional (healthcare) costs for participation	1	2	3	4	5
Additional meetings and/or investigations in relation to the trial (e.g. biopsies)	1	2	3	4	5
Expected side effects of TIL-therapy	1	2	3	4	5
Expected promising results (e.g. survival) of TIL-therapy	1	2	3	4	5
Possibility of receiving the control arm of the trial (ipilimumab)	1	2	3	4	5
It is a randomized trial (chance)	1	2	3	4	5
Clinician recommends me to participate with this trial	1	2	3	4	5
Family and or friends recommend to participate with the trial	1	2	3	4	5
The need to switch to another hospital	1	2	3	4	5

13. Please describe your level of agreement with the following statements

#	Statement	Strongly agree	Agree	Nor agree nor disagree	Disagree	Strongly disagree	I don't know
1	TIL treatment is the best possible option in my situation	1	2	3	4	5	0
2	The advantages of TIL-therapy (expected response rate) outweighs expected side effects	1	2	3	4	5	0
3	Both treatments in the TIL-trial would be a good option for me	1	2	3	4	5	0
4	I strongly prefer to receive TIL-therapy in this trial	1	2	3	4	5	0
5	I was afraid for progression if I wouldn't start with the trial.	1	2	3	4	5	0
6	The idea of randomization worried me.	1	2	3	4	5	0
7	I received sufficient information about the trial and treatment	1	2	3	4	5	0
8	I could not say no to participation	1	2	3	4	5	0
9	I wanted to stay in my own hospital or with the same clinician for my treatment.	1	2	3	4	5	0
10	I wanted to help clinicians with their research	1	2	3	4	5	0
11	I think my participation will be beneficial for other patients	1	2	3	4	5	0
12	The clinician wanted for me to participate	1	2	3	4	5	0
13	Family and friends wanted me to participate	1	2	3	4	5	0

14. Which of the above statements would be the most important one for when deciding to participate in a trial (1-13)

15. Which additional factors (would have) played a role in deciding to participate with the TIL trial?

◯ N.A.

◯ _____ (open)

16. After this questionnaire, do you consider to participate with the TIL trial?

◯ No

◯ Yes

17. Can you shortly comment on this?

_____

18. If 17 answered with yes, do you want to receive more information from one of our clinicians? Please leave your email address or telephone number and we will contact you.

- ◯ Yes, I want to receive more information via: _____
- ◯ Yes I would like to be called on: _____

19. Regarding information provision, what would you recommend us?

- ◯ Nothing, I had sufficient information
- ◯ More information via internet
- ◯ More information via specialists
- ◯ Information flyer for patients
- ◯ Designing a website for the study
- ◯ Possibility to ask questions to one of the clinicians involved in the trial.
- ◯ Other, namely _____

20. Do you have any final remarks? You can describe them here. (open)

On behalf of Stichting Melanoom and Antoni van Leeuwenhoek hospital, we want to thank you for your time and participation to this survey!

Supplement 3 - Results literature overview ATMP barriers and facilitators

Table supplement 3 - Identified barriers and facilitators from the literature overview on clinical implementation of ATMPs structured according the six CTA domains.

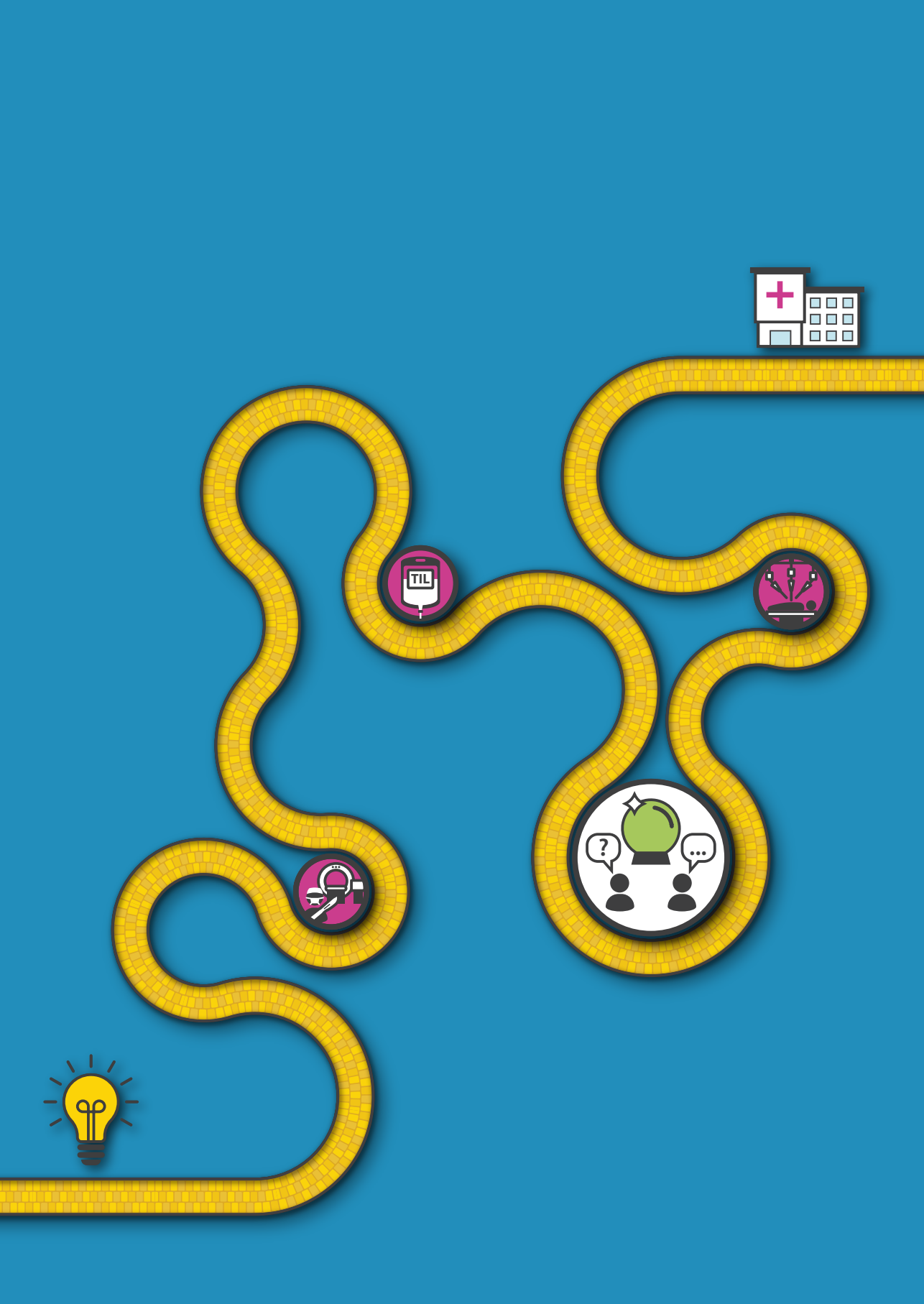
CTA domain	Barrier (general ATMPs)	(Potential) Facilitators (general ATMPs)
1. Clinical	Therapeutic risks because of using live tissue[1] Dosing and administration different from usual drugs[3] Dosing based on pre-clinical data should be handled carefully[2] Need for training of clinicians[9]	Use clinical implementation model in which trained personnel takes responsibility to integrate new therapy into routine clinical practice[1]
2. Patient-related	Therapeutic risks because of using live tissue[1] Patient enrollment[7]	
3. Organizational	MA[3,10] Study designs and feasibility of monitoring efficacy[1,2,3,4,5,6,9] : • Complex regulatory environment[6,9] • Heterogeneity in regulations and HEC implementation across Europe[2,5,12] Lack of GMP compliant facilities (adoption)[12] Institutional: Gain management support[9] Hard to change existing workflows[7,9]	Improve MA Contact regulatory authorities early in product development stage to anticipate on challenges faced per specific product[3] Drafting legislation and guidelines that provide streamlined reimbursement of EU countries[1] Engagement of HTA organization alongside development process to support value assessment[2,4] Greater coordination between EMA and national agencies[9] Develop process with a realistic reimbursable price point at all stages of scale (low costs of goods)[11] Innovative study design: accept surrogate endpoints and non-straight forward comparator[2,5] Advice from EMA on clinical trial design and MA[2] Abou-El-Enein (2016) developed a stepwise model for MA describing facilitators[1] Establish a cell therapy center of excellence to provide guidance and knowledge (adoption)[9] Introduce cryopreservation steps to control for the process[11]

(continued on next page)

Table supplement 3 (continued) - Identified barriers and facilitators from the literature overview on clinical implementation of ATMPs structured according the six CTA domains.

CTA domain	Barrier (general ATMPs)	(Potential) Facilitators (general ATMPs)
4. Technological	Lack of regulatory knowledge (GMP, ATMP)[1,6] Comply with GMP regulation to anticipate on risks with using live tissue and assure quality:[2,8,9] • In process controls[8,10] • Risk analysis[8,10] • Documentation (e.g. IMPD)[9,10,12] • Consumables, raw and starting material[6,8,10] • Assigning / hiring QP[9,10] • Training of staff[2,10] In clinical trials: establishment of safety monitoring board (IDSMB)[10] Upscaling production[2] Unclear responsibilities when exported[6] In exporting: local QP needed for release[9,12] Under developed infrastructure[9] Transport[10] Rapidly evolving field: selected ATMP overtaken by other innovative medicinal product[6] Clinicians reluctant because of potential risks:[1] • Patient enrollment[6] • Lack of motivation for implementation[9]	To improve GMP implementation: Documents on GMP guidelines:[3,6,8] • [56, 4, 57, 58, 59] Use a standard method for risk analysis: e.g. FMEA[10] Guidance on IMPD in Committee for Medicinal Products for Human Use (2006) and [10] Set up a knowledge platform to support development route of ATMPs[6,9] Education or training organized by government or academic institutions [6] Quality: Use semi-closed followed by a closed process[11] Improve scaling-up: Collaboration of Production Facilities? (in and over countries)[10] Include cryopreservation steps in the production process for minimalizing failures in the process when scaling up[11] Consider scaling out instead of scaling up[5]
5. Economical	Inadequate financial support for: Required investments for GMP[3,9,10,11,12] High manufacturing costs (small target population)[1] Clinical trial costs[4,6,10] (hard to receive research funding) Limited flexibilities by funding agencies (e.g. extended time lines)[6]	Adaptive licensing approaches e.g. accelerated assessment, exceptional circumstances licensing, conditional approval mechanisms and other risk sharing principles[2,4,5,7,9] More automative process and more productions per year could reduce costs[11]
6. Future	n.a.	n.a.

[1] Abou-El-Enein et al., 2016; [2] Ali et al 2014; [3] Boran et al 2017; [4] Bubela et al 2015; [5] Corbett et al., 2017; [6] de Wilde et al., 2016; [7] Faulkner, 2016; [8] Galli, 2016; [9] Gardner et al., 2015; [10] Hartmann-Frisch et al., 2016; [11] Heathman et al., 2015; [12] Pearce et al., 2014. **Abbreviations:** MA = market approval, GMP = good manufacturing products, ATMP = advanced therapeutic medicinal product, IMPD = investigational medicinal product dossier, QP = qualified person, QC = quality controller, IDSMB = independent data and safety monitoring board.


TIL
?
...

8

Evaluating different adoption scenarios for TIL-therapy and the influence on its (early) cost-effectiveness

Melanie Lindenberg
Valesca Retèl
Maartje Rohaan
Joost van den Berg
John Haanen
Wim van Harten

BMC Cancer (2020), Jul 31; 20 (1): 712

ABSTRACT

Background

Treatment with tumor-Infiltrating Lymphocytes (TIL) is an innovative therapy for advanced melanoma with promising clinical phase I/II study results and likely beneficial cost-effectiveness. As a randomized controlled trial on the effectiveness of TIL-therapy in advanced melanoma compared to ipilimumab is still ongoing, adoption of TIL-therapy by the field is confronted with uncertainty. To deal with this, scenario drafting can be used to identify potential barriers and enables the subsequent anticipation on these barriers. This study aims to inform adoption decisions of TIL by evaluating various scenarios and evaluate their effect on the cost-effectiveness.

Methods

First, 14 adoption scenarios for TIL-therapy were drafted using a Delphi approach with a group of involved experts. Second, the likelihood of the scenarios taking place within 5 years was surveyed among international experts using a web-based questionnaire. Third, based on the questionnaire results and recent literature, scenarios were labeled as being either "likely" or "-unlikely". Finally, the cost-effectiveness of TIL treatment involving the "likely" scored scenarios was calculated.

Results

Twenty-nine experts from 12 countries completed the questionnaire. The scenarios showed an average likelihood ranging from 29% to 58%, indicating that future developments of TIL-therapy were surrounded with quite some uncertainty. Eight of the 14 scenarios were labeled as "likely". The net monetary benefit per patient is presented as a measure of cost-effectiveness, where a positive value means that a scenario is cost-effective. For six of these scenarios the cost-effectiveness was calculated: "Commercialization of TIL production" (€-51,550), "Pharmaceutical companies lowering the prices of ipilimumab" (€11,420), "Using TIL-therapy combined with ipilimumab" (€-10,840), "Automatic TIL production" (€22,670), "TIL more effective" (€23,270), "Less Interleukin-2" (€20,370).

Conclusions

Incorporating possible future developments, as in the base case, TIL-therapy was calculated to be cost-effective compared to ipilimumab in the majority of "likely" scenarios. These scenarios could function as facilitators for adoption. If treatment with TILs is proven to be effective, the scenarios resulting in TIL-therapy not being cost-effective should be considered in the adoption decision as these may act as crucial barriers.

INTRODUCTION

Over the past decade, the treatment landscape for advanced melanoma has greatly developed due to the introduction of checkpoint inhibitors and targeted therapies. This resulted in a rise of the 5-year survival rate from 10%[1] up to 52%[2] when using the most recent and promising treatment combination of nivolumab with ipilimumab.

Despite the improved clinical outcomes, a large group of patients still fail to respond or progress after initial response upon the available treatments. Therefore, identification of additional treatment options for second-line treatment is of interest. Adoptive cell therapy with tumor-infiltrating lymphocytes (TIL) could be one of these additional treatment options. In TIL-therapy, T cells residing in patient specific tumor material are isolated and expanded ***ex vivo*** in a dedicated production facility and given back to the patient as a single intravenous infusion after a lymphodepleting non-myeloablative preparative regimen and subsequent treatment with interleukin-2 (IL-2). TIL treatment was introduced in small clinical trials in the '80s[3] and several research groups independently showed consistent objective response rates of 40-70%[4–6] and complete response rates of 10%-25%[7], in subsequent small clinical phase I/II trials. However, this therapy has not yet been widely adopted. This can mainly be explained by the lack of phase III evidence of the clinical effectiveness of TIL-therapy and the complex nature of this innovative cellular product advanced therapy medicinal product (ATMP) of which clinical implementation is known to be challenging[8,9].

Since October 2014, the Netherlands Cancer Institute (NKI) and the Herlev hospital in Denmark have been conducting the first randomized controlled trial (RCT) comparing TIL-therapy to ipilimumab as second-line treatment for advanced melanoma to evaluate its clinical and cost-effectiveness (NCT02278887). For the Netherlands, this trial is included in a coverage with evidence development (CED) program for highly promising treatments[10]. This RCT aims to provide the evidence needed to widely adopt TIL-therapy as a standard second-line treatment modality in advanced melanoma. As this trial is still ongoing, the decision for other centers and/or countries to adopt TIL-therapy is surrounded with great uncertainty or is delayed. Especially delay could affect timely patient access when TIL-therapy is proven to be effective, as clinical implementation of TIL-therapy is challenging and time-consuming[11].

In the framework of the CED program, a broad technology assessment (TA) is conducted to facilitate this clinical adoption of TIL-therapy. Within this TA, an early cost-effectiveness analysis was conducted, showing that TIL-therapy is cost-effective over ipilimumab as second-line treatment of advanced melanoma based on the currently available evidence[12]. Furthermore, a qualitative study was conducted evaluating barriers and facilitators in the

clinical implementation of TIL-therapy in light of an ATMP[11]. This study showed that its adoption can be influenced by many factors, such as attitude of clinicians and patients due to the expected therapeutic risks and the rapid evolving treatment field for advanced melanoma.

The current RCT conducted at the NKI and the final project in this TA aims to reduce the existing uncertainty surrounding the decision to clinically adopt TIL-therapy as a second-line treatment for advanced melanoma. The objective of this paper is to evaluate various adoption scenarios related to TIL-therapy and the treatment landscape of advanced melanoma. The likelihood of these scenarios to occur within five years was estimated to identify potential barriers and facilitators for the adoption of TIL-therapy and lastly, the likely adoption scenarios were evaluated in terms of cost-effectiveness.

METHODS

In this study we will often refer to "adoption scenarios", which are one-sentence descriptions of potential developments that may affect the adoption of TIL-therapy.

Drafting adoption scenarios (Delphi methodology)

A Delphi method was used to systematically generate consensus on themes related to the adoption of TIL-therapy to incorporate in the adoption scenarios. Figure 1 shows the six steps used to draft the scenarios[13,14].

First, relevant themes that could influence the adoption of TIL-therapy were identified by means of: brainstorming with internal experts, reviewing literature on TIL-therapy and research developments in treating advanced melanoma, and scanning ongoing clinical trials investigating TIL-therapy. In the second step, the identified *themes* were discussed during semi-structured interviews with stakeholders in the TIL study process at the NKI to identify their expectations on these *themes* for the coming years[11]. They were allowed to add new *themes* and were specifically asked to describe likely "what if" scenarios for the coming five and ten years[13]. The details on these semi-structured interviews are described in a previous publication[11]. The results of the interviews were discussed with the direct research group (ML, VR, WvH) in the third step, where the final *themes* were chosen to incorporate in the first (pilot) set of adoption scenarios. This first set of adoption *scenarios* (15 scenarios and two questions) was piloted in an expert group consisting of lab members, health insurers, clinicians, researchers, a representative of a patient association, a board member of the Dutch Immunotherapy Working Group for Oncology (WIN-O), and policy advisers (step 4). The set was adapted according to their given feedback. The final set consisted of 15 adoption

scenarios and 5 questions on, for example, minimal effectiveness, patients and clinicians attitude towards TIL-therapy (Table 1).

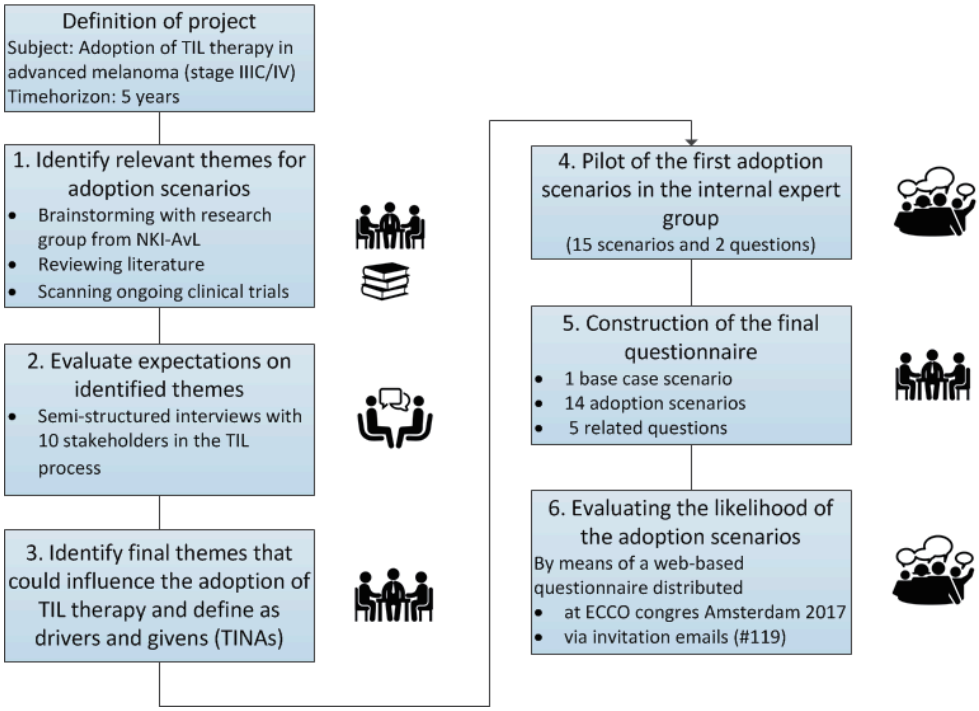


Figure 1. Schematic visualization of method and steps in drafting scenarios. This approach was based on the methods described by Shell international BV (2008) and Enserink and Hermans (2010)[14,15].

8

Estimating likelihood of scenarios

The adoption scenarios and questions were included in a web-based questionnaire (Supplement 1) and were shared among a larger group of experts to evaluate the likelihood of the scenarios happening in the coming five years. To reach international clinical experts, flyers regarding the questionnaire were distributed at the congress of the European cancer congress organization (ECCO) in Amsterdam (January 2017) after melanoma-related sessions. Additionally, the questionnaire was emailed to the scientific and clinical network of our internal experts, by which we invited 119 international experts; all were reminded after one month.

The questionnaire consisted of three parts. Part one introduced the TIL-therapy and the RCT that is currently ongoing. Part two evaluated the characteristics of the respondent (years of experience with TIL-therapy and years of experience with melanoma care, their position and their self-reported level of expertise with TIL-therapy)[16]. The third part contained the 15

adoption scenarios and the five questions as listed in Table 1. In this part, the respondents indicated their estimation on the likelihood of the scenario's occurring in the coming five years from 0% to 100%. 0% indicates that the scenario will not occur within five years, and 100% indicates that the scenario will definitely occur within five years. This method is similar to the method used in a publication focusing on the adoption of Next Generation Sequencing[17]. Table 1 lists the names of the scenarios which are in the following sections to refer to the specific scenarios.

Calculating the cost-effectiveness

Selection of scenarios

As the likelihood of the 15 adoption scenarios showed a lot of uncertainty, we followed several steps to label the scenarios as "likely" or "unlikely". The process to select the "likely" scenarios to incorporate in the cost-effectiveness analysis is visualized in Supplement 2 and described in the section below.

To start, the mean likelihood of each scenario was evaluated. A scenario with a mean likelihood of ≥55% was labeled as "likely". The scenarios that scored a likelihood <55% were stratified in two ways; first, on the answers given to the level of expert by evaluating the results of the respondents that described themselves as "familiar" and "expert" (n=23), and second for the level of experience evaluating the results from the respondents with ≥1 year experience with TIL-therapy (n=10). For the scenarios that still showed a score <55%, a recent literature review was used[18]. When a topic related to the scenario was described in the review, the scenario was labeled as "likely". Finally, if literature was also indecisive, the unlabeled scenarios were discussed and judged among experts (two clinicians, one technician and a policy adviser) involved in the TIL study at the NKI, in which also the results on the five questions were discussed. Besides, the expert panel was asked to verify the likelihood of the scenarios labelled "likely" based on the cutoff value.

As it is plausible that several scenarios will take place at the same time, the same group of experts defined possible combinations of the "likely" scenarios. These were additionally incorporated in the cost-effectiveness model.

Table 1 - Themes identified to draft scenarios and full description of scenarios.

Identified themes (result of step 2-4)	
Less or even no interleukin-2, More automatic process, Attitude of clinicians, Costs of TIL, Take-over by a commercial party, Effectiveness TIL and others, Target population, Long term effectiveness, Attitude of patients, Unexpected clinical risks, Influence of pharmacy, Placement of TIL in treatment strategy	
Name of scenario	**Full description of scenarios**
Base case	If TIL shows better survival rates (at least 10% improvement) compared to ipilimumab, TIL will be implemented in specialized melanoma centers.
Competition	Competing (immuno)therapies are equal in costs but 10% more effective compared to TIL.
TIL more effective	The effectiveness of TIL has increased with 10% (clinically relevant) due to research developments.
Biomarker	A biomarker, being able to select patients for TIL, is available.
TCR therapy	TCR therapy dominates TIL treatment in advanced melanoma, regardless other treatment modalities.
Patients unconvinced	Patients prefer the competing therapies over TIL based on complete information on toxicities and effectiveness.
2nd line treatment	TIL is implemented as a second line treatment after anti PD1 inhibitors in metastatic melanoma.
3rd line treatment	TIL is implemented as a third line (last resort) treatment in metastatic melanoma.
Combination therapy	TIL is used in combination with other immune or personalized therapies (i.e. nivolumab or vemurafenib).
Clinicians unconvinced	Clinicians are not willing to implement TIL because of one of the previous stated reasons.
Low cost competition	If TIL turns out to be cost-effective, pharmaceutical companies will lower the prices of competing immunotherapies.
Influence by companies	Arrangements between pharmaceutical companies and hospitals and/or doctors, negatively affect patient selection for TIL-therapy.
Less IL2 treatment	Additional interleukin-2 treatment after infusion of TIL is not be necessary anymore.
TIL production outsourced	Production of TIL is of interest for the pharmaceutical market and is outsourced by a commercial company.
Automatic TIL production	Production of TIL is less expensive (30% reduction) due to more automatic process steps.
Questions	
What would be the minimal effectiveness of TIL leading to accept TIL as a standard therapy for you? Expressed in one-year survival rate (%)?	
What would be the risk of developing other types of cancer such as lymphomas by activating the immune system by injecting TILs (%)?	
In which level do you agree with the following statement: "TIL treatment provides significantly better quality of life compared to ipilimumab"?	
Could you estimate the percentage of the eligible patients (metastatic melanoma patients) you think is aware of TIL-therapy as a potential treatment (in %)?	
What would be the main reason for clinicians to be unconvinced of introducing TIL-therapy?	

The base case model

A base case model is the original model used to evaluate the cost-effectiveness of an alternative treatment compared to the current standard of treatment using the best available evidence at that moment. In the current study, this is the cost-effectiveness model previously described by Retel et al. (2018)[12]. The model contained three health states: stable disease, progressive disease and death (absorbing state). The time horizon was 10 years, reflecting an average lifetime time horizon of this patient group, with a cycle time of one year. Details on this model can be found in the original research paper[12] and Supplement 3. This model assumed that TIL-therapy is available as a second-line treatment. For clarity, we assumed that there would be no changes in costs and effects of TIL-therapy and ipilimumab over the coming five years.

Incorporating the selected scenarios

The scenarios labelled as "likely" were incorporated in the cost-effectiveness model. With the experts (two clinicians, one technician and a policy adviser) involved in the TIL study at the NKI, logical consequences were defined per scenario and were then translated to input parameters for the model. For some scenarios an additional literature search was performed to feed the cost-effectiveness model. Although assumptions could be made for the efficacy of the scenario to use TIL-therapy in third line based on literature[5], no data or literature was found describing Progression Free Survival and Overall Survival data of chemotherapy after progression on PD-1 inhibitors and CTL-4 antibodies, to serve as the comparator[19]. Therefore, this scenario wasn't incorporated in the cost-effectiveness model. The scenario specific input parameters, assumptions and sources per scenario are listed in Table 2.

Table 2 - Adapted input parameters for cost-effectiveness model per scenario.

Adapted parameter	Initial deterministic value	Deterministic value	SE	Distribution	Source / Assumption
Scenario: "TIL more effective"					
PFS TIL	0.234	0.257	0.068	Beta	Assumption: 10% increase of survival rates as described in the scenario
OS TIL	0.412	0.453	0.046	Beta	
Scenario: "Combination therapy"					
PFS TIL	0.234	0.264	0.089	Beta	12mo PFS 4/13 patients[20] SE was kept the same as the initial model
OS TIL	0.412	0.499	0.098	Beta	12mo OS 9/13 patients[20] SE was kept the same as the initial model
Costs TIL	€62,000	€107,744	€13,743	Gamma	On average 2 times ipilimumab and administration costs and costs to anticipate on the side effects (€693.75 + €45,050)[20,21].

(continued on next page)

Table 2 (continued) - Adapted input parameters for cost-effectiveness model per scenario.

Adapted parameter	Initial deterministic value	Deterministic value	SE	Distribution	Source / Assumption
Scenario: "Combination therapy" (continued)					
Failure rate	0.10	0.10	0.015	Beta	1/13 received no TIL due to progression during TIL growth; 1 patient did not receive ipilimumab after TIL due to dose-limiting colitis.[20] Assumed to be similar as basecase model.
Scenario: "Low cost competition"					
Drug costs Ipilimumab	€90,100	€71,184	€9,080	Gamma	Reduced price for ipilimumab in such a way that TIL is not cost-effective anymore with a willingness to pay threshold of €30,000. A reduction of 21%.
Scenario: "Less IL2 treatment"					
Total TIL costs	€62,000	€61,450	€7,838	Gamma	Assuming the decrescendo regimen described by Andersen et al 2016 6 vials of Aldesleukin (Novartis)[21] €550 reduced compared to the initial costs.
Utility decrements for side effects in providing TIL-therapy due to toxicity	0.145	0.145	0.020	Beta	It was assumed to be the same as in the initial model because the availability of data on toxicity after a high or decrescendo dose scheme is limited.
PFS TIL	0.234	0.234	0.089	Beta	Assumed to be the same as no data shows that efficacy of TIL-therapy decreased with a lowered dose IL2.
OS TIL	0.412	0.412	0.098	Beta	
Scenario: "TIL production outsourced"					
TIL production costs	€35,500	€106,500	€11,990	Gamma	Since no commercial price is available, we made an assumption based on expert opinion (WvH and JvB) that commercial costs of TIL are at least 3 times higher. Taking into account the necessary logistical arrangements and general costs when starting a biotech company
Scenario: "Automatic TIL production"					
TIL production costs	€35,500	€24,850	€1,268	Gamma	Assumption: 30% decrease of production costs as described in the scenario.

Data analysis and visualization

The results of the scenarios incorporated in the cost-effectiveness model are expressed by the incremental cost effectiveness ratio (ICER), net monetary benefit (NMB) and the probability of TIL-therapy being cost-effective. The ICER is a deterministic statistic calculated by dividing the difference in costs by the difference in quality adjusted life uears (QALYs) for TIL-therapy and ipilimumab. An ICER, negative (less costly, more effective) or below a certain threshold (willingness to pay (WTP)), in this study €80,000, would mean that TIL-therapy is favored over ipilimumab. The WTP of €80,000 is the informal ceiling ratio in the Netherlands for diseases with the highest symptom burden[22]. As internationally different WTP thresholds are used, a second WTP threshold was used in evaluating the NMB: £30,000 (€34,821; April 2019), which is the WTP threshold used in the United Kingdom[23]. A two-way sensitivity analysis is a deterministic analysis evaluating the effect of various levels of two parameters on the ICER. We varied the 1-year progression free survival rate and the costs of TIL in a two-way sensitivity analysis.

Both NMB and probability of being cost-effective are probabilistic statistics in which uncertainty surrounding the input parameters is taken into account by randomly drawing parameter values from the parameter distributions, using Monte Carlo simulations with 1,000 iterations. The NMB was calculated using the WTP ratios and the following formula per iteration: (incremental QALYs x WTP)- incremental costs. A mean NMB ≥€0 indicates that TIL-therapy is cost-effective compared to ipilimumab, given the chosen threshold.

To calculate the probability of TIL-therapy being cost-effective, the NMB was calculated over different thresholds, ranging from €0 to €80,000 in steps of €1,000. A NMB value of ≥€0 is cost-effective, which is indicated with 1, a NMB value <€0 is not cost-effective, which is indicated with 0. This was done for all the iterations in the Monte Carlo simulation per threshold. A mean of this binary value was calculated per threshold which shows the probability of TIL being cost-effective compared to ipilimumab at that threshold. Finally, the mean of these average probability scores gives the probability of TIL-therapy being cost-effective in a WTP range of €0 to €80,000.

RESULTS

Characteristics of the respondents

Twenty-nine respondents, mainly clinicians (76%; 24% other), completed the web-based questionnaire between January and October 2017. The majority of respondents originated

from the Netherlands (n=14), fifteen experts originated from other countries, namely Belgium, Denmark, Germany, Israel, Italy, Poland, Portugal, Spain, UK, US, and Australia. Most respondents described themselves as familiar (52%), expert (28%) or a former expert (10%) with TIL-therapy and had on average 2.7 years of experience with TIL treatment. Table 3 shows the characteristics of the respondents.

Table 3 - Characteristics of the experts that participated in the scenario drafting questionnaire (n=29).

Number of respondents	29 (100%)
Function	
Medical oncologist	22 (76%)
Director	3 (10%)
Head cell production	1 (3%)
Consultant	1 (3%)
Clinical and translational research	2 (7%)
Mean experience with melanoma, years (range)	16.38 (1-35)
Mean experience with TIL-therapy, years (range)	2.72 (0-20)
Level of familiarity with TIL-therapy	
Unfamiliar	0 (0%)
Accidentally familiar	3 (10%)
Familiar	15 (52%)
Former expert	3 (10%)
Expert	8 (28%)
Employed in	
Australia	1 (3%)
Belgium	1 (3%)
Denmark	2 (7%)
Germany	3 (10%)
Israel	1 (3%)
Italy	1 (3%)
Netherlands	14 (48%)
Poland	1 (3%)
Portugal	1 (3%)
Spain	1 (3%)
UK	1 (3%)
US	1 (3%)
N/A	1 (3%)

Likelihood of the scenarios

The mean and median likelihood of each of the scenarios is presented in Table 4 and Figure 2. A large variability was seen in the expected likelihood of the scenarios suggesting that respondents are uncertain about the future developments surrounding TIL-therapy (Figure 2). On average, most of the scenarios scored a likelihood of around 50% (46%-54%). Two scenarios scored a likelihood of ≥55%: "Combination therapy" (57%) and "Automatic TIL production" (58%). Four scenarios were thought to be less likely: "Biomarker" (37%), "TCR therapy" (32%), "Low-cost competition" (30%) and "Less IL-2 treatment" (36%). Finally, the likelihood of the "Base case" in the coming five years was estimated at 54%. The results of the questions related to the adoption of TIL are listed in Supplement 6.

Table 4 - The mean and median likelihood of each scenario.

	Mean likelihood (median)		
	All respondents (n=29)	**Only familiar and experts (n=23)**	**≥ 1 year experience (n=10)**
BASE CASE SCENARIO			
"Base case"	54% (50%)	52 (45%)	55% (55%)
"WHAT IF" SCENARIOS			
"Competition"	46% (50%)	48% (50%)	43% (30%)
"TIL more effective"	52% (50%)	52% (50%)	52% (50%)
"Biomarker"	37% (35%)	38% (35%)	40% (35%)
"TCR therapy"	32% (30%)	29% (25%)	23% (20%)
"Patients unconvinced"	53% (60%)	54% (63%)	45% (50%)
"2nd line treatment"	53% (50%)	54% (50%)	54% (50%)
"3rd line treatment"	54% (50%)	57% (50%)	67% (68%)
"Combination therapy"	57% (63%)	57% (60%)	57% (60%)
"Clinicians unconvinced"	51% (50%)	52% (50%)	51% (50%)
"Low cost competition"	29% (20%)	29% (15%)	28% (23%)
"Influence by companies"	50% (58%)	52% (55%)	49% (55%)
"Less IL2 treatment"	36% (50%)	39% (50%)	41% (50%)
"TIL production outsourced"	53% (50%)	52% (50%)	44% (45%)
"Automatic TIL production"	58% (63%)	57% (60%)	62% (70%)

The first column shows the likelihood by all respondents, the second column shows the likelihood judged by the respondents that judged themselves as expert and familiar and the third column shows the respondents having ≥1 year experience with TIL-therapy. The scenarios displayed in bold were labelled as "likely" based on the evaluated likelihood (≥55% in one of these columns) (Figure 2).

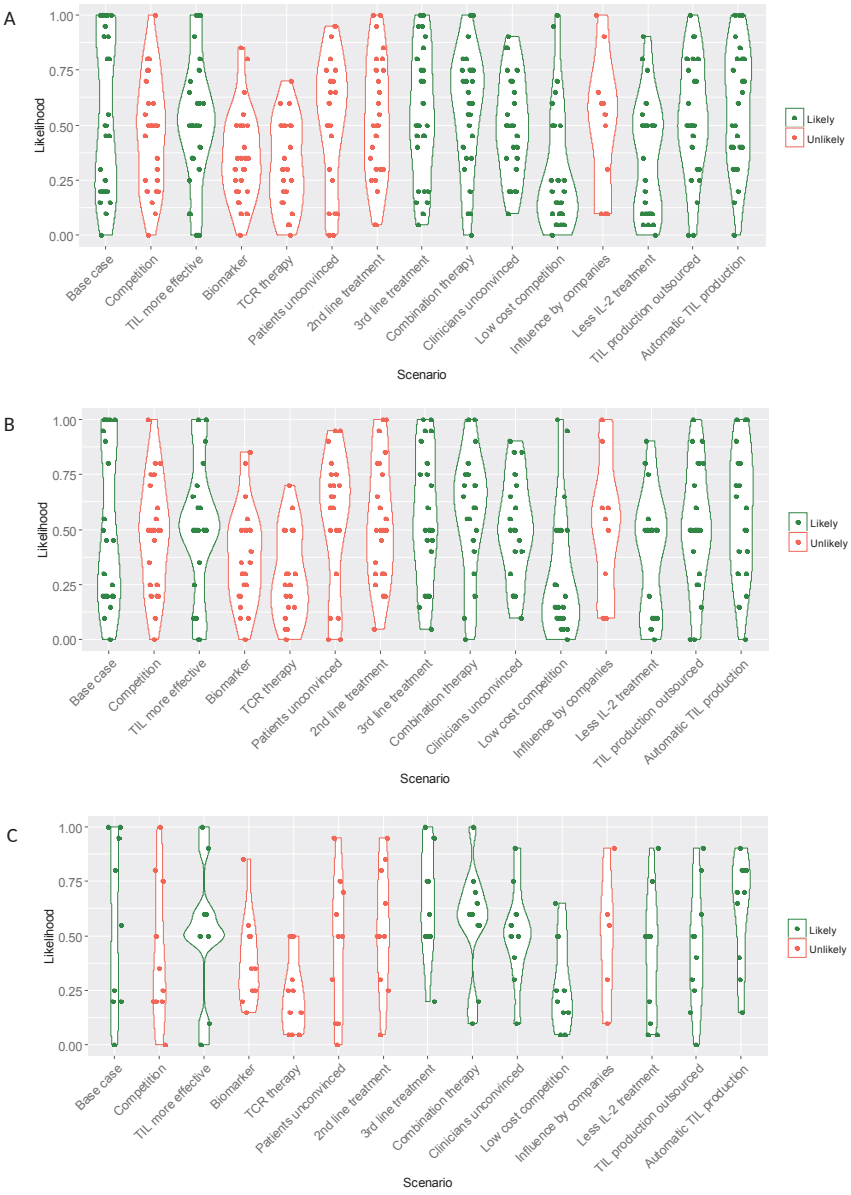


Figure 2. Likelihood of scenarios. These violin plots show all observations from the survey in points. In addition, it shows the distribution of the likelihood per scenario by making the graph wider or smaller. When a number of observations are seen at the same likelihood percentage, the plot becomes wider. Figure A shows the estimated likelihood of the future scenarios by all respondents (n=29), figure B shows the estimated likelihood by only the respondents that evaluated themselves as expert or familiar (n=23), figure C shows the estimated likelihood by only the respondents with ≥1 year of experience with TIL-therapy (n=10). The colors green ("likely") and red ("unlikely") correspond to the final label of the scenarios that followed from the steps shown in Figure 2 and according the reasons stated in Supplement 5.

Selected scenarios for incorporation in cost-effectiveness analysis

Using the cut-off value of ≥55%, "Combination therapy" and "Automatic TIL production" were labeled as "likely". Using the stratified results based on level of expertise, "Base case" and "3rd line treatment" were also labeled as "likely" (Table 4). The expert panel verified the likelihood of those four scenarios. Based on the literature review (step four in Supplement 2), "TIL more effective" and "Less IL-2 treatment" were labeled as "likely" as several studies described potential opportunities to increase the effectiveness of TIL-therapy and studies are investigating an IL-2 decreasing dose scheme[18] in order to lower the intensity of the treatment. The other scenarios or topics were not described in the recent literature review. The experts evaluated (step five) "Clinicians unconvinced", "Low-cost competition", "TIL production outsourced" as "likely" and the scenarios "Competition", "Biomarker", "TCR therapy", "Patients unconvinced" and "Influence by companies" were labelled as "unlikely". No scenario was solely labelled as "unlikely" based on the score from the survey. The arguments for labelling these scenarios as "likely" or "unlikely" are described in Supplement 5. As the base case scenario already evaluates the effect of using TIL-therapy as a second line therapy, the scenario: "2nd line treatment" was not incorporated in the cost-effectiveness analysis because it would show the same results. Eventually, scenarios resulting in no implementation of TIL-therapy e.g. "Patients unconvinced" and "Clinicians unconvinced", regardless of their likelihood, were not incorporated in the cost-effectiveness model as this results in an analysis comparing ipilimumab to ipilimumab.

Additionally, the potential combinations of scenarios were drafted and incorporated in the cost-effectiveness model. Three combinations were made related to research developments including "TIL more effective", "Automatic TIL production" and "TIL production outsourced" and three other combinations were defined related to the combination of TIL with other therapies, using "Combination therapy", "Automatic TIL production", "less IL-2" and "Low-priced competition".

Cost-effectiveness analysis

Figure 3 and 4 show the NMB and probability of TIL-therapy being cost-effective. Four out of six adoption scenarios showed a positive NMB: "TIL more effective", "Low-cost competition", "Less IL-2 treatment" and "Automatic TIL production", and a high probability of being cost-effective. Even when the total costs of the comparator (ipilimumab) are reduced with 20%, TIL-therapy had a 55% chance to be cost-effective ("Low-priced competition"). In contrast, "Combination therapy" showed a negative NMB with an ICER of €151,520 per QALY based on the first clinical results[20], and when the production of TILs is outsourced, TIL-therapy had a 0% likelihood to become cost-effective ("TIL production outsourced"). All the results from the

cost-effectiveness analysis are presented per scenario in Supplement 4. Figure 5 shows the results of the two-way sensitivity analysis and the incorporated scenarios. This graph shows for instance that the effectiveness should improve substantially when TIL production is being outsourced or TIL-therapy is combined with another therapy.

The combination of "TIL more effective" and "Automatic TIL production" showed a positive NMB as it combined the two most favourable scenarios for TIL-therapy (more effective and less expensive). The other two combinations related to research developments showed that a slight improvement for TIL-therapy in response rates does not outweigh the extra costs when TIL production is commercialized (0% probability of TIL being cost-effective), which holds true when TIL-therapy becomes more automatic (11% probability of TIL being cost-effective). The combinations that focused on the combination of TIL-therapy with a different therapy, showed a negative NMB in all combinations of scenarios. Yet, especially a reduction of the costs for TIL-therapy seem to have the highest impact on the probability that a combination therapy could become cost-effective (from 12% in base case to 28% in the combination of "combination therapy" and "TIL production outsourced".

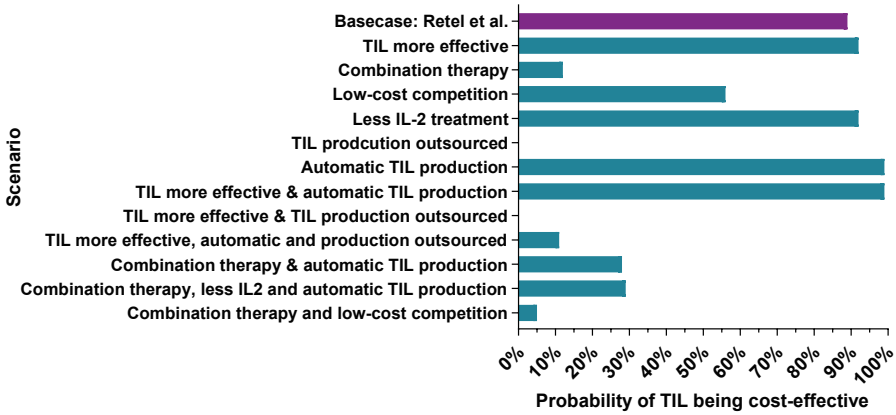


Figure 3. The probability of a scenario being cost-effective. Shows the probability of the different scenarios and the combinations of scenarios to become cost-effective when using a WTP threshold range of €0 to €80,000.

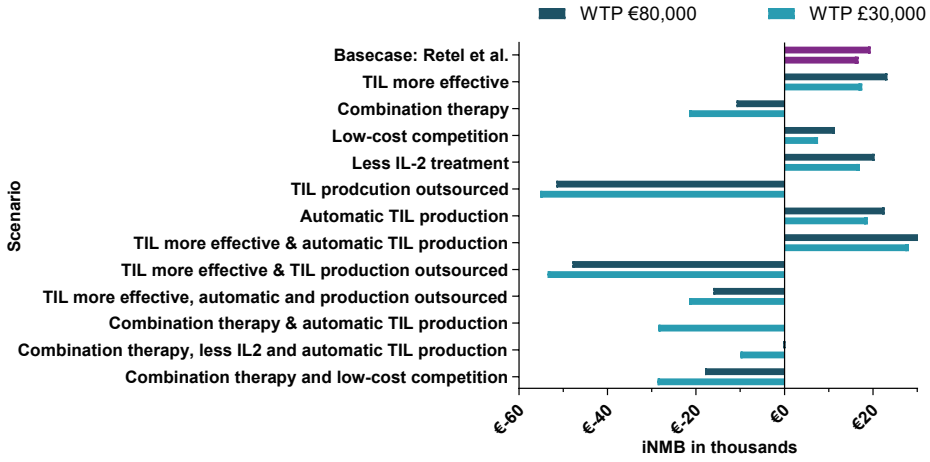


Figure 4. The incremental net monetary benefit (iNMB). Shows the iNMB for both the Dutch informal WTP threshold of €80,000 and for the WTP threshold that is mainly used in the United Kingdom of £30,000 (€34,821). A mean NMB ≥€0 indicates that TIL-therapy is cost-effective compared to ipilimumab given the chosen threshold.

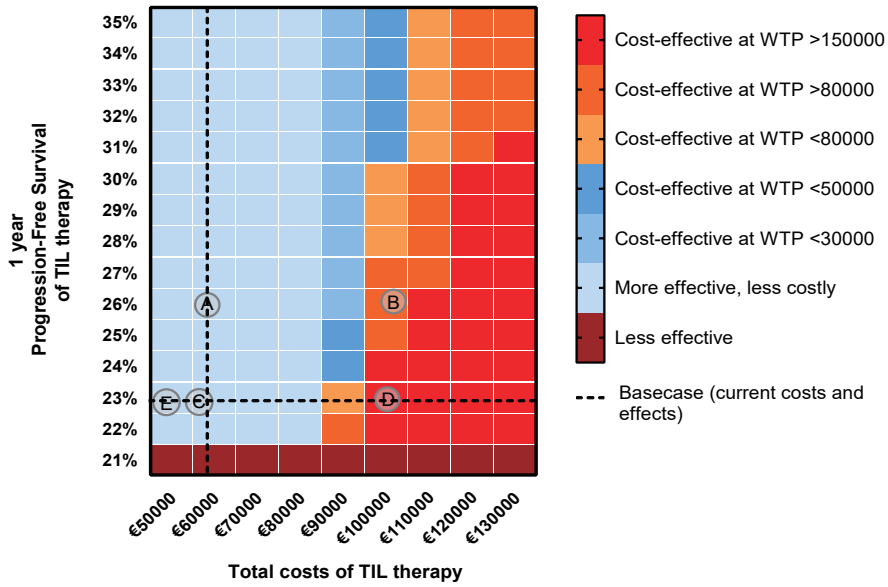


Figure 5. Two way sensitivity analysis with visualization of the incorporated scenarios. This cross table shows the levels of cost-effectiveness at different willingness to pay levels of TIL-therapy compared to ipilimumab when the Progression Free Survival (PFS) rate after 1 year changes and the costs of TIL vary. The dotted line represents the base case analysis. The incorporated scenarios are represented by letters. A = "TIL more effective", B = "Combination therapy", C = "Less IL2 treatment", D= "TIL production outsourced", E = "Automatic TIL production". The scenario "low-cost competition" was not possible to present in this graph because it affects the costs of ipilimumab instead of the costs of TIL-therapy. The colors do not always correspond with the results in Figure 3 and 4 because we evaluated the rounded numbers of costs and PFS rate.

DISCUSSION

Although a number of aspects concerning TIL-therapy are uncertain, our results show that TIL-therapy remains a promising addition to the treatment landscape of advanced melanoma as most of the "likely" scenarios resulted in TIL-therapy being cost-effective. One should, however, keep in mind that these results were based on the safety and efficacy results that are currently available (phase I/II trials)[4–7]. The ongoing RCT conducted at NKI and Herlev Hospitial (NCT02278887) is expected to bring the evidence needed to decide on its therapeutic position and adopt TIL-therapy as a standard treatment option in advanced melanoma.

Implications for clinical practice

The results of the cost-effectiveness analysis showed in most of the preferred scenarios a high probability for TIL-therapy to become cost-effective (55%-99%) and they identify aspects that could facilitate wide adoption of TIL-therapy. For example, as the scenario "Automatic TIL production" showed the highest probability for TIL-therapy to become cost-effective (99%), Research and Development focusing on optimizing the production process could facilitate the implementation and adoption of TIL-therapy. Especially as upscaling of the production process is seen as a barrier in clinical implementation of ATMPs[11,24]. Thus, automatizing of TIL production will mainly be a facilitator when the production of TILs stays in the academic setting. In contrast, the scenarios showing a reduced chance for TIL-therapy to become cost-effective identify crucial contextual factors that should be considered when deciding to adopt TIL-therapy. For example, outsourcing of the production of TILs may at first be expected to overcome known ATMP barriers as: (1) inadequate financial support for the required investments, and manufacturing costs; (2) a lack of regulatory knowledge, (3) challenging to upscale the production, and (4) to comply with Good Manufacturing Practices[9,24–27]. However, as a result of commercial pricing levels, assuming that the costs will be at least 3 times as high, this scenario resulted in a 0% probability for TIL-therapy to become cost-effective. Following our analysis, assuming a WTP threshold of €80,000, the production costs of TIL may only increase 1.5 times (~€53,000) to be cost-effective compared to ipilimumab. Within this scenario it should be kept in mind that the estimation of the commercial costs are uncertain. Especially because our assumption was based on the manufacturing costs in an academic setting and literature showed that commercial prices are mostly linked to what a society would pay instead of its expected added value or of the actual manufacturing costs[28]. However, although the estimation is uncertain the conclusion related to commercialization remains the same: Outsourcing could facilitate the implementation of TIL-therapy, however pricing agreements should be made with the commercial party to ensure cost levels that remain within the willingness to pay range of cost effectiveness. In the US setting, TIL can be

expected to be licensed and filed for FDA approval within the coming years, these insights are especially of interest to guide reimbursement decisions by insurance companies. An interesting scenario is the "Combination therapy" which showed a 12% (ICER of €151,520) probability of being cost-effective of, revealing that in this case either the treatment costs should decrease or the efficacy has to improve considerably. By automatizing the production of TILs, in our model the probability of being cost-effective increased only to 28%. Therefore, when a combination of TIL-therapy and for instance a certain checkpoint inhibitor seems promising, agreements on pricing with pharmaceutical companies for the combination therapy are necessary to remain within cost effectiveness ranges.

The results on the questions related to future developments suggest that adoption of TIL-therapy may be hampered by the attitude of patients and clinicians. First of all, clinicians nowadays seem unconvinced to apply TIL-therapy because of its perceived complexity and treatment intensity unless the therapy shows a 1-year survival rate of at least 61.3% (CI:55.2-67.5). Secondly, only a small proportion of the eligible patients seemed to be aware of TIL-therapy as a treatment option. As the attitude of stakeholders and especially clinicians, is a known barrier for implementation of ATMPs, a pro-active information strategy in anticipation of this attitude is crucial when deciding to diffuse TIL-therapy[25].

Comparison of our findings with current literature in context of an ATMP

Another barrier that ATMPs face in the translational pathway is the rapid evolving field of immunotherapy[8,9]. We therefore compared our results with the most recent developments described in literature and most of the "likely" scenarios still seem to be in line. For example, several trials investigate a combination of TIL-therapy and other targeted therapies: pretreatment with ipilimumab followed by TIL and IL-2 (NCT01701674)[20] or pretreatment with vemurafenib followed by TIL and IL-2 and followed by vemurafenib (NCT01659151)[29]. Besides, several trials investigate or investigated the effectiveness of a lower dose of IL-2 treatment[18,30] (NCT02354690), and finally research groups evaluate the optimal process of producing TILs, aiming to improve the efficacy of TIL-therapy e.g. by enriching T cell products with neo-antigens[18,31].

Some developments found in literature, however, were not incorporated. For instance, several studies evaluate different lymphodepleting preparative regimens such as total body irradiation (TBI) in combination with chemotherapy[18,32]. It is currently unclear whether such a regimen would be applied in the near future, but this scenario could influence the cost-effectiveness as TBI (requiring autologous PSC support) would significantly increase the costs. In addition, a very likely scenario that is not incorporated in this analysis, is the use of TIL-therapy in other tumors such as renal cell cancer, ovarian cancer and colorectal

cancer[33–36] (NCT01174121). This scenario should be kept in mind as it may facilitate the adoption of TIL by positively influence the clinicians attitude as both clinical experience and exposure increase and production costs may decrease. Furthermore, a recent literature review highlighted several potential agents (e.g. TIM3, GITR, OX40) that could be promising in treating advanced melanoma in the future[37]. Those agents are currently subject of the first phase I and II studies to evaluate their safety and efficacy[37]. Therefore, our study might have underestimated the likelihood of the competition scenario. However, available data on the efficacy and possible costs of those alternatives is to preliminary to incorporate these results in the cost-effectiveness analysis. Clearly, when these agents are proven to be safe, effective and more effective compared to TIL-therapy, those new treatments could hamper the adoption of TIL-therapy.

Observations from the scenario method and future directions

A wide range in the expected likelihood of the scenarios was identified (Figure 2), which challenged the labelling process for likely and unlikely scenarios. This may be explained by several factors. First, when a respondent is not (yet) involved in the TIL-therapy process, it is harder to have an opinion on the likelihood of these scenarios as theoretical models describe that some extent of experience is needed to evaluate the future adoption process[38]. Second, faced with uncertainty, respondents could be hesitant in choosing extreme options such as 0% and 100% likelihood. Finally, it is likely that the expected timing of these scenarios, if they are likely, differ across countries and hospitals as the adoption process and attitude towards TIL differs per site. The respondents originated from 12 different countries which could thus explain some of the wide ranges, as in one country a scenario may be likely in the coming five years (e.g. commercialization in the US) and in another country not at all.

Furthermore, we are aware that the scenarios labelled as “unlikely” and therefore not incorporated in the cost-effectiveness analysis, could still play a role in the adoption process of TIL-therapy (e.g. biomarker development, possible dominance of T-cell receptor (TCR) gene therapy over TIL-therapy, influence by companies and competition). These factors should not be neglected and it would be valuable to incorporate these in future decision-making processes.

Additionally, as the chance on the development of other types of cancer by using TIL-therapy was thought to be 6.4% (CI:4.5%-8.3%; Supplement 6) on average, clinical studies having a longer follow-up time than the current observational studies should evaluate the actual risk. When the risk is shown to be evident, it should be ethically discussed whether TIL treatment may still be preferred over ipilimumab. Finally, based on the currently available clinical evidence, data are lacking for one of the most likely scenarios, “Third-line treatment”, to

evaluate the relative cost-effectiveness of TIL-therapy. When estimating the expected costs of palliative chemotherapy we can estimate the incremental QALYs needed to become cost-effective at a certain willingness to pay threshold. When estimating the costs of on average 3.5 doses of chemotherapy (dacarbazine)[39] at €17,102 based on a three-weekly dosage of 200mg/m^2 for 5 days[21,40] compared to the costs of TIL-therapy, the difference in QALY's should be at least 0.561. This means that, TIL-therapy has to show a substantial gain in survival and/or quality of life or a reduction in follow-up costs to become cost-effective in third line. Such a calculation is informative but in order to inform decision makers on the effects of this –per likely scenario, clinical outcomes after progression on both PD-1 inhibitors and CTL-4 antibodies based on e.g. clinical registries should be obtained. Next we should compare these to clinical outcomes of TIL-therapy in patients that progressed on multiple treatment strategies, such as reported in the study of Sarnaik and colleagues[41].

Strengths and limitations

The main strength of this study is that we systematically drafted future scenarios (qualitatively) with internal and external experts, using a Constructive Technology Assessment framework[11,42] and using multiple Delphi rounds. This provides a comprehensive insight in the potential future developments that could influence TIL adoption and provides research and development teams with valuable information to anticipate on possible future developments. Since the landscape of immunotherapy for melanoma is continuously developing, the expectations of the experts were compared to the most recent literature reviews and ongoing clinical trials to select the "likely" scenarios and discuss our results. The main limitation is obviously, that the scenarios may not even keep up with actual developments. Other limitations are related to the early nature of this analysis. For example, to simulate the combination therapy, the input for the model was based on a first observational study in which only 13 patients were enrolled that received the combination of TIL-therapy and ipilimumab[20]. Additionally, the chosen cut-off value of ≥55% to evaluate scenarios as "likely" could be questioned due to the high uncertainty surrounding the likelihood scores. However, since the expert opinions and recent literature verified that the "likely" scenarios based on the cut-off value were "likely", a different cut-off value is not expected to have altered our conclusions. Furthermore, the cost-effectiveness analyses were conducted from a Dutch perspective similar to the original model by Retèl et al. 2018. The costs for both TIL-therapy and ipilimumab are expected to differ between countries[43] which limits the generalizability of the cost-effectiveness results to different settings. The generalizability may also be limited by fact that mainly experts from European countries completed the questionnaire. However, by verifying the likelihood results with the most recent literature, the identified crucial contextual factors are expected to hold true also in other countries because similar (financial) challenges are expected regarding e.g. outsourcing and providing a combination of therapies.

Conclusion

The results of our scenario study can support the implementation and adoption process of TIL-therapy as they identified crucial contextual factors that require anticipation and identified potential facilitators. As implementation of TIL-therapy is complex and could be time consuming, clinicians and/or other decision makers may decide to adapt the implementation process to possible developments in an early stage to anticipate and grant timely patient access when TIL-therapy shows to be effective.

ACKNOWLEDGEMENTS

The authors want to thank all the experts that participated in the web-based questionnaire.

REFERENCES

1. Garbe, C., Eigentler, T. K., Keilholz, U., Hauschild, A. & Kirkwood, J. M. Systematic Review of Medical Treatment in Melanoma: Current Status and Future Prospects. *Oncologist* **16**, 5–24 (2011).
2. Larkin, J. *et al.* Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **381**, 1535–1546 (2019).
3. Rosenberg, S. A. *et al.* Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N. Engl. J. Med.* **319**, 1676–1680 (1988).
4. Wu, R. *et al.* Adoptive T-Cell Therapy Using Autologous Tumor-Infiltrating Lymphocytes for Metastatic Melanoma. *Cancer J.* **18**, 160–175 (2012).
5. Andersen, R. *et al.* Long-Lasting complete responses in patients with metastatic melanoma after adoptive cell therapy with tumor-infiltrating lymphocytes and an attenuated il2 regimen. *Clin. Cancer Res.* **22**, 3734–3745 (2016).
6. Rosenberg, S. A. *et al.* Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clin. Cancer Res.* **17**, 4550–4557 (2011).
7. Merhavi-Shoham, E., Itzhaki, O., Markel, G., Schachter, J. & Besser, M. J. Adoptive Cell Therapy for Metastatic Melanoma. *Cancer J.* **23**, 48–53 (2017).
8. Hartmann-Fritsch, F., Marino, D. & Reichmann, E. About ATMPs, SOPs and GMP: The Hurdles to Produce Novel Skin Grafts for Clinical Use. *Transfus. Med. Hemother.* **43**, 344–352 (2016).
9. de Wilde, S. *et al.* Hurdles in clinical implementation of academic advanced therapy medicinal products: A national evaluation. *Cytotherapy* **18**, 797–805 (2016).
10. Zorginstituut Nederland (Dutch institute of healthcare). Conditional approval for treatment with TILs in stage IIIc and IV advanced melanoma in the Netherlands. (2015). Available at: https://www.zorginstituutnederland.nl/werkagenda/kanker/voorwaardelijke-toelating-behandeling-met-tumor-infiltrerende-lymfocyten-til-van-uitgezaaid-melanoom-irresectabel-stadium-iiic-en-stadium-iv. (Accessed: 4th September 2019)
11. Lindenberg, M. A. *et al.* Treatment With Tumor-infiltrating Lymphocytes in Advanced Melanoma: Evaluation of Early Clinical Implementation of an Advanced Therapy Medicinal Product. *J. Immunother.* **41**, 413–425 (2018).
12. Retel, V. P. *et al.* Early cost-effectiveness of tumor infiltrating lymphocytes (TIL) for second line treatment in advanced melanoma: a model-based economic evaluation. *BMC Cancer* **18**, 895 (2018).
13. Shell international BV. *Scenarios: An Explorer's Guide*. (2008).
14. Enserink, B. & Hermans, L. *Policy Analysis of Multi-Actor Systems chapter 5: Exploring the Future*. (Eleven International Publishing, 2010).
15. Shell international BV. *Scenarios: An Explorer 's Guide*. (2008).
16. Lipinski, A. & Loveridge, D. *Multiple Perspective: Concept, Application and User Guidelines. Systems Practice* **2**, (1982).
17. Joosten, S. E. P., Retèl, V. P., Coupé, V. M. H., van den Heuvel, M. M. & van Harten, W. H. Scenario drafting for early technology assessment of next generation sequencing in clinical oncology. *BMC Cancer* **16**, 66 (2016).
18. Rohaan, M. W., van den Berg, J. H., Kvistborg, P. & Haanen, J. B. A. G. Adoptive transfer of tumor-infiltrating lymphocytes in melanoma: a viable treatment option. *J. Immunother. Cancer* **6**, 102 (2018).
19. Wilson, M. A. & Schuchter, L. M. Chemotherapy for Melanoma. in *Melanoma* (eds. Kaufman, H. L. & Mehnert, J. M.) 209–229 (Springer International Publishing, 2016). doi:10.1007/978-3-319-22539-5_8
20. Mullinax, J. E. *et al.* Combination of Ipilimumab and Adoptive Cell Therapy with Tumor-Infiltrating Lymphocytes for Patients with Metastatic Melanoma. *Front. Oncol.* **8**, (2018).
21. Zorginstituut Nederland (Dutch institute of healthcare). Medicijnkosten (costs of pharmaceuticals). (2019). Available at: http://www.medicijnkosten.nl/. (Accessed: 2nd April 2020)
22. Versteegh, M. M. *et al.* Severity-Adjusted Probability of Being Cost Effective. *Pharmacoeconomics* **37**, 1155–1163 (2019).
23. McCabe, C., Claxton, K. & Culyer, A. J. The NICE Cost-Effectiveness Threshold. *Pharmacoeconomics* **26**, 733–744 (2008).
24. Ali, R., Hollander, A., Kemp, P., Webster, A. & Wilkins, M. Regulating cell-based regenerative medicine: the challenges ahead. *Regen. med.* **9**, 81–87 (2014).

25. Abou-El-Enein, M., Elsanhoury, A. & Reinke, P. Overcoming Challenges Facing Advanced Therapies in the EU Market. *Cell Stem Cell* **19**, 293–297 (2016).
26. Galli, M. C. ATMPs for Cancer Immunotherapy: A Regulatory Overview. in *Methods in Molecular Biology* **1393**, 1–9 (2016).
27. Gardner, J., Faulkner, A., Mahalatchimy, A. & Webster, A. Are there specific translational challenges in regenerative medicine? Lessons from other fields. *Regen. Med.* **10**, 885–895 (2015).
28. Morgan, S. G., Bathula, H. S. & Moon, S. Pricing of pharmaceuticals is becoming a major challenge for health systems. *BMJ* **368**, l4627 (2020).
29. Deniger, D. C. *et al.* A Pilot Trial of the Combination of Vemurafenib with Adoptive Cell Therapy in Patients with Metastatic Melanoma. *Clin. Cancer Res.* **23**, 351–362 (2017).
30. Nguyen, L. T. *et al.* Phase II clinical trial of adoptive cell therapy for patients with metastatic melanoma with autologous tumor-infiltrating lymphocytes and low-dose interleukin-2. *Cancer Immunol. Immunother.* **68**, 773–785 (2019).
31. Tran, E. *et al.* Cancer Immunotherapy Based on Mutation-Specific CD4+ T Cells in a Patient with Epithelial Cancer. *Science (80-.).* **344**, 641–645 (2014).
32. Dudley, M. E. *et al.* Adoptive Cell Therapy for Patients With Metastatic Melanoma: Evaluation of Intensive Myeloablative Chemoradiation Preparative Regimens. *J. Clin. Oncol.* **26**, 5233–5239 (2008).
33. Lee, H. J. *et al.* Expansion of tumor-infiltrating lymphocytes and their potential for application as adoptive cell transfer therapy in human breast cancer. *Oncotarget* **8**, 113345–113359 (2017).
34. Baldan, V., Griffiths, R., Hawkins, R. E. & Gilham, D. E. Efficient and reproducible generation of tumour-infiltrating lymphocytes for renal cell carcinoma. *Br. J. Cancer* **112**, 1510 (2015).
35. Andersen, R. *et al.* Tumor infiltrating lymphocyte therapy for ovarian cancer and renal cell carcinoma. *Hum. Vaccin. Immunother.* **11**, 2790–2795 (2015).
36. Mayor, P., Starbuck, K. & Zsiros, E. Adoptive cell transfer using autologous tumor infiltrating lymphocytes in gynecologic malignancies. *Gynecol. Oncol.* **150**, 361–369 (2018).
37. Luther, C., Swami, U., Zhang, J., Milhem, M. & Zakharia, Y. Advanced stage melanoma therapies: Detailing the present and exploring the future. *Crit. Rev. Oncol. Hematol.* **133**, 99–111 (2019).
38. Rogers, E. M. The Diffusion of Innovations. *Diffus. Innov.* 576 (2003).
39. Middleton, M. R. *et al.* Randomized Phase III Study of Temozolomide Versus Dacarbazine in the Treatment of Patients With Advanced Metastatic Malignant Melanoma. *J. Clin. Oncol.* **18**, 158–158 (2000).
40. Dutch Healthcare Authority (NZa). DBC-zorgproduct 029499025 (DRG open tariff). *2018* Available at: https://www.opendisdata.nl/msz/zorgproduct/029499025?
41. Sarnaik, A. *et al.* Safety and efficacy of cryopreserved autologous tumor infiltrating lymphocyte therapy (LN-144, lifileucel) in advanced metastatic melanoma patients who progressed on multiple prior therapies including anti-PD-1. *J. Clin. Oncol.* **37**, 2518–2518 (2019).
42. Douma, K. F. L., Karsenberg, K., Hummel, M. J. M., Bueno-de-Mesquita, J. M. & van Harten, W. H. Methodology of constructive technology assessment in health care. *Int. J. Technol. Assess. Health Care* **23**, 162–168 (2007).
43. van Harten, W. H., Wind, A., de Paoli, P., Saghatchian, M. & Oberst, S. Actual costs of cancer drugs in 15 European countries. *Lancet Oncol.* **17**, 18–20 (2016).
44. Besser, M. J. *et al.* Clinical Responses in a Phase II Study Using Adoptive Transfer of Short-term Cultured Tumor Infiltration Lymphocytes in Metastatic Melanoma Patients. *Clin. Cancer Res.* **16**, 2646–2655 (2010).
45. Radvanyi, L. G. *et al.* Specific Lymphocyte Subsets Predict Response to Adoptive Cell Therapy Using Expanded Autologous Tumor-Infiltrating Lymphocytes in Metastatic Melanoma Patients. *Clin. Cancer Res.* **18**, 6758–6770 (2012).
46. Wolf, B. *et al.* Safety and Tolerability of Adoptive Cell Therapy in Cancer. *Drug Saf.* **42**, 315–334 (2019).

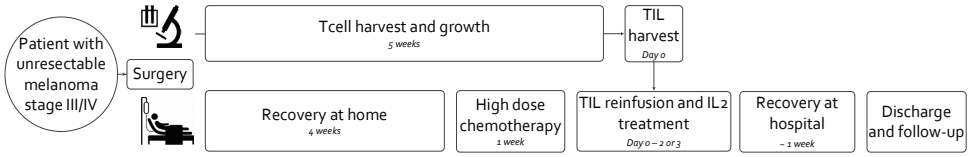
SUPPLEMENTARY MATERIAL

Supplement 1 - Questionnaire

Welcome to this short questionnaire in which we try to explore the future of the tumor infiltrating lymphocytes (TIL) treatment in advanced melanoma.

"An exploration of the future, is not to predict the future but to explore plausible futures and become aware of the uncertainties."

Scenario drafting - the way we explore the future - is performed in order to identify plausible barriers and facilitators and to estimate the full potential of this treatment by means of a cost-effectiveness analysis. Filling in this questionnaire will take approximately 15 minutes.
TIL, as you may know, is a personalized treatment for IIIC and IV stage melanoma in which at least a resectable lesion of 2-3 cm is available. In this treatment a tumor is resected and will be transported to a production facility. In this lab, T cells will be harvested and will grow approximately 4-5 weeks until a number of a couple billion cells is reached. After this period of growth, the patient will receive the infusion product of TILs. In addition interleukin-2 is given to stimulate the TILs inside the body. Below a schematic display of the procedure.



This questionnaire includes several questions and scenarios. For these scenarios we are interested in your estimation of the likelihood for this scenario to happen within 5 years.

Thank you for your time and effort!
With kind regards,
Melanie Lindenberg (m.lindenberg@nki.nl)
(on behalf of Prof. J. Haanen MD and Prof. W.H. van Harten MD)

Part 1

1. What is your function (work)? (open question)
2. To what company / hospital / institute are you affiliated to? (open question)
3. How many years of experience do you have with melanoma? (open question)
4. How many years of experience do you have with TIL? (open question)

5. To what extend do you find yourself familiar with the tumor infiltrating lymphocytes (TIL) treatment? (before the explanation that was given at the beginning of the questionnaire)

- ◯ Unfamiliar - When mentioning the subject, it does not recall any memories or it does not give lead to saying something sensible about it.
- ◯ Accidentally familiar - When you know what it is about, you have read something about it or you heard or saw something about it on the radio, television or other media.
- ◯ Familiar - When you know most arguments pro and contra the most controversial elements of the subject, when you have read a lot about it and when you have formed an opinion.
- ◯ Former expert - Used to be an expert on the subject some time ago, but your knowledge is

somewhat outdated because other activities came up. But you are still reasonably well-informed about recent developments, which provides you with a broad overview of the subject as opposed to deep detailed knowledge.

◯ Expert- When you belong to the small community of people who, at this moment, study, work on and are dedicated to this subject. You typically know who else works on this subject, you know the domestic literature and probably also the international literature about this subject, you go to conferences and seminars and when possible you publicize about the subject.

Part 2

Background information

Below, scenarios and questions are given. In these questions it is important to know the following definitions:

- Intervention of interest: TIL-therapy
- Standard of care: ipilimumab
- Competing therapies: combination of therapies (ipilimumab and nivolumab or pembrolizumab), solely pembrolizumab, solely nivolumab or potential treatments currently in the pipeline

Furthermore, when italic text is given, this is extra information to help filling in the likelihood of the scenario or the question.

Scoring scenarios

0% I think this is not likely at all

50% This might as well, will or will not happen within 5 years

100% I am sure that this will be the case in 5 years

Testcase

Scenario: In the western countries, more than half of the car users will be using an electrical or hybrid car.

How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)

TIL is part of a coverage with evidence development program in the Netherlands.

Scenario basecase 1

If TIL shows better survival rates (at least 10% improvement) compared to ipilimumab, TIL will be reimbursed.

How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)

◯ I don't know

TIL is part of a coverage with evidence development program in the Netherlands.

Scenario basecase 2

If TIL shows better survival rates (at least 10% improvement) compared to ipilimumab, TIL will be implemented in specialized melanoma centers.

How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)

◯ I don't know

Scenario effectiveness 1
Competing (immuno)therapies are equal in costs but 10% more effective compared to TIL.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

Question effectiveness 1
Effectiveness of TIL Is likely to improve due to research developments such as gene modification or selective TIL, currently the one year survival rate is: 45% - 60% (based on phase 1 and 2 trials).

What would be the minimal effectiveness of TIL leading to accept TIL as a standard therapy for you, expressed in one-year survival rate? (%)

☐

Scenario effectiveness 2
The effectiveness of TIL has increased with 10% (clinically relevant) due to research developments.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

Question effectiveness 2
What would be the risk of developing other types of cancer such as lymphomas by activating the immune system by injecting TILs? (in %)

☐

Question effectiveness 3
TIL is known as an intensive treatment, but compared to PD1 and CTL-4 inhibitors it may be argued that it is less intensive on longer term. First, the duration of TIL treatment is short compared to the other treatments: 3 weeks of hospital admission vs. 2-3 weekly sessions for a longer time period (ranging between 12 weeks and 1-2 years). Secondly, adverse events of TIL treatment are reversible and thus of a short duration, whereas competing therapies show substantial chances on high (long-term) toxicities (20%-40%).
In what extend do you agree with the following statement:
TIL treatment provides significantly better quality of life compared to ipilimumab.
◯ Strongly agree
◯ Agree
◯ Nor agree or disagree
◯ Disagree
◯ Strongly disagree

Currently a lot of research is focused on identifying factors that can predict a response on immunotherapy.

Scenario research 1
A biomarker, being able to select patients for TIL, is available.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

T cell receptor (TCR) gene therapies are currently in development in treating several types of cancer.

Scenario research 2
TCR therapy dominates TIL treatment in advanced melanoma, regardless other treatment modalities.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

Scenario patient perspective 1
Patients prefer the competing therapies over TIL based on complete information on toxicities and effectiveness.
How likely is this scenario in the coming 5 years? (0-100%) (slider)
◯ I don't know

Question patient perspective 1
In the RCT TIL vs ipilimumab, patient recruitment is going slow. We think this is partly because of limited awareness of patients for this treatment.
Could you estimate the percentage of the eligible patients (metastatic melanoma patient) you think is aware of TIL as a potential treatment? (in %)
[ ]

Literature suggests that TIL is still effective after treatment with PD1 inhibitors.

Scenario implementation 1
TIL is implemented as a second line treatment after anti PD1 inhibitors in metastatic melanoma.
How likely is this scenario for the coming 5 years? (0-100%) (slider)
◯ I don't know

Scenario implementation 2
TIL is implemented as a third line (last resort) treatment in metastatic melanoma.
How likely is this scenario for the coming 5 years? (0-100%) (slider)
◯ I don't know

Scenario implementation 3
TIL is used in combination with other immune or personalized therapies (i.e. nivolumab or vemurafenib).
How likely is this scenario for the coming 5 years? (0-100%) (slider)
◯ I don't know

Question implementation
If TIL proves to be effective, what would be the main reason for clinicians to be unconvinced of introducing TIL-therapy in the coming 5 years?
◯ Complexity of TIL (understanding of TIL growth and the clinical process)
◯ Intensiveness of IL2 and expected adverse events
◯ User-friendliness of TIL (practical issues in implementation)
◯ This is not the case: clinicians will treat patients with TIL
◯ Other namely:
[ ]

Scenario implementation 4
Clinicians are not willing to implement TIL because of one of the previous stated reasons.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

Scenario interactions with the pharmaceutical market 1
If TIL turns out to be cost-effective, pharmaceutical companies will lower the prices of competing immunotherapies.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

Scenario interactions with pharmaceutical market 2
Arrangements between pharmaceutical companies and hospitals and/or doctors, negatively affect patient selection for TIL-therapy.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

Scenario process changes 1
Additional interleukin-2 treatment after infusion of TIL is not be necessary anymore.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

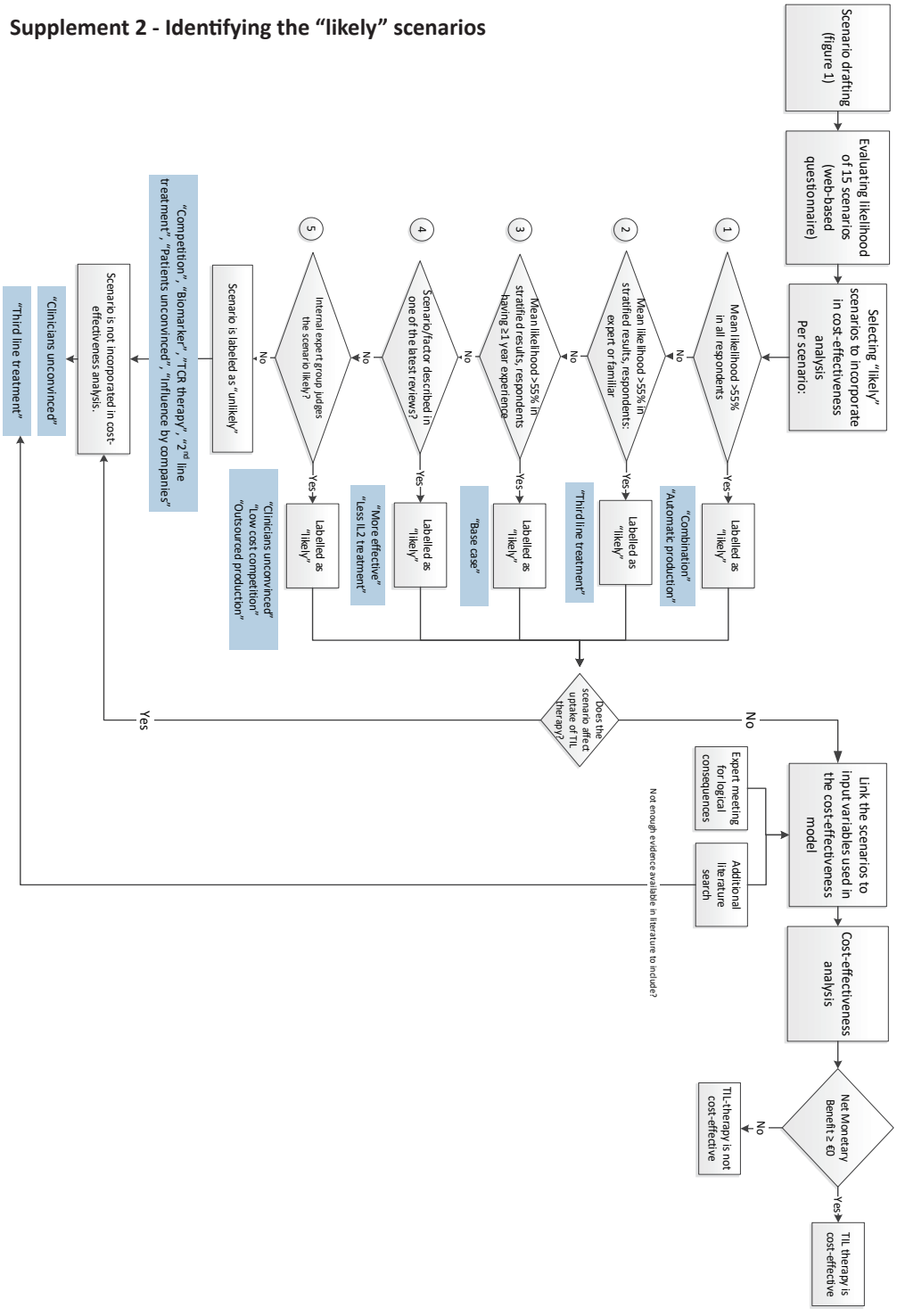
Scenario process changes 2
Production of TIL is of interest for the pharmaceutical market and is outsourced by a commercial company.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

Scenario process changes 3
Production of TIL is less expensive (30% reduction) due to more automatic process steps.
How likely is this scenario for the coming 5 years in percentages (0-100%)? (slider)
◯ I don't know

This is the end of the questionnaire, thank you for your time and effort!

With kind regards,
Melanie Lindenberg
m.lindenberg@nki.nl

Supplement 2 - Identifying the "likely" scenarios



Supplement 3 - Information on the input parameters for the base case analysis

The original analysis by Retèl et al. 2018 presents a full overview of the methods for the base case analysis[1]. In this supplement we provide data corresponding to this analysis.

Data on progression free and overall survival

The survival parameters were based on two studies describing both progression free and overall survival one year after TIL infusion, Besser et al. 2010 and Radvanyi et al. 2012[2,3]. Below the Kaplan Meier curves from the study by Radvanyi et al. 2012 are presented which are included as supplementary material. We did not adapted those graphs. The input for the pooled survival analysis using CMA Software, version 3, Biostat, US, was based on the two tables included in the manuscripts describing the months of progression free survival and overall survival.

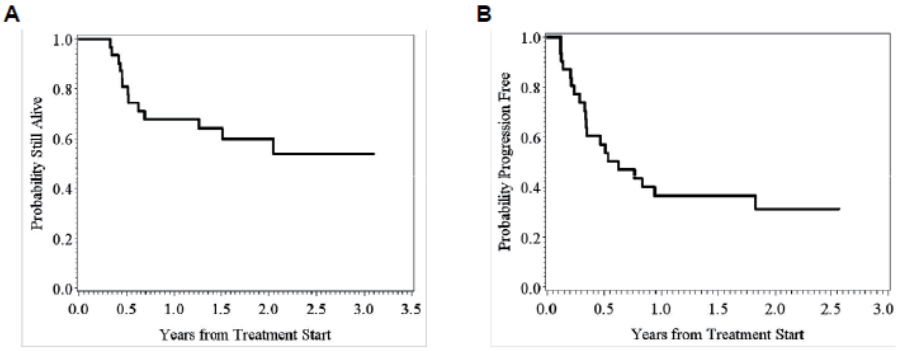


Figure S4. Kaplan-Meier curves of overall survival and progression-free survival for all 31 treated patients. Median overall survival was not reached (A), while median progression-free survival (B) for the cohort of patients was 7.6 months (95% CI: 4.1-22.2 months).

The input for the pooled analysis was as follows:

Study	Events	Sample size	Event rate at one year	Standard error
Progression free survival				
Besser 2010	3	20	0.150	0.626
Radvanyi 2012	10	31	0.323	0.384
Overall survival				
Besser 2010	7	20	0.350	0.469
Radvanyi 2012	20	31	0.645	0.375

This resulted in the following pooled results for OS and PFS after 1 year. These rates were transformed in transition probabilities assuming a linear distribution. Those are presented in the next table.

	Pooled event rate	Lower limit	Upper limit
PFS	0.266	0.160	0.408
OS	0.531	0.389	0.667

1. Retel VP, Steuten LMG, Geukes Foppen MH, Mewes JC, Lindenberg MA, Haanen JBAG, et al. Early cost-effectiveness of tumor infiltrating lymphocytes (TIL) for second line treatment in advanced melanoma: a model-based economic evaluation. BMC Cancer. 2018 Sep;18(1):895.
2. Besser MJ, Shapira-Frommer R, Treves AJ, Zippel D, Itzhaki O, Hershkovitz L, et al. Clinical responses in a phase II study using adoptive transfer of short-term cultured tumor infiltration lymphocytes in metastatic melanoma patients. Clin Cancer Res. 2010;
3. Radvanyi LG, Bernatchez C, Zhang M, Fox PS, Miller P, Chacon J, et al. Specific lymphocyte subsets predict response to adoptive cell therapy using expanded autologous tumor-infiltrating lymphocytes in metastatic melanoma patients. Clin Cancer Res. 2012;

All input parameters

This table shows the input parameters of the base case model. The table is copied from the original manuscript. The sources are listed below the table. (Supplement 2, Table 3)

Parameters	Mean	SE	Distribution	Source
Survival probabilities per year				
Ipilimumab				
PFS	0.175	0.012	Beta	[6]
OS	0.366	0.018	Beta	[6]
TIL				
PFS	0.234	0.089*	Beta	[10, 11]
OS	0.412	0.098*	Beta	[10, 11]
Utilities and side effects				
ipilimumab				
Stable disease	0.850	0.020	Beta	[18]
Progression	0.590	0.020	Beta	[18]
TIL				
Stable disease	0.850	0.020	Beta	[18]
Progression	0.590	0.020	Beta	[18]
Utility decrements				
Fatigue	0.090	0.020	Beta	[18]
Diarrhea	0.060	0.020	Beta	[18]
Colitis	0.130	0.020	Beta	[18]
Neutropenia	0.130	0.020	Beta	[18]
Dyspnea	0.100	0.020	Beta	[18]
Flu-like syndrome (grade I/II))	0.090	0.020	Beta	[18]
Anaemia	0.110	0.020	Beta	[18]
Likelihood of side effects				
ipilimumab				
Fatigue	0.070	0.015	Beta	[6]
Diarrhea	0.060	0.015	Beta	[6]
Colitis	0.060	0.015	Beta	[6]

(Continued on next page)

Supplement 2 Table 3 (continued)

Parameters	Mean	SE	Distribution	Souce
ipilimumab (continued)				
Dyspnea	0.040	0.015	Beta	[6]
Immune	0.100	0.015	Beta	[6]
Anaemia	0.030	0.015	Beta	[6]
TIL*				
Fatigue	0.001	0.001	Beta	[24]
Diarrhea	0.001	0.001	Beta	[24]
Neutropenia	0.560	0.100	Beta	[24]
Dyspnea	0.020	0.015	Beta	[24]
Immune	0.220	0.100	Beta	[24]
Anaemia	0.440	0.100	Beta	[24]
Failures, non-compliance TIL				
Failures	0.100	0.015	Beta	[20] Expert opinion
Non-compliance	0.100	0.015	Beta	[21]
Costs in euros				
Ipilimumab-total	91,487.50	+/-25%	Gamma	
Drug	90,100.00	+/-25%	Gamma	[22]
Administration	473.00	+/-25%	Gamma	[23]
Management of side effects	914.50	+/-25%	Gamma	[6,16]
TIL**	62,000.00	+/-25%	Gamma	NKI-AVL
Interleukin treatment within the TIL-therapy regimen	2,130.24	+/-25%	Gamma	NKI-AVL and [22]
Follow-up costs stable disease***	516.00	+/-25%	Gamma	[25]
Costs progressive disease****	9,125.00	+/-25%	Gamma	[31]
Side effects				
Fatigue	198.00	+/-25%	Gamma	[16]
Diarrhea	580.00	+/-25%	Gamma	[16]
Colitis/neutropenia*****	1115.00	+/-25%	Gamma	[16]
Dyspnea	100.00	+/-25%	Gamma	Assumption
Immune	7,680.00	+/-25%	Gamma	[16]
Anaemia	898.00	+/-25%	Gamma	[16]

* Modeled in the first cycle of "stable disease"

**TIL costs including management of side effects, production costs and hospitalization

***based on 4* follow-up visit physician+CT scan (stable)

****cost for palliative care or end-stage disease care was based on the per diem cost of a palliative care unit

*****resembles 2-5 days hospitalization for severe toxicity (grade III-IV)

Abbreviations: PFS = progression free survival, OS = overall survival, SE = standard error.

Input cost price calculation NKI-AVL: based on *N*= 10 patients from the pilot study

Inclusion criteria of the pilot study were: a resectable metastasis of at least 2-3cm; a sufficient heart, lung and kidney function; a maximum of 2 asymptomatic brain metastasis smaller than 1cm; not concurrently being treated with immune function-suppressing medication; not having auto-immune disorders; and a minimum expected life span of 3 months.

References corresponding to the table

6. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med. 2010;363:711–723. doi: 10.1056/NEJMoa1003466.

10. Besser MJ, Shapira-Frommer R, Treves AJ, Zippel D, Itzhaki O, Hershkovitz L, et al. Clinical responses in a phase II study using adoptive transfer of short-term cultured tumor infiltration lymphocytes in metastatic melanoma patients. Clin Cancer Res. 2010;16:2646–2655. doi: 10.1158/1078-0432.CCR-10-0041.

11. Radvanyi LG, Bernatchez C, Zhang M, Fox PS, Miller P, Chacon J, et al. Specific lymphocyte subsets predict response to adoptive cell therapy using expanded autologous tumor-infiltrating lymphocytes in metastatic melanoma patients. Clin Cancer Res. 2012;18:6758–6770. doi: 10.1158/1078-0432.CCR-12-1177.

16. National Institute for Health and Care Excellence (NICE). Ipilimumab for previously treated advanced (unresectable or metastatic) melanoma (TA268). In; 2012.

18. Beusterien KM, Szabo SM, Kotapati S, Mukherjee J, Hoos A, Hersey P, et al. Societal preference values for advanced melanoma health states in the United Kingdom and Australia. Br J Cancer. 2009;101:387–389. doi: 10.1038/sj.bjc.6605187.

20. Besser MJ, Shapira-Frommer R, Itzhaki O, Treves AJ, Zippel DB, Levy D, et al. Adoptive transfer of tumor-infiltrating lymphocytes in patients with metastatic melanoma: intent-to-treat analysis and efficacy after failure to prior immunotherapies. Clin Cancer Res. 2013;19:4792–4800. doi: 10.1158/1078-0432.CCR-13-0380.

21. Geukes Foppen MH, Donia M, Svane IM, Haanen JB. Tumor-infiltrating lymphocytes for the treatment of metastatic cancer. Mol Oncol. 2015;9:1918–1935. doi: 10.1016/j.molonc.2015.10.018.

22. Dutch National Health Care Institute. Dutch Pharmaceutical Guidelines, consulted in 2016; avaliable at: https://www.medicijnkosten.nl/.

23. Barzey V, Atkins MB, Garrison LP, Asukai Y, Kotapati S, Penrod JR. Ipilimumab in 2nd line treatment of patients with advanced melanoma: a cost-effectiveness analysis. J Med Econ. 2013;16:202–212. doi: 10.3111/13696998.2012.739226.

24. Ellebaek E, Iversen TZ, Junker N, Donia M, Engell-Noerregaard L, Met O, et al. Adoptive cell therapy with autologous tumor infiltrating lymphocytes and low-dose Interleukin-2 in metastatic melanoma patients. J Transl Med. 2012;10:169. doi: 10.1186/1479-5876-10-169.

25. Hakkaart-van Roijen L vdLN, Bouwmans C, Kanters T, Tan SS,. Kostenhandleiding. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. (in Dutch). In opdracht van Zorginstituut Nederland. 2015.

31. Crott, R. Cost effectiveness and cost utility of adjuvant interferon α in cutaneous melanoma: A review. *PharmacoEconomics* (2004). doi:10.2165/00019053-200422090-00002

Supplement 4 - Results from the cost-effectiveness analysis per incorporated scenario

	# patients in the model	Costs	QALYs	LYs	iCosts
Base case results[12]					
TIL	1000	€81,085	0.43	0.68	
Ipilimumab	1000	€94,705	0.38	0.58	
					€-13,620
Scenarios					
"Base case" - If TIL shows better survival rates (at least 10% improvement) compared to ipilimumab, TIL will be implemented.					
TIL	46/58*	€81,085	0.43	0.68	
Ipilimumab	89	€94,705	0.38	0.58	
					€-13,620
"TIL more effective" - The effectiveness of TIL has increased with 10% (clinically relevant) due to research developments.					
TIL	1000	€81,668	0.49	0.78	
Ipilimumab	1000	€94,705	0.38	0.58	
					€-13,037
"Combination therapy"- TIL is used in combination with other immune or personalized therapies (i.e. nivolumab or vemurafenib)					
TIL	1000	€123,853	0.57	0.91	
Ipilimumab	1000	€94,705	0.38	0.58	
					€29,148
"Low cost competition"- If TIL turns out to be cost-effective, pharmaceutical companies will lower the prices of competing immunotherapies.					
TIL	1000	€77,478	0.43	0.68	
Ipilimumab	1000	€75789	0.38	0.58	
					€1,512
"Less IL2 treatment"- Additional interleukin-2 treatment after infusion of TIL is not be necessary anymore.					
TIL	1000	€80,590	0.43	0.68	
Ipilimumab	1000	€94,705	0.38	0.58	
					€-14,115
"TIL production outsourced"- Production of TIL is of interest for the pharmaceutical market and is outsourced by a commercial company.					
TIL	1000	€152,085	0.43	0.68	
Ipilimumab	1000	€94,705	0.38	0.58	
					€57,380
"Automatic TIL production"- Production of TIL is less expensive (30% reduction) due to more automatic process steps.					
TIL	1000	€70,435	0.43	0.68	
Ipilimumab	1000	€94,705	0.38	0.58	
					€-24,270

* only 46 patients patients actually receive TIL-therapy as a part of the patient fails on TIL-therapy and would still receive ipilimumab.

	iQALYs	ICER	Conclusion	NMB	Probability of TIL being cost-effective
	0.050	€-270,281	Dominant	€19,473	89%
	0.050	€-270,281	Dominant	€19,693	90%
	0.114	€-114,128	Dominant	€23,270	92%
	0.192	€151,520	Dominated	€-10,837	12%
	0.050	€30,014	Threshold	€11,418	56%
	0.049	€-287,058	Dominant	€20,366	92%
	0.050	€1,138,642	Dominated	€-51,551	0%
	0.050	€-481,620	Dominant	€22,667	99%

Supplement 5 - Reasons for in and excluding scenarios for cost-effectiveness modelling

Scenario	Labelled as	Reason
"Base case"	Likely	The mean likelihood in respondents having 1 or more years of experience was above 55%. (Figure 2) Besides, the respondents described a minimal 1-year survival of 61.3% (CI: 55.2%-67.5%) to be the minimal acceptable effectiveness of TIL-therapy to adopt it, which is shown by the published results from the observational studies.
"Competition"	Unlikely	This scenario was discussed in the internal research group. In the literature review used for further selection, no medicines were described to be in development that are expected to be more effective and/or equal in costs than TIL-therapy[18]. This could however be explained by the scope of the review.
"TIL more effective"	Likely	In the review used to identify likely scenarios several developments were discussed that would result in TIL-therapy being more effective[18].
"Biomarker"	Unlikely	A biomarker to specifically select patients that would benefit from TIL-therapy seems not likely to be discovered in the coming years as similar variables seem to be prognostic for selecting patients for CTL-4 and PD1 antibodies[18]. Also the expected mean likelihood of this scenario was scored below 40%, indicating that this scenario is unlikely to happen in the coming 5 years.
"TCR therapy"	Unlikely	Results of TCR gene therapy are still too preliminary and therefore it as unlikely that this treatment would become available and would dominate TIL-therapy within 5 years[18,46]. Also the expected mean likelihood of this scenario was scored below 40%, indicating that this scenario is unlikely to happen in the coming 5 years.
"Patients unconvinced"	Unlikely	This scenario was discussed in the internal research group. Based on a previous analysis in which we evaluated factors that may influence the choice of receiving TIL-therapy among eligible patients, patients were positive about receiving TIL-therapy[11].
"2nd line treatment"	Unlikely	This scenario was discussed in the internal research group. This scenario is currently similar to the base case model which was already simulated. The addition of this scenario in the cost-effectiveness analysis would thus not add information. Furthermore, in light of the developments and new treatment options available it is more likely that TIL-therapy will be placed after progression on 1st line treatment with anti-PD1 and 2nd line treatment with CTL-4 antibodies. Therefore this scenario was labelled "unlikely".
"3rd line treatment"	Likely	The mean likelihood in respondents describing themselves as familiar or an experts and in respondents having 1 or more years of experience was above 55%. (Figure 2)
"Combination therapy"	Likely	The mean likelihood in all respondents was above 55%.

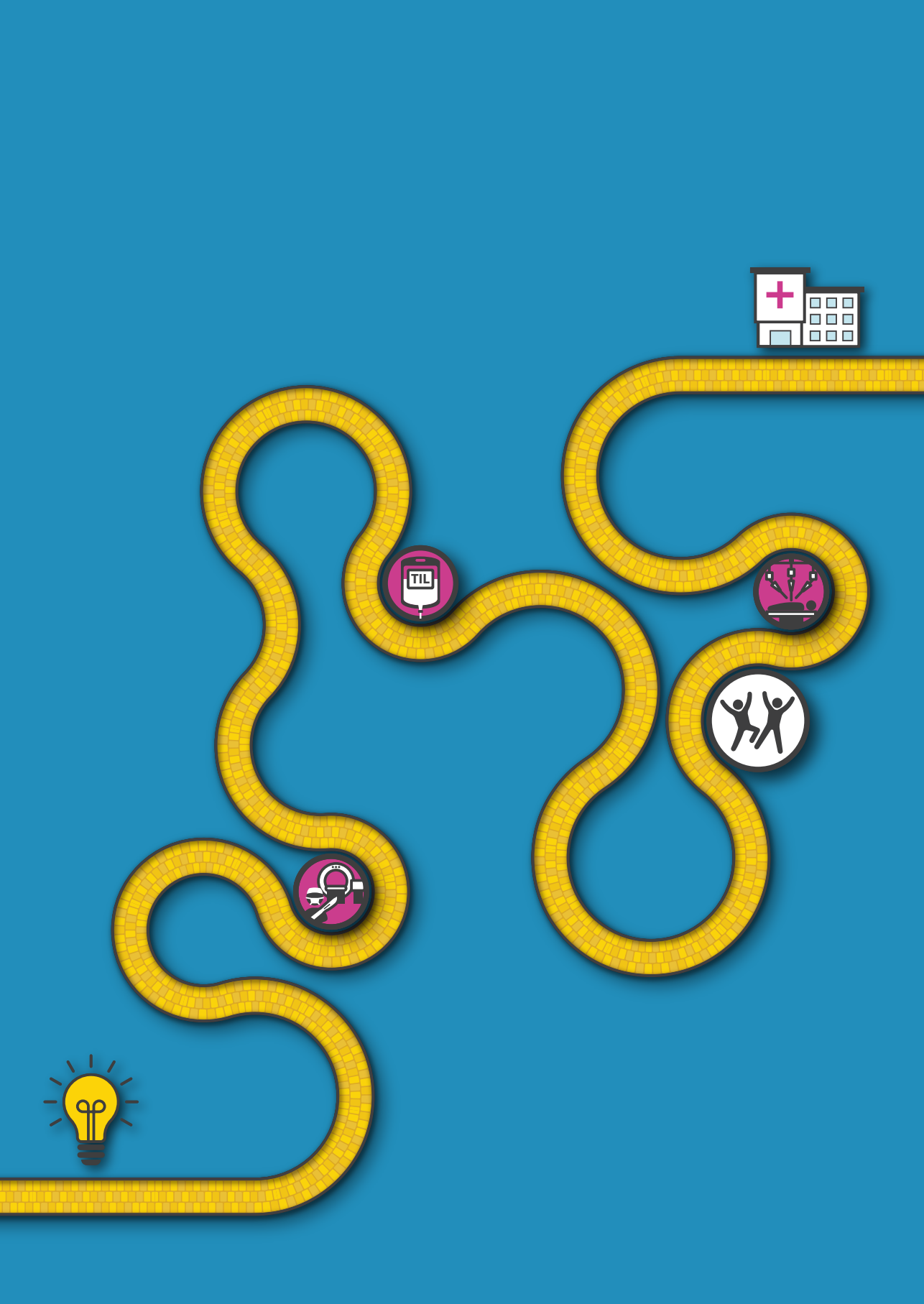
Scenario	Labelled as	Reason
"Clinicians unconvinced"	Likely	This scenario was discussed in the internal research group. In the additional questions that were asked in the web-based questionnaire only 3 of 29 respondents described that none of the stated reasons (complexity, user-friendliness or intensity of IL2 treatment) applied to be unconvinced on the use of TIL-therapy and that clinicians would apply TIL-therapy. Besides, they expressed to be mainly indifferent (31%) or to disagree (37%) with the statement that TIL-therapy would improve quality of life of patients compared to ipilimumab. Therefore it was thought very likely that clinicians remain unconvinced in the coming years.
"Low cost competition"	Likely	This scenario was discussed in the internal research group. Following the interests of pharma and trends seen in decreasing medicines it is thought to be very likely that pharma would lower the prices of competing immunotherapies when TIL-therapy turns out to be a true and less expensive competitor.
"Less IL2 treatment"	Likely	This scenario was discussed in the literature review showing that several research group are evaluating different IL2 dosing schemes[18], therefore this scenario was labelled as likely.
"Influence by companies"	Unlikely	This scenario was discussed in the internal research group. By our clinical experts it is not expected that this would have a significant influence. If TIL-therapy would be (cost-)effective, clinicians would provide this treatment regardless their agreements with pharmaceutical companies.
"TIL production outsourced"	Likely	This scenario was discussed in the internal research group. The head of the production facility explained that this scenario has been seen in other countries and therefore it seems a very realistic scenario.
"Automatic TIL production"	Likely	The mean likelihood in all respondents was above 55%.

Supplement 6 - Results from the questions included in the questionnaire

	# Respondents (%)	Mean score (95% CI)
Theme: effectiveness		
What would be the minimal effectiveness of TIL leading to accept TIL as a standard therapy for you? Expressed in one-year survival rate (%)?	26 (89%)	61.3% (55.2- 67.5)
What would be the risk of developing other types of cancer such as lymphomas by activating the immune system by injecting TILs (%)?	24 (83%)	6.4% (4.5- 8.3)
Agreement with statement: TIL treatment provides significantly better quality of life compared to ipilimumab.	29 (100%)	
• Strongly agree	2 (7%)	
• Agree	7 (24%)	
• Nor agree or disagree	9 (31%)	
• Disagree	10 (34%)	
• Strongly disagree	1 (3%)	
Theme: Patient perspective		
Could you estimate the percentage of the eligible patients (metastatic melanoma patients) you thinks is aware of TIL as a potential treatment (in %)	25 (86%)	20.8% (14.7- 26.9)
Theme: Clinician's attitude		
What would be the main reason for clinicians to be unconvinced of introducing TIL-therapy? (more options were possible)	29 (100%) 47 answers	
• Complexity of TIL (understanding of TIL growth and the clinical process)		18/47 (38%)
• Intensiveness of IL2 and expected adverse events		15/47 (32%)
• User-friendliness of TIL (practical issues in implementation)		11/47 (23%)
• This is not the case: clinicians will treat patients with TIL		3/47 (6%)

PART IV

Mainstream HTA


TIL

9

Long-term functional outcomes after robot-assisted prostatectomy compared to laparoscopic prostatectomy: Results from a national retrospective cluster study

Melanie Lindenberg
Valesca Retèl
Jacobien Kieffer
Carl Wijburg
Laurent Fossion
Henk van der Poel*
Wim van Harten*
* Shared last author

Submitted

ABSTRACT

Introduction

Despite multiple studies on the efficacy of robot-assisted radical prostatectomy (RARP), there is no definitive conclusion about the added value of RARP. A retrospective cluster study was conducted to evaluate long-term sexual and urinary functioning after RARP and laparoscopic radical prostatectomy (LRP) based on real-world data from 12 Dutch hospitals.

Patients and methods

Data was collected from patients who underwent surgery between 2010 and 2012. A mixed effect model was used to evaluate group differences on urinary and sexual functioning measured with the EPIC-26 questionnaire. A regression analysis was conducted to evaluate the relationship between these functional outcomes and hospital volume, age, neurovascular bundle preservation, D'Amico risk score and receiving radiotherapy.

Results

In total, 1,370 (65.1%) patients participated, 907 underwent RARP and 463 LRP, with a median follow-up time of 7.08 years (SD=0.98) and median age of 71.5 years (SD=5.8). Patients who underwent RARP showed a statistically and clinically significant better urinary functioning compared to patients who underwent LRP (p=0.002). They also showed a shorter procedure time (p=<0.001), reduced blood loss (p=<0.001), and a higher chance of neurovascular bundle preservation (39.8% vs 29.1%; p=<0.01). Being younger, receiving a nerve-sparing procedure and not receiving adjuvant radiotherapy were independently associated with better urinary and sexual functioning in both groups.

Conclusion

RARP resulted in better urinary function compared to LRP. Based on the peri-operative and long-term functional outcomes, RARP is preferred over LRP. Given these results, decision-makers are advised to reconsider their position on coverage, especially when RARP proves to be cost-effective.

INTRODUCTION

To guide treatment decisions among prostate cancer patients, knowledge about the impact of treatments on their health-related quality of life (HRQoL) and their preferences is important[1]. Radical prostatectomy (RP) is known for its negative impact on urinary and sexual functioning[1,2]. The introduction of the Da Vinci® (Intuitive Surgical) robot in prostate cancer care was expected to improve HRQoL and survival by providing better sight and a greater range of motion.

Although no benefits have been proven in recurrence-free survival[3], the introduction of robot-assisted RP (RARP) has shown improvements in hospital stay, blood loss, urinary incontinence, and erectile functioning compared to open (ORP) and laparoscopic (LRP) RP[4–6]. To date, systematic reviews are still unable to draw definitive conclusions from studies on the efficacy of RARP due to high variability in patient selection, study design, and outcome measurements[7–9].

More recently, population-based studies compared functional outcomes after RARP and ORP[10–12]. Showing better sexual functioning after two years for RARP, but no long-term difference in functional outcomes was seen[10,11]. As multiple studies have suggested that hospital volume is associated with better functional outcomes[13,14], hospital volume should be taken into account when evaluating RARP. Note, however, that the previous and other recent observational and randomized studies mainly compared RARP to ORP[15,16]. Therefore, the clinical evidence base to decide on the position of RARP in the current treatment landscape, especially in comparison to LRP, is inconclusive[17].

In this study, we evaluated the long-term (6-9 years) urinary and sexual functioning in 1370 prostate cancer survivors after RARP and LRP based on real-world data from the Netherlands, collected in 12 hospitals.

METHODS

Study design and patient population

Prostate cancer patients who underwent surgery between 2010 and 2012 were invited to participate in this retrospective cluster study. This timeframe was specifically chosen to involve high volume hospitals that still performed LRP as well as larger hospitals that already adopted RARP and had performed at least 50 RARPs. We selected hospitals with different hospital volumes for both interventions. In total, 12 hospitals participated in our study, eight that performed RARPs, and seven LRPs during our timeframe. Four hospitals provided data for both procedures. From these hospitals, patients were invited when (i) their vital status was known or could be validated with the general practitioner, (ii) they were not part of the first 50 RARPs, (iii) they were living in the Netherlands, and (iiii) they had sufficient command of the Dutch language. General clinical information was collected from deceased patients.

The study was approved by the medical ethical committee of the Netherlands Cancer Institute (NKI-AVL) and the institutional review boards of all recruiting hospitals. All participants gave consent to use and evaluate the sampled data as described in the informed consent.

Procedure

Figure 1 shows the CONSORT diagram of the study. From the 2626 patients assessed for eligibility, 2117 were invited by their treating physicians to participate between January 2018 and March 2019.

Study measures

Primary outcome measures were the Urinary Incontinence domain and Sexual domain of the expanded prostate cancer index composite short form (EPIC-26)[18]. Besides, being incontinent and having erectile dysfunction was evaluated by one question per domain: number of pads used (use of ≥1 pad) and the quality of the erection (not firm enough for any sexual activities), respectively (Supplement 5).

Secondary outcome measures were: bowel, hormonal and urinary irritative/obstructive of the EPIC-26, the Summary score of the EORTC QLQ C30 version 3.0, and utilities measured by EQ5D-5L for overall quality of life. All these questionnaires were incorporated in one survey that was sent to the patients.

Additionally, clinical characteristics were retrieved from the medical record (see Table 1 and Supplement 2). Besides, the survey incorporated questions on social-demographics, complications (Clavien Dindo classification[19]), hormonal treatment or radiotherapy within 6 months after treatment with or without PSA rise, and the use of additional care, pharmaceuticals or instruments for complaints related to erectile dysfunction and incontinence. Furthermore, five questions from the EPIC-26 and EORTC-QLQ-PR25 were included in the survey to evaluate the preoperative status of the patients. Baseline continence was defined as no pads used and no unintentional release of urine. Baseline potency was defined as having no problem at all with getting or maintaining an erection. Finally, for patients that deceased between surgery and inclusion, the date of death and cause of death was retrieved from the medical record.

Statistical analysis

The domain scores of the EPIC-26 were calculated according to published scoring algorithms. Some of the questions had to be recoded because an additional answer option was given: "Not applicable (because I was not sexually active)". The recoding procedure is provided in Supplement 1.

To analyse the difference in the primary and secondary outcomes between RARP and LRP a mixed effects modelling approach with random intercept was used. The primary analysis included only patients who were defined as continent and potent at baseline. Clustering based on hospital was included as a random factor. The models were adjusted for possible confounders: age at inclusion, D'Amico risk score[20], receiving radiotherapy, neurovascular bundle preservation, use of pharmaceuticals or instruments for erectile dysfunction, hospital type, and hospital volume. The confounders were added stepwise as fixed factors. Details on the evaluation of the best model were incorporated in Supplement 3. The P-value for the overall model effects was set at .05. A difference of 6-9 points on the Urinary Incontinence domain and a difference of 10-12 on the Sexual Domain were considered clinically significant[21].

Additionally, the socio-demographic and clinical characteristics of the groups were compared using chi-square tests and independent samples t-test. The survival of the total patient population receiving RP was compared with Kaplan Meier curves and a log-rank test. Patients that died after the 1st of March 2018 were excluded from survival analysis because patient recruitment in the first recruiting hospital was then completed.

Finally, regression analyses were conducted using mixed effect models with random intercept and random clustering of hospital to evaluate the influence of hospital volume, age, D'Amico risk score, receiving radiotherapy, and neurovascular bundle preservation on better urinary and sexual functioning.

RESULTS

Study sample

The total set of potential patients was n=2626. In total 202/2626 patients died before inclusion, of which 164 died before March 1, 2018. Overall mortality in the LRP group (n=72) was significantly higher than in the total RARP group (n=92) (8.7%; 5.1% log-rank: 0.003). Prostate cancer-specific mortality was also higher in the LRP group (RARP: N=17, 0.95%, LRP: N=12, 1.44%, log-rank 0.326), though not significant.

From the 2117 invited patients, 1378 patients completed the questionnaire showing an overall response rate of 65.1%. Eight patients were removed from the study sample because of various reasons (Figure 1), resulting in a final sample of 1370 patients. 907 underwent RARP, and 463 LRP.

Patient and hospital characteristics

All patient and clinical characteristics are listed in Table 1. The median age of the study sample at inclusion was 71.5 years (46.6-85.1), and the median time to follow-up was 7.08 years (5.27–9.86). At baseline, 3.6% and 18.9% of the patients were considered incontinent and impotent respectively, which did not significantly differ between the groups. In the RARP group, patients were more often operated in high volume (p<0.01) and academic hospitals (p<0.001) compared to the LRP group.

Clinical characteristics and per- and postoperative outcomes

In the RARP group, a higher proportion of patients was classified as clinical high-risk[20] (33.6%; 26.6%, p=0.02). Furthermore, RARP showed a shorter procedure time (159 min; 191 min, p=<0.001), less blood loss (156 ml; 250 ml, p=<0.001), and a higher chance of neurovascular bundle preservation (39.8%; 29.1%, p=<0.01).

Positive surgical margin rate (RARP: 27.3%; LRP: 25.9%, p=0.59) and biochemical recurrence (RARP: 33.6%; LRP: 33.7%, p=0.99) was similar between the groups. Notably, a higher number of LRP patients received hormonal therapy compared to RARP (10.8%; 7.5%, p=0.07).

Follow-up characteristics

The complication rate (RARP: 18.5%; LRP: 16.4%, p=0.34) and the severity of the complications

was similar between the groups (p=0.49). The LRP group had more often complaints of incontinence (52.1%; 67.3%, p<0.001) and of erectile dysfunction directly after surgery (74.4%; 81.2%, p=0.02). Table 1 also presents the proportion of patients that used additional care for those complaints.

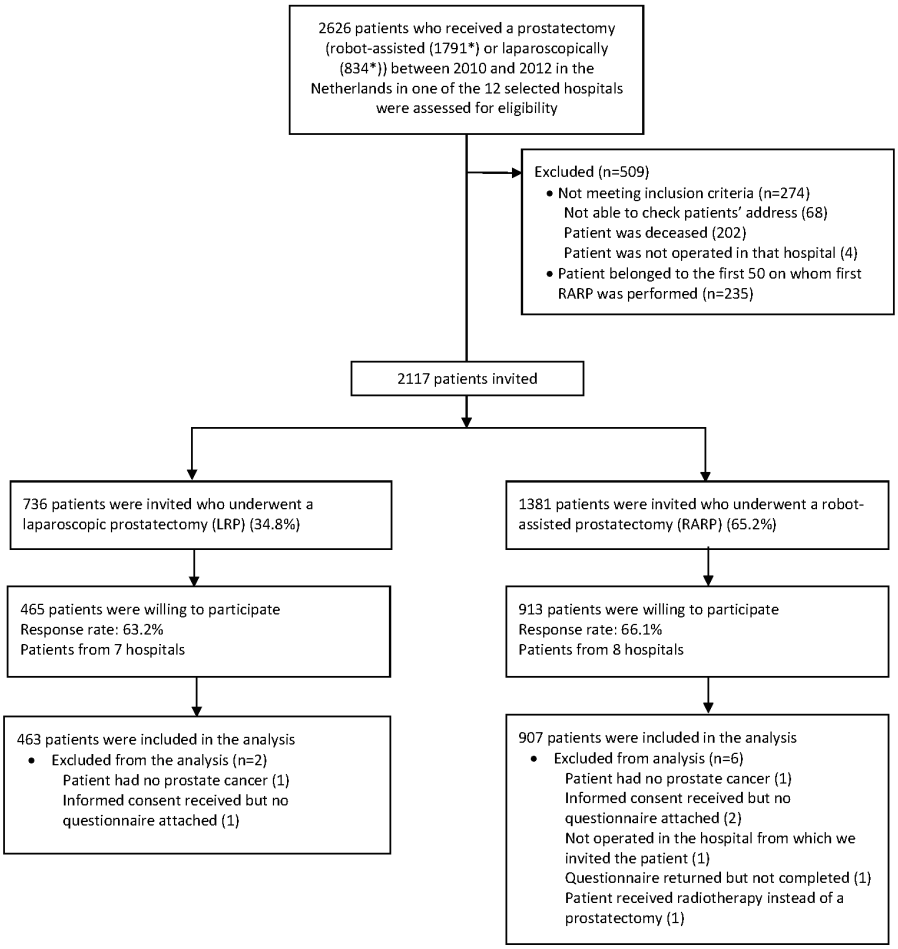


Figure 1. CONSORT diagram. * of one of the eligible patients it was unknown whether he underwent LRP or RARP, eventually this patient did not participate; RARP = robot-assisted radical prostatectomy, LRP = laparoscopic radical prostatectomy.

Table 1 - Sociodemographics, clinical characteristics and peri and postoperative measurements.

	Robot-assisted prostatectomy (n=907)	**Laparoscopic prostatectomy (n=463)**	**Sig.**
Sociodemographics and general characteristics			
Age			
at surgery (median, range)	64 (39-79)	64 (45-75)	0.51
at filling in questionnaire (median, range)	71.21 (46.6-85.1)	72.08 (51.8-83.7)	0.06
Marital status			
Married or living together with partner	792 (87.7%)	407 (88.3%)	0.90
Missing	4	2	
Highest completed education level			0.32
Primary education	41 (4.6%)	29 (6.4%)	
Secondary education vocational education	523 (58.5%)	253 (56%)	
Higher education	330 (36.9%)	170 (37.6%)	
Volume hospital (number of prostatectomies per year)			<0.01
< 50 patients per year	86 (9.5%)	90 (19.4%)	
50-100 patients per year	113 (12.5%)	243 (52.5%)	
100-150 patients per year	243 (26.8%)	130 (28.1%)	
>150 patients per year	465 (51.3%)	0	
Type of hospital			<0.001
General hospital	407 (44.9%)	337 (72.8%)	
Academic or specialized hospital	500 (55.1%)	126 (27.2%)	
Baseline incontinent (%)	29 (3.2%)	21 (4.6%)	0.21
Missing	7	5	
Baseline impotent (%)	161 (18.2%)	98 (21.7%)	0.12
Missing	20	12	
Clinical characteristics			
Preoperative prostate volume (mL) median, range	41 (12-220)	38 (0-170)	0.06
Missing	197	235	
Preoperative PSA level (ng/mL) median, range	8.5 (1-254)	9 (0.7-80)	0.36
Missing	18	14	
Clinical stage			<0.01
cT1a-1b	12 (1.4%)	5 (1.1%)	
cT1c	338 (40.6%)	241 (54.5%)	
cT2a	219 (26.3%)	105 (23.8%)	
cT2b	96 (11.5%)	34 (7.7%)	
cT2c	86 (10.3%)	27 (6.1%)	
cT3	81 (9.7%)	30 (6.8%)	
Missing	75	21	
cGleason Score			0.34
≤6	493 (55.4%)	247 (53.8%)	
7	293 (32.9%)	167 (36.4%)	
>7	104 (11.7%)	45 (9.8%)	
Missing	17	4	

(Continued on next page)

Table 1 (continued) - Sociodemographics, clinical characteristics and peri and postoperative measurements.

	Robot-assisted prostatectomy (n=907)	**Laparoscopic prostatectomy (n=463)**	**Sig.**
D'Amico risk classification			0.02
Low risk	264 (29.5%)	138 (29.8%)	
Intermediate risk	330 (36.9%)	202 (43.6%)	
High risk	300 (33.6%)	123 (26.6%)	
Missing	13	0	
Skin-to-skin procedure time (minutes) median, range	159 (70- 412)	191 (72- 300)	<0.001
Missing	48	151	
Perioperative blood loss (ml), median, range	156 (0- 3,200)	250 (0- 3,300)	<0.001
Missing	54	15	
Neurovascular bundle preservation			<0.01
Bilateral	356 (39.8%)	133 (29.1%)	
Unilateral	275 (30.8%)	151 (33%)	
None	263 (29.4%)	173 (37.9%)	
Missing	13	6	
Pathologic characteristics			
Pathological T-stage			<0.001
pT0	8 (0.9%)	0	
pT1	8 (0.9%)	1 (0.2%)	
pT2a	95 (10.7%)	121 (26.3%)	
pT2b	63 (7.1%)	13 (2.8%)	
pT2c	439 (49.54%)	162 (35.2%)	
pT3	254 (28.7%)	157 (34.1%)	
pT4	19 (2.1%)	6 (1.3%)	
Missing	21	3	
pGleason Sum			0.11
≤6	317 (35.3%)	151 (32.6%)	
7	436 (48.5%)	251 (54.2%)	
>7	146 (16.2%)	61 (13.2%)	
Missing	8	0	
Prostate volume (g), median, range	55 (5-718)	54 (12-200)	0.33
Missing	90	157	
Positive resection margin (%)	246 (27.3%)	120 (25.9%)	0.59
Missing	5	0	
Lymph node dissection performed (%)	343 (37.9%)	124 (26.8%)	<0.01
Missing	2	1	
Number of lymph nodes removed median, range	9 (1-38)	12 (1-56)	<0.001
Missing	31	45	
Positive lymph nodes (% of patients that received a lymph node dissection)	10.8%	7.0%	0.35
Missing	0	23	

(Continued on next page)

Table 1 (continued) - Sociodemographics, clinical characteristics and peri and postoperative measurements.

	Robot-assisted prostatectomy (n=907)	**Laparoscopic prostatectomy (n=463)**	**Sig.**
Characteristics of follow-up			
Occurrence of BCR when at least 3 years of follow-up data is available (%)	190 (33.6%) n=565	101 (33.7%) n=300	0.99
Received radiotherapy (%)	188 (20.7%)	97 (21.0%)	0.96
Received hormonal therapy (%)	68 (7.5%)	46 (10.8%)	0.07
Duration of admission (days) median, range	3.0 (2-27)	3.0 (2-27)	0.11
Complications (patient-reported) classified by Clavien-Dindo grading system	168 (18.5%)	76 (16.4%)	0.34
Grade 1	70 (41.7%)	34 (44.7%)	0.49
Grade 2	41 (24.4%)	13 (17.1%)	
Grade 3	46 (27.4%)	21 (27.6%)	
Grade 4	11 (6.5%)	8 (10.5%)	
Patients having incontinence complaints after surgery°	461 (52.1%)	307 (67.3%)	<0.001
Missing	22	7	
Among the patients with complaints; patients that used additional care$	399 (86.6%)	285 (92.8%)	<0.01
Physiotherapy	376 (94.2%)	268 (94.2%)	0.912
Number of visits (median, range)	6 (1-60)	8 (1-60)	0.015
Visiting the general practitioner	24 (6.0%)	17 (6.0%)	0.98
Number of visits (median, range)	2 (1-20)	2 (1-40)	0.31
Surgical procedure (e.g. male sling)	22 (5.5%)	39 (13.7%)	<0.01
Among the patients with complaints; the number of pads used in the previous 4 weeks			0.02
None	205 (44.7%)	103 (34.0%)	
1 per day	154 (33.6%)	115 (38.0%)	
2 per day	60 (13.1%)	44 (14.5%)	
3 or more per day	40 (8.7%)	41 (13.5%)	
Missing	2	4	
Patients having complaints of erectile dysfunction after surgery°	653 (74.4%)	362 (81.2%)	0.02
Missing	29	17	
Among the patients with complaints, patients that used additional care$	195 (29.9%)	104 (28.7%)	0.68
Physiotherapy$	29 (14.9%)	17 (16.3%)	0.8
Number of visits (median, range)	8.00 (1-25)	9.00 (2-30)	
Visiting the general practitioner$	45 (23.1%)	14 (13.5%)	0.047
Number of visits (median, range)	2.00 (1-12)	2.00 (1-4)	

(Continued on next page)

Table 1 (continued) - Sociodemographics, clinical characteristics and peri and postoperative measurements.

	Robot-assisted prostatectomy (n=907)	**Laparoscopic prostatectomy (n=463)**	**Sig.**
Visiting a different specialist$	127 (65.1%)	78 (75.0%)	0.08
Most frequent described specialties: $			
• Urologist or urology department	94 (74.0%)	63 (80.8%)	
• Sexologist or outpatient clinic for sexuality	23 (18.1%)	5 (6.5%)	
Number of visits for all the described specialists (median, range)	3 (1-80)	3 (1-18)	
Patients that used pharmaceuticals or other medical instruments for complaints of erectile dysfunction in the whole population	326 (36.5%)	146 (32.3%)	0.13
Missing	13	11	
Using a tablet (e.g. Cialis, Viagra, Levitra)$	207 (63.5%)	88 (60.3%)	0.5
Using an intra-urethral injection (e.g. Muse)$	6 (1.8%)	14 (9.6%)	<0.01
Using an intra-cavernous injection (e.g. Androskat)$	116 (35.6%)	55 (37.7%)	0.66
Prothesis$	4 (1.2%)	0 (0%)	0.32
Vacuum constriction device$	51 (15.6%)	32 (21.9%)	0.1

Sociodemographics, clinical characteristics and peri and postoperative measurements including the use of care after surgery of the study population of 1370; other numbers apply when indicated. ° These questions asked patients whether they had experienced complaints directly after surgery and whether they used additional care for those complaints. $ These percentages represent the number of patients that used a certain type of care of the total number of patients that described to use care or pharmaceuticals for certain complaints. Patients were allowed to choose multiple answers therefore the numbers do not add up.

Primary outcome measurements

The RARP group showed a statistically and clinically significant better urinary function compared to the LRP group (estimated means: 73.34; 64.98, p=0.002) (Table 2). No significant differences were found in sexual functioning between RARP and LRP (28.89; 24.77, p=0.12). Based on the number of pads used, RARP patients showed a higher chance to be continent (p=0.002). Based on the firmness of the erection, the RARP group showed a non-significant higher chance for being potent (p=0.052) (Supplement 5). Supplement 4 shows the observed scores of the Urinary Incontinence and Sexual domain.

Table 2 - Primary and Secondary outcomes.

A. Primary outcomes				
Outcome measurement	Robot-assisted prostatectomy (907)	Laparoscopic prostatectomy (463)	Absolute difference	P-value
EPIC domain: Urinary incontinence (0-100) (SE)	73.34 (1.33) N=823#	64.98 (1.63) N=419#	8.35	0.002
Being incontinent based on Question 27 from Urinary Incontinence domain (SE) $	41.5% (3.7%) N=842#	52.9% (4.2%) N=427#	11.4%	0.002
EPIC domain: Sexual (0-100) (SE)	28.89 (1.87) N=659#	24.77 (2.18) N=327#	4.12	0.12
Having erectile dysfunction based on Question 59 from Sexual domain (SE) $	67.7% (2.9%) N=682#	76.2% (3.4%) N=336#	8.5%	0.052

B. Secondary outcomes			
	Robot-assisted prostatectomy (907)	Laparoscopic prostatectomy (463)	P-value
EPIC domains (0-100) (SE)			
• Urinary irritative / obstructive*	95.75 (0.43) N=868#	95.08 (0.56) N=451#	0.36
• Bowel~	94.83 (0.55) N=870#	94.70 (0.65) N=452#	0.88
• Hormonal~	92.78 (0.47) N=859#	91.39 (0.61) N=441#	0.04
EORTC summary score (0-100) (SE)*	92.33 (0.39) N=867#	91.37 (0.50) N=444#	0.09
EQ5D-5L (0-1) (SE)*	0.918 (0.005) N=872#	0.914 (0.006) N=450#	0.54

Reports the sample size included in the model as for some patients information on incorporated confounders was missing, those were left out of the analysis. A. Shows the primary outcomes noted in estimated marginal means returned from the mixed effect model which are controlled for hospital (cluster), age at the time of completing the questionnaire, D'Amico risk score, radiotherapy received at any time during follow-up (both salvage and adjuvant) and neurovascular bundle preservation. In this analysis, patients being incontinent and impotent before surgery were excluded. The analysis on the Urinary Incontinence was also controlled for hospital type and for the Sexual domain we additionally controlled for the use of pharmaceuticals or instruments used when patients had complaints after surgery. The addition of hospital volume depressed the fit of the model in both domain scores and was therefore not included. Incontinence was defined as use of 1 or more pads per day. Having erectile dysfunction was defined as: erection not firm enough for any sexual activity. The observed results from the two separate questions of the EPIC-26 are presented in Supplement 4. $ The analysis was controlled for cluster, D'Amico risk score, Radiotherapy, nerve-sparing and age.

B. Shows the estimated marginal means of the secondary outcomes. ~ controlled for cluster, D'Amico risk score, and radiotherapy. Age depressed the model fit and was not included. * controlled for cluster, D'Amico risk score, and nerve-sparing. Age depressed the model fit and was not included.

Secondary outcome measurements

On the EPIC domains: urinary irritative/obstructive and bowel, both groups showed high and similar scores (Table 2). On the EPIC hormonal domain, a significant better score was seen for RARP, corresponding with the lower number of patients receiving hormonal treatment (RARP: 92.78; LRP: 91.39, p=0.04). The EORTC-C30 summary score (RARP: 92.33; LRP: 91.37) and the utility values were comparable between the groups (RARP: 0.918; LRP: 0.914).

Factors influencing functional outcome scores

Table 3 presents the results from the regression analysis. Patients in both groups show a statistically and clinically significantly better urinary functioning when not receiving radiotherapy (RARP:+7.55, p<0.001; LRP:+9.39, p=0.005), and when having a nerve-sparing procedure (both groups: p<0.05). Furthermore, a larger hospital volume was a clinically significant predictor of better urinary functioning in the LRP group.

Only for patients undergoing RARP, a statistically significantly better sexual functioning was seen when not receiving radiotherapy (+6.66, p=0.007). Patients in both groups show significantly better sexual functioning when being younger of age (both groups: p<0.01) and when having neurovascular bundle preservation. In the LRP group, a higher hospital volume was predictive of better sexual functioning.

Table 3 - Regression analysis for primary outcome measures: EPIC-26 Urinary Incontinence and Sexual domain by means of a mixed effect model corrected for cluster (hospital).

Urinary Incontinence domain						
	Robot-assisted radical prostatectomy			**Laparoscopic radical prostatectomy**		
Parameter	**Estimate**	**SE**	**Sig.**	**Estimate**	**SE**	**Sig.**
Intercept	99.03	10.20	.000	54.75	17.73	0.002
D'Amico risk score						
low risk	-2.18	2.32	0.347	5.38	3.87	0.165
intermediate risk	-1.06	2.06	0.605	1.65	3.41	0.628
high risk	0[ref]			0[ref]		
Radiotherapy						
No	7.55	2.13	<0.001	9.39	3.30	0.005
Yes	0[ref]			0[ref]		
Neurovascular bundle preservation						
Bilateral	3.71	2.24	0.098	7.81	3.66	0.034
Unilateral	4.75	2.22	0.033	4.36	3.31	0.189
None	0[ref]			0[ref]		
Hospital volume						
>150 patients	3.70	3.01	0.220	-	-	-
100-150 patients	-5.12	3.18	0.108	10.81	4.07	0.008
50-100 patients	-0.91	3.71	0.807	14.05	3.56	<0.001
0-50 patients	0[ref]			0[ref]		
Age at surgery	-0.49	0.15	0.001	-0.209	0.255	0.412
Sexual domain						
	Robot-assisted radical prostatectomy			**Laparoscopic radical prostatectomy**		
Parameter	**Estimate**	**SE**	**Sig.**	**Estimate**	**SE**	**Sig.**
Intercept	84.67	11.78	<0.001	77.06	15.99	<0.001
D'Amico risk score						
low risk	1.86	2.75	0.501	4.62	3.57	0.196
intermediate risk	0.38	2.44	0.876	5.44	3.21	0.091
high risk	0[ref]			0[ref]		
Radiotherapy						
No	6.66	2.48	0.007	5.38	3.16	0.090
Yes	0[ref]			0[ref]		
Neurovascular bundle preservation						
Bilateral	18.81	2.68	<0.001	11.27	3.34	0.001
Unilateral	12.31	2.66	<0.001	11.16	3.01	<0.001
None	0[ref]			0[ref]		
Hospital volume						
>150 patients	4.09	3.82	0.285			
100-150 patients	-4.37	3.98	0.273	9.97	3.88	0.011
50-100 patients	-0.75	4.56	0.869	2.90	3.27	0.376
0-50 patients	0[ref]			0[ref]		
Age at surgery	-1.09	0.172	.000	-1.125	.227	<0.001

Abbreviations: SE = standard error.

DISCUSSION

To the best of our knowledge, this is the first large cohort study that shows a clear clinical long-term benefit concerning urinary functioning after RARP compared to LRP. Over the years, many studies showed a trend towards short-term better sexual and urinary functioning after RARP[7–12,22,23]. Consistent with the literature, our results indicated that undergoing RARP is associated with shorter procedure time, reduced blood loss, and a higher chance of neurovascular bundle preservation[12,23,24]. Furthermore, RARP could be beneficial for sexual functioning as well, since our data showed that nerve-sparing procedures were closely related to better sexual functioning[25]. Our data do not show significance in better sexual functioning after RARP, although patients seemed to have a lower risk for having erectile dysfunction compared to LRP. This may be explained by aspects other than erectile functioning e.g. partner support and mental health that relate to sexual functioning[26]. Furthermore, age and longer follow-up are likely to affect the valuation of being potent[25].

We also found that irrespective of the type of procedure, being younger, receiving neurovascular bundle preservation, and not receiving adjuvant radiotherapy were associated with having better urinary and sexual functioning. These factors should be discussed with patients to optimally guide their treatment decision since survival between treatment options for localized prostate cancer is comparable[27].

In our study the observed functioning scores in both interventions were lower compared to that of earlier studies with shorter follow up[15]. This can be explained by the fact that our study was executed in different volume hospitals and patients were not excluded based on certain clinical characteristics, possibly more closely resembling daily clinical practice. Besides, as functional outcomes are known to worsen over time because of age[28], a longer follow-up period may also be an explanation for these differences. This argument is strengthened by two studies showing comparable domain scores for RARP after 6 and 3 years respectively[11,12]. In accordance with literature, LRP showed lower scores on the Urinary Incontinence and Sexual domains compared to ORP[22], which suggests that functional outcomes after LRP are worse than after ORP.

Our data also showed that higher hospital volume was associated with better functional outcomes after LRP. Such a relationship was not found among patients undergoing RARP. This could be explained by not having an equal distribution of hospital volume among the groups, as RARP was more often performed in high volume hospitals, and using a relatively short learning curve of only 50 procedures, where a minimum of 200 has been suggested[29.] When using a longer learning curve we would have missed a substantial number of patients

undergoing RARP in lower-volume hospitals (50-100/year), since the majority of these hospitals shifted within our timeframe. Furthermore, it seems that other hospital-specific characteristics e.g. surgeon experience play an important role since we found that hospitals performing "50-100 procedures/year" in the LRP group showed better urinary functioning compared to hospitals performing "100-150 procedures/year".

The retrospective design of our study has some disadvantages. We had a lack of firm baseline information on incontinence and erectile dysfunction, which could have led to recall bias. We also had more missing data in the RARP group, due to the referral system in the Netherlands. Despite a carefully chosen timeframe, we were unable to include patients from very large-volume hospitals (>150 procedures/year) for the LRP group resulting in a selection bias[11,14]. The latter was controlled for by including a cluster variable for hospital. We lacked information on surgeon expertise as some of the operating surgeons have since retired or currently work in a different hospital. Concerning blood loss, we had no information on more reliable measures such as hemoglobin levels or the number of blood transfusions needed. We also did not have information on comorbidities in both groups. Furthermore, although we did not expect differences in comorbidities between the groups[12,22], a difference in all-cause mortality was found, but which could partly be explained by comorbidities. Finally, a response rate of 65% could have led to selection bias.

A great strength of the present analysis is being the first national study evaluating long-term functional outcomes after RARP in a large cohort of prostate cancer patients. Further strengths include the incorporation of healthcare usage for incontinence and erectile dysfunction complaints, the inclusion of patients operated within a narrow timeframe evaluating the early introduction phase of the Da Vinci robot, and controlling for cluster effects by using mixed-effect modelling.

In light of recent developments, e.g. centralization of prostate cancer care, comparison with more recent data is necessary to be more conclusive on the relationship between hospital volume or surgeon experience and improved functional outcomes after RARP. Furthermore, a cost-effectiveness analysis is necessary to decide on coverage for RARP, as RARP comes with substantial extra costs[30]. Finally, the findings in overall and prostate-cancer specific mortality are noteworthy but no conclusions can be drawn yet; this aspect merits further study, taking in-depth medical file data and population registry data into account.

Conclusions

We conclude that RARP is preferred over LRP when it comes to perioperative outcomes and long-term urinary functioning. Therefore, guidelines concerning the preferred surgery type

may change, and decision-makers have to reconsider their position on coverage, especially when RARP proves to be cost-effective compared to LRP.

ACKNOWLEDGEMENTS

We want to thank the urologists and the supportive staff from each participating hospital for supporting us in conducting the research project. Above all, we want to thank all the participating patients for their time and effort to complete the questionnaire. Finally, we want to thank Intuitive surgical for their financial support by means of an Intuitive research grant in the round of 2019-2020 to perform the present analysis.

REFERENCES

1. Lardas, M. *et al.* Quality of Life Outcomes after Primary Treatment for Clinically Localised Prostate Cancer: A Systematic Review. *Eur. Urol.* **72**, 869–885 (2017).
2. O'Leary, E., Drummond, F. J., Gavin, A., Kinnear, H. & Sharp, L. Psychometric evaluation of the EORTC QLQ-PR25 questionnaire in assessing health-related quality of life in prostate cancer survivors: a curate's egg. *Qual. Life Res.* **24**, 2219–2230 (2015).
3. Ritch, C. R. *et al.* Biochemical recurrence-free survival after robotic-assisted laparoscopic vs open radical prostatectomy for intermediate- and high-risk prostate cancer. *Urology* **83**, 1309–1315 (2014).
4. Bolenz, C. *et al.* Cost Comparison of Robotic, Laparoscopic, and Open Radical Prostatectomy for Prostate Cancer. *Eur. Urol.* **57**, 453–458 (2010).
5. Wilson, T. & Torrey, R. Open versus robotic-assisted radical prostatectomy: which is better? *Curr. Opin. Urol.* **21**, 200–5 (2011).
6. Basto, M. *et al.* Patterns-of-care and health economic analysis of robot-assisted radical prostatectomy in the Australian public health system. *BJU Int.* **117**, 930–939 (2016).
7. Ramsay, C. *et al.* Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol. Assess.* **16**, 313 (2012).
8. Health Quality Ontario. Ontario health technology assessment series: Robotic surgical system for radical prostatectomy: A health technology assessment. *Ont. Health Technol. Assess. Ser.* **17**, 1–172 (2017).
9. Ilic, D. *et al.* Laparoscopic and robot-assisted vs open radical prostatectomy for the treatment of localized prostate cancer: a Cochrane systematic review. *BJU Int.* **121**, 845–853 (2018).
10. Nyberg, M. *et al.* Functional and Oncologic Outcomes Between Open and Robotic Radical Prostatectomy at 24-month Follow-up in the Swedish LAPPRO Trial. *Eur. Urol. Oncol.* **1**, 353–360 (2018).
11. Baunacke, M. *et al.* Long-term functional outcomes after robotic vs. retropubic radical prostatectomy in routine care: a 6-year follow-up of a large German health services research study. *World J. Urol.* **38**, 1701–1709 (2020).
12. Herlemann, A., Cowan, J. E., Carroll, P. R. & Cooperberg, M. R. Community-based Outcomes of Open versus Robot-assisted Radical Prostatectomy. *Eur. Urol.* **73**, 215–223 (2018).
13. Gershman, B. *et al.* Redefining and Contextualizing the Hospital Volume-Outcome Relationship for Robot-Assisted Radical Prostatectomy: Implications for Centralization of Care. *J. Urol.* **198**, 92–99 (2017).
14. Leow, J. J. *et al.* Systematic Review of the Volume-Outcome Relationship for Radical Prostatectomy. *Eur. Urol. Focus* **4**, 775–789 (2018).
15. Coughlin, G. D. *et al.* Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: 24-month outcomes from a randomised controlled study. *Lancet Oncol.* **19**, 1051–1060 (2018).
16. Wallerstedt, A. *et al.* Quality of Life After Open Radical Prostatectomy Compared with Robot-assisted Radical Prostatectomy. *Eur. Urol. Focus* **5**, 389–398 (2019).
17. European Association of Urology. *Guidelines on Prostate cancer 6.1.2.3. Surgical techniques*. (2020).
18. Szymanski, K. M., Wei, J. T., Dunn, R. L. & Sanda, M. G. Development and Validation of an Abbreviated Version of the Expanded Prostate Cancer Index Composite Instrument for Measuring Health-related Quality of Life Among Prostate Cancer Survivors. *Urology* **76**, 1245–1250 (2010).
19. Clavien, P. A. *et al.* The Clavien-Dindo Classification of Surgical Complications. *Ann. Surg.* **250**, 187–196 (2009).
20. D'Amico, A. V *et al.* Biochemical Outcome After Radical Prostatectomy, External Beam Radiation Therapy, or Interstitial Radiation Therapy for Clinically Localized Prostate Cancer. *JAMA* **280**, 969 (1998).
21. Skolarus, T. A. *et al.* Minimally Important Difference for the Expanded Prostate Cancer Index Composite Short Form. *Urology* **85**, 101–106 (2015).
22. Nossiter, J. *et al.* Robot-assisted radical prostatectomy vs laparoscopic and open retropubic radical prostatectomy: functional outcomes 18 months after diagnosis from a national cohort study in England. *Br. J. Cancer* **118**, 489–494 (2018).
23. Haglind, E. *et al.* Urinary Incontinence and Erectile Dysfunction After Robotic Versus Open Radical Prostatectomy: A Prospective, Controlled, Nonrandomised Trial. *Eur. Urol.* **68**, 216–225 (2015).

24. Koike, H. *et al.* Health-related quality of life after robot-assisted radical prostatectomy compared with laparoscopic radical prostatectomy. *J. Robot. Surg.* **11**, 325–331 (2017).
25. Sanda, M. G. *et al.* Quality of Life and Satisfaction with Outcome among Prostate-Cancer Survivors. *N. Engl. J. Med.* **358**, 1250–1261 (2008).
26. Guercio, C. & Mehta, A. Predictors of Patient and Partner Satisfaction Following Radical Prostatectomy. *Sex. Med. Rev.* **6**, 295–301 (2018).
27. Hamdy, F. C. *et al.* 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N. Engl. J. Med.* **375**, 1415–1424 (2016).
28. Resnick, M. J. *et al.* Long-Term Functional Outcomes after Treatment for Localized Prostate Cancer. *N. Engl. J. Med.* **368**, 436–445 (2013).
29. Secin, F. P. *et al.* The Learning Curve for Laparoscopic Radical Prostatectomy: An International Multicenter Study. *J. Urol.* **184**, 2291–2296 (2010).
30. Barbash, G. I. & Glied, S. A. New Technology and Health Care Costs — The Case of Robot-Assisted Surgery. *N. Engl. J. Med.* **363**, 701–704 (2010).

SUPPLEMENTARY MATERIAL

Supplement 1 - Recoding variables for statistical analysis

With the recoding step, we aimed to have a minimum number of missing values. We checked the influence of recoding by comparing our results with recoding to the results where the additional answer option was evaluated as missing. The results or conclusions did not alter by using the recoding steps described in this supplement.

The following questions in the EPIC questionnaire had an additional answering option: "not applicable": 28, 29, 30, 31, 33, 49, 50, 52, 53, 54, 57, 58.

The questions 28 till 54 were symptom scales asking how big a certain problem was. When in these questions the answer: "not applicable" was selected it was assumed that the specific problem was "no problem" and it was recoded as "0".

For question 28 (part of the urinary incontinence domain) our recoding was checked with the answer on question 23 ("Over the past 4 weeks, how often have you leaked urine?"). A cross table is presented below (Supplemental Table 1). 342 patients described having leaked urine rarely or never, however, 17 patients filling in "not applicable" described a certain level of urine leakage. Therefore, only the 342 patients having rarely to no urine leakage were recoded to have no problem. For the other 17 patients this question was seen as missing data.

Table 1 Supplement 1

		Question 23: Over the past 4 weeks, how often have you leaked urine?					
		More than once a day	About once a day	More than once a week	About once a week	Rarely to never	Total
Question 28: How big a problem, if any, has each of the following been for you during the last 4 weeks?: Dripping or leaking urine	Big problem	28	1	3	2	2	36
	Moderate problem	52	15	8	3	6	84
	Small problem	59	21	13	11	5	109
	Very small problem	124	88	44	69	53	378
	No problem	50	68	23	71	171	383
	Not applicable	5	4	2	6	342	359
Total		318	197	93	162	579	1349

For questions 57 and 58, part of the Sexual domain, a specific recoding needed to be created as the answer was: "Not applicable, because I was not sexually active". The answer option thus provides more information on the sexual function of the patient.

For question 57: "How would you rate each of the following during the last 4 weeks, your ability to have an erection?" the patients answering: "not applicable" were matched with their answer on question 59:

"How would you describe the usual quality of your erections during the last 4 weeks?". When patients answered "not applicable because I was not sexually active" AND described having no erection at all, their answer to question 57 was recoded to: "very poor to none". 414 of 466 answers were recoded to "Very poor to none".
Below a cross table is presented (Supplemental Table 2) with the answers of the patients on questions 57 and 59.

Table 2 Supplement 1

		Question 59: How would you describe the usual quality of your erections during the last 4 weeks?				
		None at all	Not firm enough for any sexual activity	Firm enough for masturbation and foreplay only	Firm enough for intercourse	Total
Question 57: How would you rate your ability to have an erection during the last 4 weeks?	Very poor to none	288	119	62	1	470
	Poor	2	23	75	10	110
	Fair	1	12	69	36	118
	Good	1	0	26	90	117
	Very good	1	0	1	18	20
	Not applicable because I was not sexually active	414	40	8	4	466
	Missing	1	0	2	0	3
Total		708	194	243	159	1304

The latter 52 answers were matched with a second control question: Question 60: "How would you describe the frequency of your erections during the last 4 weeks?".
Patients that did not have a match with question 59, but that answered "I never had an erection when I wanted one" on question 60 were recoded to "very poor to none" on question 57.
A cross table is presented below (S1.Table 3). 27 answers to question 57 that were "not applicable because I was not sexually active" were recoded according to the information in question 60.
The 25 patients that answered "not applicable because not sexually active" and did not result in a match with either question 59 or 60, were recoded as missing.

Table 3 Supplement 1

		Question 60: How would you describe the frequency of your erections during the last 4 weeks?					
		I never had an erection when I wanted one	I had an erection less than half the time I wanted one	I had an erection about half the time I wanted one	I had an erection more than half the time I wanted one	I had an erection whenever I wanted one	Total
Question 57 recoded: How would you rate your ability to have an erection during the last 4 weeks?	Very poor to none	757	61	14	5	4	841
	Poor	14	38	24	23	8	107
	Fair	4	10	25	52	26	117
	Good	2	1	3	31	80	117
	Very good	2	0	0	1	17	20
	N.A. and missing answer on question 59	2	1	0	0	0	3
	Not matched based on previous matching	27	13	2	1	7	50
Total		808	124	68	113	142	1255

For question 58: "How would you rate each of the following during the last 4 weeks? Your ability to reach orgasm (climax)?"
In total 506/1370 patients answered "not applicable because I was not sexually active". A control question of the EORTC QLQ PR25 was used: "Did you have ejaculation problems (eg dry ejaculation)?" in which we recoded the answer options as follows:

Table 4 Supplement 1

PR25 EORTC 54	EPIC-26 question 58
	Very good
Not at all	Good
A Little	Fair
Quite a bit	Poor
Very much	Very poor to none

Very good was left out of the options.

Since the question from the PR25 questionnaire was only required to complete when patients were sexually active, we had only information for 85/506 patients to recode this question.
To calculate the mean score for the sexual domain, 1 question may be missing.
Therefore we first evaluated the number of full domain scores missing if we would recode the other 421 answers as missing.

Secondly, we evaluated whether patients that answered "not applicable because not sexually active"

differed on other aspects from the patients that were randomly missing or completed these questions. In this second analysis we found that the patients answering "not applicable because not sexually active" had a lower chance to receive a neurovascular bundle preserved procedure, on average were older, received LRP more often, were more often not able to have an erection, scored their sexual function as "bad", and finally evaluated the lack of sexual function as less as a problem than patients that completed the questions and those that randomly did not complete these questions.
Especially because of the three latter findings we thought that evaluating the answers of these 421 patients as missing would neglect valuable information. Therefore, we chose to recode the latter 421 patients that chose the option: "not applicable because sexually not active" to a score of 0 on question 58.

Our results were consistent also when we had recoded the 421 answers as missing:
Estimated marginal means when 421 were recoded as missing: RARP: 29.93 (1.95), LRP: 26.41 (2.25) N= 979 in total. (p=0.195)

Supplement 2 - Clinical characteristics retrieved

The following clinical data was retrieved from the medical records of the hospitals:

- Operation date
- Initial PSA
- Clinical tumor stage
- Clinical Gleason score
- Duration of the operation (skin-to-skin)
- Lymph node dissection performed
- Number of lymph nodes taken
- Number of positive lymph nodes
- Neurovascular bundle preservation (bilateral, unilateral, none)
- Prostate volume (grams)
- Pathologic tumor stage
- Pathologic Gleason score
- Surgical margin status
- Length of stay including the day of surgery
- Biochemical recurrence (BCR) defined by two consecutive PSA measurements of ≥0.2 ng/mL or secondary treatment received, and hormonal and/or radiotherapy treatment received after surgery
- The risk on recurrence (low, intermediate, high) was defined by the D'Amico risk classification[1]

1. D'Amico, A. V et al. Biochemical Outcome After Radical Prostatectomy, External Beam Radiation Therapy, or Interstitial Radiation Therapy for Clinically Localized Prostate Cancer. JAMA 280, 969 (1998).

Supplement 3 - More detailed information on the statistical analysis

Additional information on the statistical analysis

The scoring algorithms used to calculate the domain and sum scores of specific questionnaire are listed below:

For the EPIC-26[1]

For the Summary score of the EORTC QLQ C30 version 3.0[2]

For the EQ5D-5L[3]

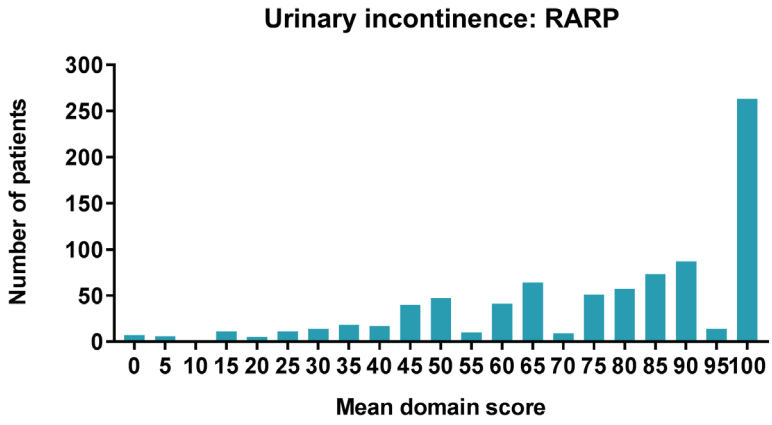
For the EORTC-QLQ-PR25[4]

For the primary analysis, after stepwise inclusion of the confounders as fixed factors. Each of the models was compared using the Bayesian Information Criterion (BIC) and the Akaike's Information Criterion (AIC) after adding a confounder[5,6]. Both criteria are used to compare non-nested models and both penalize the number of model parameters. The BIC also penalizes small sample sizes[7]. Models with lower BIC or AIC values are considered to be better fitting models[8].

All analyses were conducted in SPSS version 25.0.

1. Szymanski, K. M., Wei, J. T., Dunn, R. L. & Sanda, M. G. Development and Validation of an Abbreviated Version of the Expanded Prostate Cancer Index Composite Instrument for Measuring Health-related Quality of Life Among Prostate Cancer Survivors. ***Urology*** **76**, 1245–1250 (2010).
2. Aaronson, N. K. ***et al.*** The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. ***J. Natl. Cancer Inst.*** **85**, 365–376 (1993).
3. Ramos-Goni, J. M. & Rivero-Arias, O. Eq5d: A command to calculate index values for the EQ-5D quality-of-life instrument. ***Stata J.*** **11**, 120–125 (2011).
4. van Andel, G. ***et al.*** An international field study of the EORTC QLQ-PR25: A questionnaire for assessing the health-related quality of life of patients with prostate cancer. ***Eur. J. Cancer*** **44**, 2418–2424 (2008).
5. Schwarz, G. Estimating the Dimension of a Model. ***Ann. Stat.*** **6**, 461–464 (1978).
6. Akaike, H. Information theory and an extension of the maximum likelihood principle. Proceedings of the 2nd international symposium on information theory. in ***Second International Symposium on Information Theory*** (1973).
7. Singer, J. D. & Willett, J. B. ***Applied Longitudinal Data Analysis***. ***Applied Longitudinal Data Analysis*** (2009). doi:10.1093/acprof:oso/9780195152968.001.0001
8. Raftery, A. E. Bayesian Model Selection in Social Research. ***Sociol. Methodol.*** **25**, 111 (1995).

Supplement 4 - Distribution of the domain scores on the EPIC domains: Urinary Incontinence and Sexual



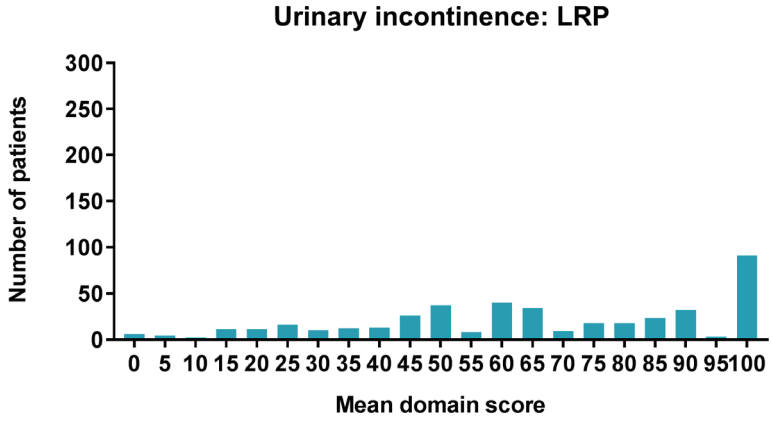


Figure 1A (Supplement 4). shows the distribution on the urinary incontinence domain of the EPIC-26. In the RARP group, 31.1% (N=263) had a perfect score of 100 meaning having no problems with unintentional urinary leakage and no need for wearing pads, whereas in the LRP group 21.5% (n=91) of the patients had a perfect score.

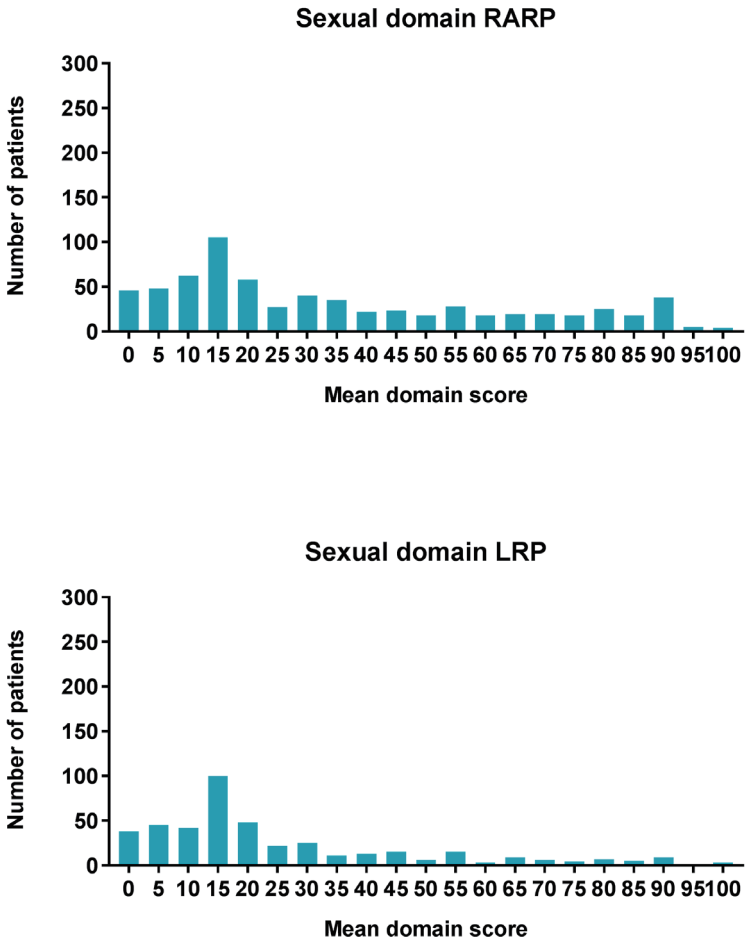


Figure 1B (Supplement 4). Having a perfect score on the sexual domain was rare (RARP: 0.6%, LRP 0.9%). In the RARP group, 9.6% (n=72) patients had a score above 80. In the LRP group, 4.8% (n=17) had a score above 80.

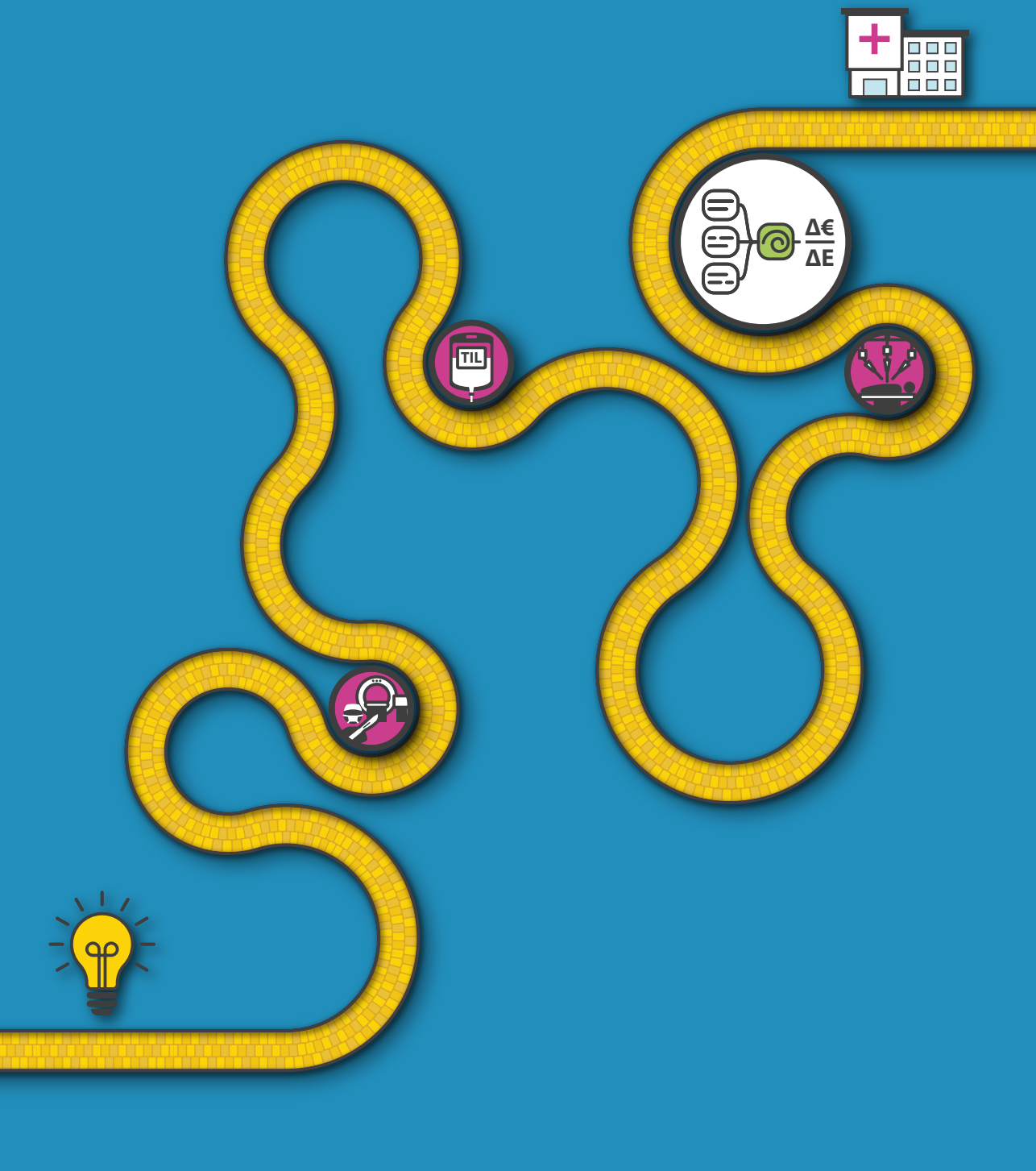
Supplement 5 - Results on two specific questions from the EPIC-26 Urinary Incontinence and Sexual domain to label patients as incontinent and impotent

Table 1 Supplement 5 - shows the uncorrected answers of the study sample on a single item of the EPIC-26 questionnaire related to pad use. We excluded patients that were incontinent at baseline.

How many pads or adult diapers per day did you usually use to control leakage during the last 4 weeks? (EPIC-26 question number 27)	**RARP (n=864)**	**LRP (n=432)**
None (continent)	573 (66.3%)	213 (49.3%)
One pad per day (incontinent)	184 (21.3%)	130 (30.1%)
Two pads per day (incontinent)	65 (7.5%)	48 (11.1%)
Three or more pads per day (incontinent)	42 (4.9%)	41 (9.5%)

Table 2 Supplement 5 - shows the uncorrected answers of the study sample on a single item of the EPIC-26 questionnaire related firmness of the erection for intercourse. We excluded patients that were impotent at baseline.

How would you describe the usual quality of your erections during the last 4 weeks? (EPIC-26 question number 59)	**RARP (n=701)**	**LRP (n=341)**
None at all (impotent)	324 (46.2%)	207 (60.7%)
Not firm enough for any sexual activity (impotent)	102 (14.6%)	49 (14.4%)
Firm enough for masturbation and foreplay only (potent)	158 (22.5%)	57 (16.7%)
Firm enough for intercourse (potent)	117 (16.7%)	28 (8.2%)


TIL
ΔE
ΔE

10

Cost-utility analysis on robot-assisted and laparoscopic prostatectomy based on long-term (6-9 years after surgery) functional outcomes

Melanie Lindenberg
Valesca Retèl
Henk van der Poel
Ferdau Bandstra
Carl Wijburg
Wim van Harten

Submitted

ABSTRACT

Introduction

Robot-assisted radical prostatectomy (RARP) is one of the standard treatment options for prostate cancer. However, controversy still exists on its added value. Based on a recent large-sample retrospective cluster study from the Netherlands showing significantly improved long-term urinary functioning after RARP compared to laparoscopic RP (LRP), we evaluated the cost-effectiveness of RARP compared to LRP.

Methods

A decision tree was constructed to measure the costs and effects from a Dutch societal perspective over a ~7 year time horizon. The input was based on the aforementioned study, including patient-reported consumption of additional care and consumed care for ergonomic issues reported by surgeons. Intervention costs were calculated using a bottom-up costing analysis in 5 hospitals. Finally, a probabilistic-, one-way sensitivity- and scenario analyses were performed to show possible decision uncertainty.

Results

The intervention costs were €9,964 for RARP and €7,253 for LRP. Total trajectory costs were €12,078 for RARP and €10,049 for LRP. RARP showed higher QALYs compared to LRP (6.17 vs 6.11). The incremental cost-utility ratio (ICUR) was €34,206 per QALY gained, in favour of RARP. As a best-case scenario, when RARP is being centralized (>150 cases/year), total trajectory costs decreased to €10,377 having a higher utilization, a shorter procedure time, and shorter length of stay resulting in an ICUR of €3,495.

Discussion

RARP showed to be cost-effective compared to LRP based on data from a population-based, large scale study with 7 years of follow-up. This is a clear incentive to fully reimburse RARP, especially when hospitals provide RARP centralized.

INTRODUCTION

Radical prostatectomy is recommended as a front-line treatment for men diagnosed with localized prostate cancer who have a life expectancy greater than 10 years[1]. In many countries, this procedure is currently performed using robot-assisted radical prostatectomy (RARP), showing improvements compared to open (ORP) and laparoscopic (LRP) radical prostatectomy in urinary incontinence, erectile functioning, hospital stay, and blood loss[2–4], but showing no benefits on oncological outcomes[5]. Additionally, RARP showed improved ergonomics compared to ORP and LRP[6]. However, based on the current evidence base, systematic reviews and meta-analyses concluded that the quality of the evidence is too limited to draw definite conclusions on the advantages of RARP compared to LRP[7–10]. For the Dutch National Health Care Institute and many other national reimbursement bodies, this is the reason to reimburse RARP not for its actual costs but for the costs of ORP or LRP. Therefore, hospitals are faced with substantial additional costs, money that otherwise could be used for improvements in quality of care within a hospital.

Aiming at filling this research gap, a retrospective cluster study was conducted evaluating real-world data from 12 hospitals in the Netherlands (n=1370) to evaluate long-term (median follow-up of 7.08 years) functional and oncologic outcomes and besides evaluate perioperative outcomes, and healthcare usage[11]. This study showed similar prostate cancer-specific survival and oncologic outcomes, but better perioperative outcomes and significantly improved urinary functioning after RARP compared to LRP.

As a part of this retrospective cluster study, the present analysis aimed to comprehensively evaluate the intervention costs of RARP and LRP, and evaluate the cost-effectiveness of RARP compared to LRP from a Dutch societal perspective.

METHODS

Research design and study sample

The design of this study follows the aforementioned retrospective cluster study[11]. In total 1370 patients were included undergoing either RARP or LRP between 2010 and 2012 in 12 hospitals in the Netherlands[11]. In this study, data were collected at one moment in time at least 5 years after surgery.

A decision tree was constructed in Microsoft Excel (Supplement 1) starting with prostate

cancer patients undergoing RARP or LRP. As no significant differences in oncologic outcomes and prostate cancer-specific survival were found[11], the analysis focussed on functional outcomes. After RARP and LRP, patients could end up in the following health states: "continent and potent", "continent and impotent" and "incontinent and impotent".

The analysis was performed from a societal perspective in the Netherlands and the time horizon corresponds with the median follow-up period of 7.08 (range: 5.27 – 9.86) years[11].

Input parameters

All input parameters are presented in Table 1.

Transition probabilities

To define whether a patient ended-up in a certain health state the following definitions were used: patients using no pads (EPIC-26 question 27) were considered continent, patients having a score of ≥17 on the sexual health inventory for men (SHIM) questionnaire were considered potent. Since no cut-off value is known for the EPIC-26 sexual domain (primary outcome of the retrospective cluster study) to define patients having erectile dysfunction, the SHIM questionnaire was also included in the survey[11]. Supplement 2 shows the observed scores on the SHIM. The analysis assumed that patients were in those states for the complete time horizon.

As the combination of being incontinent and potent was not common according to our experts and this group was too small to perform separate analyses on (2.6%), this combination was not taken into account.

We also incorporated the risk of having complications, receiving homecare after surgery, use of additional care for complaints of incontinence and erectile dysfunction directly after surgery (e.g. physiotherapy, sphincter placement) and for a longer period (e.g. pad use and pharmaceuticals)[11].

Utility values

Utilities, values between 0 and 1 where a higher score indicates better health, were evaluated by the EQ5D-5L questionnaire. For each health state, a utility value was calculated (Table 1). The utility value was assumed to be stable over the follow-up period. The utility values were multiplied with the median follow-up time of 7.08 years to obtain the quality adjusted life years (QALYs).

Table 1 - Input parameters for the cost-effectiveness analysis.

Input parameters	RARP		LRP			
Parameter name	**Det value**	**SE**	**Det Value**	**SE**	**Distribution**	**Source**
Probability						
Of being in a certain health state						
"continent and potent"	13.0%	0.012	5.3%	0.011	Dirichlet	[11]
"continent and impotent"	55.6%	0.017	45.1%	0.025	Dirichlet	[11]
"incontinent and impotent"	31.3%	0.016	49.7%	0.025	Dirichlet	[11]
Of having complications after surgery						
Clavien-Dindo grade 1	7.7%	0.010	7.3%	0.014	Beta	[11]
Clavien-Dindo grade 2	4.5%	0.007	2.8%	0.009	Beta	[11]
Clavien-Dindo grade 3	5.1%	0.008	4.5%	0.011	Beta	[11]
Clavien-Dindo grade 4	1.2%	0.004	1.7%	0.007	Beta	[11]
Of receiving home care after surgery						
Receiving home care	1.8%	0.004	3.5%	0.009	Beta	[11]
Hours per week (mean)	7.03	2.54	5.5	2.35	Gamma	[11]
Number of weeks (mean)	4.2	0.85	8.5	3.5	Gamma	[11]
Of receiving additional care for incontinence complaints after surgery						
Receiving physiotherapy	42.5%*	0.023	58.5%*	0.023	Beta	[11]
Number of visits (mean)	7.85	0.315	9.40	0.519	Gamma	[11]
Consulting a General Practitioner (GP)	2.7%*	0.008	3.7%*	0.009	Beta	[11]
Number of visits (mean)	3.40	0.85	5.88	2.38	Gamma	[11]
Sphincter placement	2.5%*	0.007	8.6%*	0.013	Beta	[11]
Of number of pads used in the "incontinent and impotent" health state (measured at follow-up)						
1 pad	61.2%	0.017	61.3%	0.025	Beta	[11]
2 pads	23.3%	0.015	20.9%	0.021	Beta	[11]
3 or more pads	15.5%	0.013	17.8%	0.019	Beta	[11]
Of receiving additional care for complaints of erectile dysfunction after surgery						
Receiving physiotherapy	2.31%	0.005	2.91%	0.008	Beta	[11]
Number of visits (mean)	6.72	1.03	8.67	1.79	Gamma	[11]
Consulting a General Practitioner (GP)	3.6%	0.006	2.4%	0.007	Beta	[11]
Number of visits (mean)	2.88	0.44	2.00	0.23	Gamma	[11]
Consulting a different specialist	10.1%	0.010	13.4%	0.016	Beta	[11]
Number of visits (mean)	3.80	0.78	3.29	0.38	Gamma	[11]
Place a prosthesis	0.3%	0.004	0%	0.000	Beta	[11]
Use a vacuum constriction device	4.1%	0.012	5.3%	0.019	Beta	[11]
Of receiving pharmaceuticals for erectile dysfunction after surgery and during follow-up						
Initial use of an PDE-5 inhibitor	16.7%	0.016	15.5%	0.023	Beta	[11]
Success rate PDE-5 inhibitor	36.6%	0.037	27.4%	0.035	Beta	[11]
Continuous use of an PDE-5 inhibitor†	6.1%	-	4.2%	-	-	-

(Continued on next page)

Table 1 (continued) - Input parameters for the cost-effectiveness analysis.

Input parameters	**RARP**		**LRP**			
Parameter name	**Det value**	**SE**	**Det Value**	**SE**	**Distribution**	**Source**
Initial use of ICI	9.9%	0.004	9.7%	0.022	Beta	[11]
Success rate ICI	70%	0.036	70%	0.036	Beta	Estimation[12]
Continuous use of ICI†	7.0%	-	6.8%	-	-	-
Initial use of IUI	0.5%	0.016	2.5%	0.014	Beta	[11]
Success rate IUI	56%	0.029	56%	0.029	Beta	[13]
Continuous use of an IUI†	0.3%	-	1.4%	-	-	-
Frequency per year (PDE-5 inhibitor)	156	-	156	-	-	EAU Guidelines
Frequency per year (IUI and ICI)	104	-	104	-	-	
Frequency for initial use	5		5			Expert opinion
Utility values	RARP and LRP		SE			
"continent and potent"	0.9638 (n=149)		0.01		Beta	[11]
"continent and impotent"	0.9309 (n=904)		0.00		Beta	[11]
"incontinent and impotent"	0.8964 (n=169)		0.01		Beta	[11]
Cost parameters	RARP Det value	SE	LRP Det value	SE		
Intervention costs†	€9,963.71	€147.24	€7,253.36	€182.62	Gamma	See Table 2
State costs incontinence in "incontinent and impotent" (pad use)#	€2,086.18	€266.09	€2,115.93	€269.89	Gamma	-
State costs for having complaints of erectile dysfunction (medicine use)#	€1,076.89	€137.36	€1,021.48	€130.29	Gamma	-
Cost parameters (unit costs)	**RARP and LRP**		**SE**		**Distribution**	**Source**
Receiving homecare (per hour)	€65.68		€8.38		Gamma	[14]
Costs complication grade 1	€579.39		€73.90		Gamma	Expert opinion [14,15] ‡
Costs complication grade 2	€1,158.79		€147.80		Gamma	
Costs complication grade 3	€3,949.85		€503.81		Gamma	
Costs complication grade 4	€10,760.18		€1,372.47		Gamma	
Costs of one pad	€0.52		€0.07		Gamma	Abena Man (Dutch company)
Physiotherapy consult	€35.24		€4.50		Gamma	[14]
GP consult	€35.24		€4.50		Gamma	[14]
Consult with a specialist	€117.59		€14.99		Gamma	[14]
Sphincer placement	€2,455.00		€313.14		Gamma	Δ
Prothesis placement	€10,003.17		€1275.91		Gamma	$
Vacuum constrictor	€250.00		€31.89		Gamma	[16]
PDE-5 inhibitor (50mg)	€8.06		€1.03		Gamma	[17]

(Continued on next page)

Table 1 (continued) - Input parameters for the cost-effectiveness analysis.

Input parameters	**RARP**		**LRP**			
Parameter name	**Det value**	**SE**	**Det Value**	**SE**	**Distribution**	**Source**
ICI (Phentolamine/papaverine 15mg/0.5ml)	€9.80		€1.25		Gamma	[17]
IUI first time (alprostadil, 1000µg)	€15.36		€1.96		Gamma	[17]
ICI (Phentolamine/papaverine 15mg/0.5ml)	€9.80		€1.25		Gamma	[17]
IUI first time (alprostadil, 1000µg)	€15.36		€1.96		Gamma	[17]
Parameters for including use of care by surgeons and sick-leave of surgeons due to pain complaints						
Additional care used by surgeons because of neck and/or back complaints						
% used additional care for pain complaints	7.1%	0.07	21.4%	0.11	Beta	Supplement B and C
Frequency of care used	5	0.64	5	0.64	Gamma	
Total costs for additional care per treatment arm	€294.42		€475.16			Section 2.3
Sick leave of surgeons because of neck and/or back complaints (friction cost method)						
Surgeons with sick leave because of pain complaints	1.0%	0.026	7.1%	0.066	Beta	
Proportion surgical activities of total work activities (% to replace when sick)	20%		20%		-	Supplement C
Duration of sick leave (weeks)	10	1.28	10	1.28	Gamma	
Frequency of sick leave	2	0.26	2	0.26	Gamma	
Friction period	12.6	-	12.6	-	-	[14]
Costs per hour	€124	-	€124	-	-	[14]
Friction costs	€99,111.2	-	€99,111.2	-	-	Replaced for 20%
Total costs per treatment arm	€991.1		€7,076.5			Section 2.3
Additional modelling parameters						
Discounting rate costs	0.04					[14]
Discounting rate QALYs	0.015					[14]
Time horizon	7.08 years					[11]

Abbreviations: Det. Value = deterministic value, ICI = intra-cavernous injection, IUI = intra-urethral injection, GP = general practitioner.

* Shows the percentages of patients that used an additional type of care of the whole population. For this purpose, the percentages based on Lindenberg et al 2020 (Table 1) describing the use of additional care and having complaints were multiplied.

† More information on calculation of this parameter is presented in Table 2.

These costs are the result of combining the percentages of pads use per intervention and the unit costs, and combining the percentages of continuous use (initial use multiplied with the success rate) of an PDE-5 inhibitor, ICI, IUI with the unit costs of the pharmaceuticals.

† continuous use was found by multiplying the initial use times the success rates.

‡ DRG code 182199024, Δ DRG code 149999079, $DRG code 149899005. The costs for the DRGs were retreived from https://www.opendisdata.nl/msz/zorgproduct (Dutch website).

Surgeon effects

As part of the retrospective study, a questionnaire (Supplement 3) was distributed among surgeons (n=20) that operated in the selected hospitals between 2010 and 2012 evaluating complaints of back and neck pain after or related to LRP and RARP. Supplement 4 shows the results of the questionnaire, and Supplement 5 describes how these effects were translated in monetary values to incorporate the effects in the analysis per treatment arm.

Intervention costs

The intervention costs were evaluated bottom-up by an Activity-Based Costing (ABC) analysis in 5 hospitals, 2 performing LRP, and 3 performing RARP[18]. The following cost categories were included: personnel, material, use of the OR, medical devices, hospitalization, and overhead costs. Because an additional lymph node dissection (LND) resulted in a longer procedure time, and the percentage differed between interventions[11], the costs were calculated with and without LND. The cost categories personnel, material, and medical devices were evaluated per hospital. The costs for using the OR were based on a previous study from a Dutch perspective[19]. The hospitalization costs were calculated by taking the average length of stay per intervention multiplied with the reference costs for an admission day[14]. Finally, a weighted mean of the intervention costs with and without LND was calculated[11]. Table 2 shows the input parameters for the intervention costs. In Supplement 5 more detailed information for the calculation of several cost categories (e.g. health state costs, homecare costs) is provided.

Costs of additional care directly after surgery

Costs for complications were based on expert opinion and a previous evaluation by National Institute for Health and Care Excellence[15]. For homecare costs, a weighted average of the unit costs for personal care, and nursing care was calculated[14].

For the costs for using additional care for complaints of incontinence and erectile dysfunction after surgery, the activities and/or pharmaceuticals taking into account the duration and/or frequency of activities were linked to unit costs or costs for DRGs which were corrected for inflation[14,21,22] (Table 1). For pharmaceuticals, an initial starting dose of 5 tablets or injections was assumed based on expert opinion.

Health state costs

The health state costs included the use of pads and pharmaceuticals used for erectile dysfunction complaints (see Supplement 5 for more information).

Table 2 - Intervention costs.

Input variables	RARP (95% CI)	LRP (95% CI)	Source
Input for RP without LND			
Procedure time (mean hours)	3.47 (3.37-3.56)	3.61 (3.53-3.69)	[11]
Skin-to-skin procedure time (mean hours)	2.77 (2.68-2.85)	3.06 (2.99-3.12)	[11]
Length of stay (mean days)	3.25 (3.13-3.38)	2.99 (2.86-3.13)	[11]
Input for RP with LND			
Procedure time (mean hours)	3.67 (3.54-3.80)	4.25 (4.07-4.42)	[11]
Skin-to-skin procedure time (mean hours)	2.98 (2.87-3.10)	3.74 (3.60-3.88)	[11]
Length of stay (mean days)	3.24 (3.02-3.45)	4.59 (4.03-5.14)	[11]
Input regardless of with or without LND			
% receiving LND	37.9% (35%-41%)	26.8% (23%-31%)	[11]
Costs of OR usage per hour	€238.20	€238.20	[19]
Personnel costs per hour: Anaesthetist (0.5), Surgeon (1-2), OR assistant (2.2), Medical assistant (1) on average per hour	€323.66	€366.60	Real time observation and [20]
Hospitalization costs per day	€505.32	€505.32	[14]
Intervention costs results	**RARP**	**LRP**	**Source / calculation**
Intervention costs without LND (1 procedure)			
Personnel costs	€1,036.17 (10%)	€1,225.25 (18%)	[20]
Costs for OR usage	€825.90 (8%)	€859.88 (12%)	[19]
Hospitalization costs	€1,643.87 (17%)	€1,512.97 (22%)	[14]
Material costs (e.g. surgical tools, suture material, Da Vinci materials)	€2,786.85 (28%)	€2,417.67 (35%)	LRP: [7]*; RARP: based on internal costs per hospital
Medical devices costs (equipment costs and service costs)	€2,571.22 (26%)	-	Interviews / internal cost information of 3 hospitals
Overhead costs	€1,059.00 (11%)	€918.71 (13%)	[7,14]
Intervention costs with LND (1 procedure)			
Personnel	€1,103.53 (11%)	€1,459.54 (18%)	[20]
Costs for OR usage	€874.53 (9%)	€1,011.32 (12%)	[19]
Hospitalization costs	€1,635.30 (16%)	€2,317.08 (29%)	[14]
Material costs (e.g. surgical tools, suture material, Da Vinci materials)	€2,786.85 (28%)	€2,417.67 (30%)	For LRP: [7]*; RARP based on internal costs
Medical devices costs (equipment costs and service costs)	€2,571.22 (26%)	-	Internal cost information
Overhead costs	€1,059.00 (11%)	€918.71 (11%)	[7,14]
Total costs without LND	€9,923.01	€6,934.48	
Total costs with LND	€10,030.42	€8,124.32	
Total costs per intervention (used in the CUA)	**€9,963.71**	**€7,253.36**	

* Exchange rate from pound to euro of €1.23 (average rate of 2012) costs were corrected for inflation (1.105 from 2012 to 2019).

Analysis and sensitivity analyses

In the analysis, the costs were discounted at a rate of 4%, and effects at a rate of 1.5% according to Dutch guidelines23. The outcome of the decision tree is the incremental cost-utility ratio (ICUR) calculated by dividing the incremental costs by the incremental QALYs. Furthermore, a deterministic sensitivity analysis (DSA) and a probabilistic sensitivity analysis (ProbSA) were performed to evaluate the impact of parameter uncertainty. For the DSA, all parameters were varied over their upper and lower limits to evaluate the impact on the ICUR. Besides, two different definitions of having no erectile dysfunction (SHIM>22) and being continent (0-1 pad used) were evaluated.

For the ProbSA, Table 1 shows the distributions used for the parameters in the Monte Carlo simulation (drawing 1000 random samples). All potential outcomes are plotted in a cost-effectiveness (CE-) plane. Furthermore, cost-effectiveness acceptability curves (CEAC) were drafted, indicating the probability that RARP is cost-effective compared to LRP given a certain willingness to pay (WTP) ratio. In the Netherlands, the informal WTP ratio is €80,000 per QALY[24].

Scenario analysis

Finally, in a scenario analysis, three scenarios were evaluated. The first scenario evaluated the best-case scenario (centralization) by evaluating data from the two hospitals performing >150 RARPs per year, including potential effects on clinical outcomes. Supplement 6 shows the detailed calculation and input used for this scenario. In the second scenario, the same intervention costs were included but the potentially improved clinical outcomes were not taken into account as the accompanied study showed no linear relationship between hospital volume and improved functional outcomes[11]. In the third scenario, the Da Vinci robot was also used for other indications, evaluating the ICUR over a range of 100 to 850 procedures a year, by only adjusting the medical device costs.

RESULTS

Base case analysis results

Total intervention costs were €9,964 for RARP and €7,253 for LRP. The categories medical devices (26%) and material (28%) contributed the most to the intervention costs of RARP. For LRP, the categories material (30%-35%), personnel (18%), and hospitalization (22%-29%) contributed the most.

Total trajectory costs were €12,078 for RARP and €10,049 for LRP. Regarding the follow-up costs, incontinence complaints accounted for the largest difference between LRP and RARP (€629) (Table 3A). Total QALYs found for RARP were 6.17 and 6.11 after LRP. Showing incremental costs of €2,029 and incremental QALYs of 0.059 for RARP. RARP shows to be cost-effective at an ICUR of €34,206 as this is below the informal WTP threshold of €80,000 (Table 3B).

Table 3 - The deterministic results of the base case and scenario analysis.

A. Costs per category resulting from the base case analysis (per patient) (not discounted)

	RARP	LRP	Difference RARP - LRP
Surgery	€9,963.71	€7,253.36	€2,710.36
Complications after surgery	€426.61	€439.05	€-12.43
Home care after surgery	€34.91	€107.22	€-72.31
Additional care and sick leave of surgeons	€2.33	€18.57	€-16.25
Incontinence complaints after surgery	€181.54	€412.73	€-231.19
Complaints regarding erectile dysfunction after surgery	€108.16	€89.59	€18.57
Costs for being incontinent over the total time horizon	€653.68	€1,051.55	€-397.87
Costs for having complaints related to erectile dysfunction over the total time horizon	€936.57	€967.83	€-31.26
Not discounted total costs	€12,307.52	€10,339.90	€1,967.62
Discounted total costs	€12,078.01	€10,048.73	€1,490.55

B. Results from the base case and scenario analysis

	Costs	QALYs	iCosts	iQALY	ICUR
Deterministic results from the base case analysis					
RARP	€12,078.01	6.17			
LRP	€10,048.73	6.11			
			€2,029.28	0.059	€34,206.26
Deterministic results from the centralization scenario (scenario 1)					
RARP	€10,377.21	6.20			
LRP	€10,048.73	6.12			
			€328.48	0.094	€3,495.36
Deterministic results from the centralization scenario without taking into account potential clinical benefits (scenario 2)					
RARP	€10,599.91	6.17			
LRP	€10,048.73	6.11			
			€551.18	0.059	€9,290.88

For the scenario analysis, both a scenario analysis evaluating a centralization scenario with and without potential clinical improvements because of centralization of care were presented.

Sensitivity analyses

Figure 1 shows that the ICUR was most sensitive to uncertainty surrounding the utility values, intervention costs, and the two other definitions used. Although using another definition for incontinence (€44,596) and erectile dysfunction (€42,867) would show a substantial higher ICUR, it did not alter our conclusion. Uncertainty surrounding other parameters such as surgeon effects and additional care used for incontinence and erectile dysfunction had a limited effect.

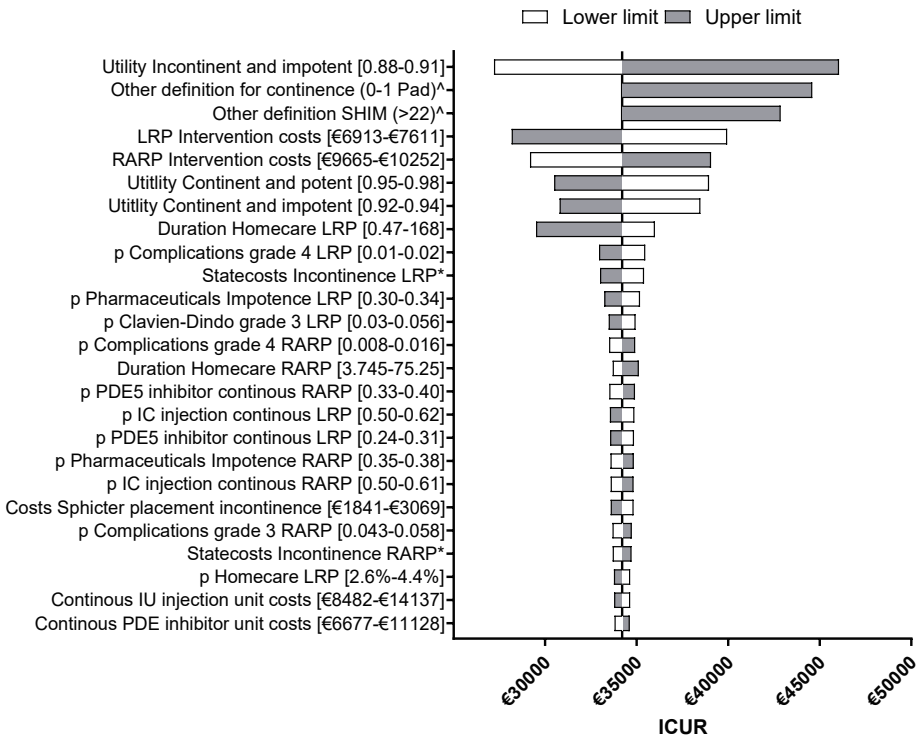


Figure 1. Results from the one-way sensitivity analysis. This figure shows the influence of the observed uncertainty (lower and upper value) surrounding a specific parameter on the main outcome measure. All parameters starting with a "p" indicate a probability. From this figure we learn that the uncertainty surrounding the intervention costs, definitions, and utility values showed the largest deviation from the base case ICUR. However this uncertainty does not affect our conclusion. **Abbreviations:** ICUR = incremental cost-utility ratio.
* the uncertainty from this parameter was a combined value, the uncertainty surrounding the chance of using 1, 2 and 3 or more pads were changed at the same time. The SE surrounding these parameters can be found in Table 1.

Scenario analyses

Table 3B shows the results of scenario 1 and 2. Total trajectory costs of scenario 1 were €10,377 and we found 6.20 QALYs for RARP, resulting in an ICUR of €3,495. For scenario 2, we found total trajectory costs of €10,600 and 6.17 QALYs, resulting in an ICUR of €9,291.

Figure 3 shows that when a hospital performs ≥250 procedures with the Da Vinci robot, the ICUR comes below €20,000, when a hospital has ≥800 procedures a year, RARP is becoming cost-saving compared to LRP.

The ProbSA showed that all possible outcomes indicate that RARP is more effective at higher costs (Figure 2). According to the CEAC, RARP had a 99.8% probability to become cost-effective at a WTP threshold of €80,000.

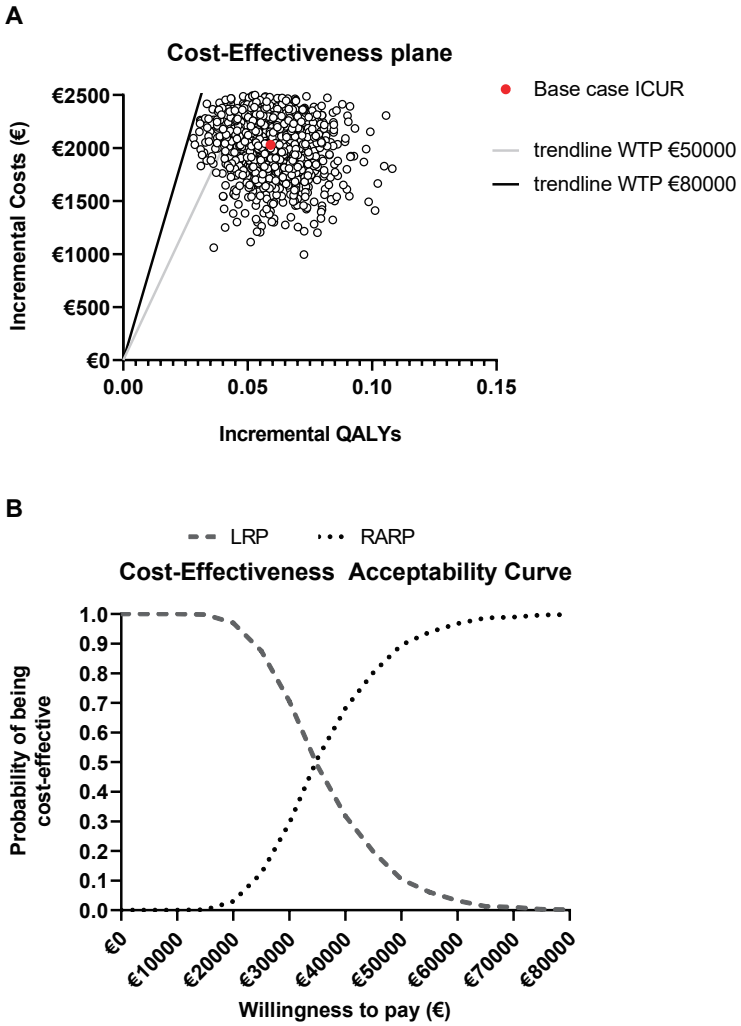


Figure 2. The results from the probabilistic sensitivity analysis. Figure 2A presents all potential outcomes given the distribution surrounding the parameter. The trend lines show the WTP thresholds. All potential outcomes are below the WTP threshold of €80,000. The majority of outcomes also fall below the WTP threshold of €50,000. Figure 3B shows the probability of RARP being cost-effective, given a certain WTP threshold. The probability of RARP being cost-effective at a WTP threshold of €80,000 is 99.8%.

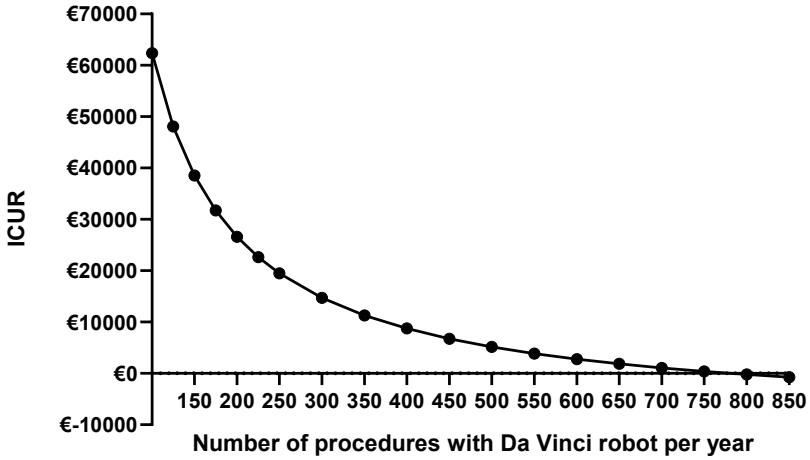


Figure 3. Results from scenario 3. Figure 3 Presents the incremental cost-utility ratio (ICUR) when the Da Vinci is used more often. For example when also used for other indications. Showing an ICUR below €20,000 when ≥250 procedures are performed per year with the robot. When fully optimizing the robot, RARP even shows the potential to be cost-saving compared to LRP.

DISCUSSION

RARP showed to be cost-effective compared to LRP when evaluating long-term functional outcomes, presenting an ICUR of €34,206. These results strengthen the conclusions from the clinical study showing that RARP was more effective compared to LRP on the long-term[11].

The costs found for RARP (€9,964) and LRP (€7,253) were in line with previously published estimates[25,26]. Compared to LRP, the OR costs, personnel costs, and hospitalization costs were lower for RARP due to shorter procedure times and length of stay. In evaluating the intervention costs of RARP we created a rather negative scenario by assuming the use of the Da Vinci robot only for prostatectomies, although many hospitals use the robot in multiple indications where it also suggests to be cost-effective[27,28]. When increasing the utilization of the robot, the ICUR decreased substantially because of lower per-patient costs as seen in the scenario analysis. Based on our data, centralization of RARP (Table 3B) resulted in a decreased length of stay, shorter procedure times, and better outcomes, as has been suggested by literature[29]. We should mention that these scenarios represent a best case example: results from a large volume hospital (>150 procedures/year) and experienced surgeons, showing ICURs between €3,495-€9,291. The effect of centralization on the cost-effectiveness may even be underestimated because we evaluated data from the early introduction phase of the Da Vinci robot[30] and outcomes are expected to improve with surgeon experience[31,32].

Finally, as the material costs are a large driver of the intervention costs, critical appraisal of the instruments used per surgery may be useful. This could result in a cost reduction of ~€250 per surgery[33], with substantial influence on the cost-utility (Figure 1).

The influence of surgeon effects on the cost-effectiveness was limited, although surgeons experienced substantially more pain complaints after LRP compared to RARP (69% vs 21%) (Supplement 3). As similar attempts to incorporate ergonomic differences of interventions on physicians in cost-effectiveness analyses are scarce, we (pragmatically) translated the costs per surgeon having sick leave to costs per patient. In this method the costs for one surgeon having sick leave was divided over ±38 patients. Although we used the most common approach to incorporate ergonomic effects as financial effect[34], it could be argued that our approach underestimates its impact, especially when one would adopt a hospital perspective.

Our findings and conclusions seem to be in line with previous literature showing that RARP was more costly ($7,504-$9,737) compared to LRP ($6,320-$10,991), resulting in ICURs ranging between $28,801-$31,673[26]. Comparison with the findings from another review (including 38 cost-effectiveness studies) was more challenging because in these studies various methods were used to incorporate the costs (e.g. evaluation of the costs based on cost-to-charge ratios or hospital charges) and/or authors only presented incremental costs or savings[10]. However, in general, their results seem to point in the same direction: RARP could be cost-saving when optimal outcomes can be achieved, and the medical equipment is optimally used[10]. Yet, we should note that when the cost-effectiveness of RARP was compared to ORP, RARP is expected to show a smaller chance to be cost-effective, as the costs of ORP are lower compared to LRP[10,26], but outcomes are expected to be similar to LRP[35].

The strength of the present analysis is that it is the first analysis comparing RARP to LRP using long-term functional outcome data and incorporating additional care for complaints of incontinence and erectile dysfunction. Besides, this is one of the few analyses adopting a societal perspective[10], and as far as we know, the first analysis incorporating costs related to homecare and ergonomic complaints of surgeons. A final strength is the bottom-up cost analysis of the intervention and follow-up costs as this provides an accurate and transparent overview of the costs[36].

Several limitations should be acknowledged. First, the generalizability of our results may be limited by the focus on the Dutch healthcare system. We, therefore, presented all cost input parameters transparently to enable calculation of reliable estimates for other countries as well. Furthermore, the cost-effectiveness of RARP may be underestimated because we had no data on the recovery of functional outcomes in the years after surgery, and the recovery duration was suggested to be in favor of RARP[37,38]. Also we did not include costs of hormonal therapy,

although a higher proportion of patients received hormonal treatment after LRP compared to RARP[11]. Contrary, the functional outcomes found for LRP could be underestimated due to the chosen time frame, since the larger hospitals – having more advanced urologists on average – are expected to have shifted earlier to RARP. However, incorporating several confounders in the clinical analysis, did not alter our conclusion[11], for which we are confident that our results point in the right direction.

Conclusion

We conclude that RARP is cost-effective compared to LRP when evaluating long-term health and economic effects at most acceptable WTP ratios. When RARP is centralized and surgeons are experienced with the Da Vinci robot and/or the Da Vinci robot is used in multiple indications, RARP becomes cost-effective at all WTP ratios and has the potential to be cost-saving. Therefore, our results are a clear incentive to fully reimburse RARP, especially when hospitals provide RARP centralized.

ACKNOWLEDGEMENTS

We want to acknowledge the five hospitals that agreed to participate in the cost analysis. Furthermore, we want to thank Intuitive Surgical for providing a research fund (round of 2019-2020) to perform the current study.

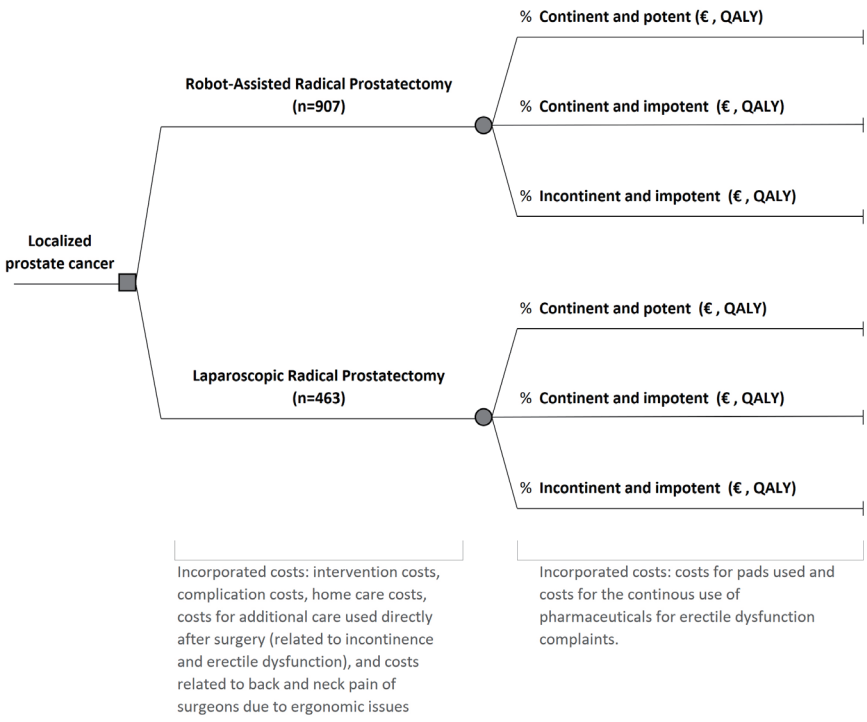
REFERENCES

1. Hamdy, F. C. *et al.* 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N. Engl. J. Med.* **375**, 1415–1424 (2016).
2. Basto, M. *et al.* Patterns-of-care and health economic analysis of robot-assisted radical prostatectomy in the Australian public health system. *BJU Int.* **117**, 930–939 (2016).
3. Nyberg, M. *et al.* Functional and Oncologic Outcomes Between Open and Robotic Radical Prostatectomy at 24-month Follow-up in the Swedish LAPPRO Trial. *Eur. Urol. Oncol.* **1**, 353–360 (2018).
4. Herlemann, A., Cowan, J. E., Carroll, P. R. & Cooperberg, M. R. Community-based Outcomes of Open versus Robot-assisted Radical Prostatectomy. *Eur. Urol.* **73**, 215–223 (2018).
5. Ritch, C. R. *et al.* Biochemical recurrence-free survival after robotic-assisted laparoscopic vs open radical prostatectomy for intermediate- and high-risk prostate cancer. *Urology* **83**, 1309–1315 (2014).
6. Bagrodia, A. & Raman, J. D. Ergonomic Considerations of Radical Prostatectomy: Physician Perspective of Open, Laparoscopic, and Robot-Assisted Techniques. *J. Endourol.* **23**, 627–633 (2009).
7. Ramsay, C., Pickard, R., Robertson, C., Close, A. & Vale, L. Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol. Assess. (Rockv).* **16**, 313 (2012).
8. Health Quality Ontario. Ontario health technology assessment series: Robotic surgical system for radical prostatectomy: A health technology assessment. *Ont. Health Technol. Assess. Ser.* **17**, 1–172 (2017).
9. Ilic, D. *et al.* Laparoscopic and robot-assisted vs open radical prostatectomy for the treatment of localized prostate cancer: a Cochrane systematic review. *BJU Int.* **121**, 845–853 (2018).
10. Schroeck, F. R. *et al.* Cost of New Technologies in Prostate Cancer Treatment: Systematic Review of Costs and Cost Effectiveness of Robotic-assisted Laparoscopic Prostatectomy, Intensity-modulated Radiotherapy, and Proton Beam Therapy. *Eur. Urol.* **72**, 712–735 (2017).
11. Lindenberg, M. A. *et al.* Long-term functional outcomes after robot-assisted prostatectomy compared to laparoscopic prostatectomy: results from a national retrospective cluster study (submitted). (2020).
12. Coombs, P. G., Heck, M., Guhring, P., Narus, J. & Mulhall, J. P. A review of outcomes of an intracavernosal injection therapy programme. *BJU Int.* **110**, 1787–1791 (2012).
13. Guay, A. T., Perez, J. B., Velásquez, E., Newton, R. A. & Jacobson, J. P. Clinical Experience with Intraurethral Alprostadil (MUSE®) in the Treatment of Men with Erectile Dysfunction. *Eur. Urol.* **38**, 671–676 (2000).
14. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. *Manual for cost research: methods and standard cost prices for economic evaluations in health care*. (2015).
15. Ramsay, C. *et al.* Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol. Assess.* **16**, 313 (2012).
16. University hospital Ghent (Belgium). *Erectiele dysfunctie (erectile dysfunction) pateint brochure*.
17. Zorginstituut Nederland (Dutch institute of healthcare). Medicijnkosten (costs of pharmaceuticals). (2019). Available at: http://www.medicijnkosten.nl/. (Accessed: 2nd April 2020)
18. Lievens, Y., van den Bogaert, W. & Kesteloot, K. Activity-based costing: a practical model for cost calculation in radiotherapy. *Int. J. Radiat. Oncol.* **57**, 522–535 (2003).
19. Patel, S. *et al.* Understanding the Costs of Surgery: A Bottom-Up Cost Analysis of Both a Hybrid Operating Room and Conventional Operating Room. *Int. J. Heal. Policy Manag.* (2020). doi:10.34172/ijhpm.2020.119
20. Dutch Federation of Academic Medical Centers. *Collective labor agreement 2018-2020 for academic medical centers*. (2018).
21. StatLine. Dutch price index numbers. (2020). Available at: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?ts=1534249924434. (Accessed: 3rd February 2020)
22. Dutch Healthcare Authority (NZa). Open data from the Dutch health authority on DRGs (title is translated). Available at: https://www.opendisdata.nl/. (Accessed: 15th March 2020)
23. Zorginstituut Nederland (Dutch institute of healthcare). *Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg*. (2016).
24. Versteegh, M. M. *et al.* Severity-Adjusted Probability of Being Cost Effective. *Pharmacoeconomics* **37**, 1155–1163 (2019).

25. Close, A. *et al.* Comparative Cost-effectiveness of Robot-assisted and Standard Laparoscopic Prostatectomy as Alternatives to Open Radical Prostatectomy for Treatment of Men with Localised Prostate Cancer: A Health Technology Assessment from the Perspective of the UK Natio. *Eur. Urol.* **64**, 361–369 (2013).
26. Tandogdu, Z., Vale, L., Fraser, C. & Ramsay, C. A Systematic Review of Economic Evaluations of the Use of Robotic Assisted Laparoscopy in Surgery Compared with Open or Laparoscopic Surgery. *Appl. Health Econ. Health Policy* **13**, 457–467 (2015).
27. Kukreja, J. B. *et al.* Cost-Effectiveness of Robot-assisted Radical Cystectomy Using a Propensity-matched Cohort. *Eur. Urol. Focus* **6**, 88–94 (2020).
28. Ho, C. *et al. Robot-Assisted Surgery Compared with Open Surgery and Laparoscopic Surgery: Clinical Effectiveness and Economic Analyses*. *Canadian Agency for Drugs and Technologies in Health* (2011).
29. Xia, L. *et al.* Associations Between Hospital Volume and Outcomes of Robot-Assisted Radical Prostatectomy. *J. Urol.* **203**, 926–932 (2019).
30. Abrishami, P., Boer, A. & Horstman, K. Understanding the adoption dynamics of medical innovations: Affordances of the da Vinci robot in the Netherlands. *Soc. Sci. Med.* **117**, 125–133 (2014).
31. Fossati, N. *et al.* Assessing the Impact of Surgeon Experience on Urinary Continence Recovery After Robot-Assisted Radical Prostatectomy: Results of Four High-Volume Surgeons. *J. Endourol.* **31**, 872–877 (2017).
32. Bravi, C. A. *et al.* The Impact of Experience on the Risk of Surgical Margins and Biochemical Recurrence after Robot-Assisted Radical Prostatectomy: A Learning Curve Study. *J. Urol.* **202**, 108–113 (2019).
33. Ramirez, D., Ganesan, V., Nelson, R. J. & Haber, G.-P. Reducing Costs for Robotic Radical Prostatectomy: Three-instrument Technique. *Urology* **95**, 213–215 (2016).
34. Tompa, E., Dolinschi, R., De Oliveira, C., Amick, B. C. & Irvin, E. A systematic review of workplace ergonomic interventions with economic analyses. *Journal of Occupational Rehabilitation* (2010). doi:10.1007/s10926-009-9210-3
35. Nossiter, J. *et al.* Robot-assisted radical prostatectomy vs laparoscopic and open retropubic radical prostatectomy: functional outcomes 18 months after diagnosis from a national cohort study in England. *Br. J. Cancer* **118**, 489–494 (2018).
36. Najjar, P. A., Strickland, M. & Kaplan, R. S. Time-Driven Activity-Based Costing for Surgical Episodes. *JAMA Surg.* **152**, 96 (2017).
37. Koike, H. *et al.* Health-related quality of life after robot-assisted radical prostatectomy compared with laparoscopic radical prostatectomy. *J. Robot. Surg.* **11**, 325–331 (2017).
38. Huang, X., Wang, L., Zheng, X. & Wang, X. Comparison of perioperative, functional, and oncologic outcomes between standard laparoscopic and robotic-assisted radical prostatectomy: a systemic review and meta-analysis. *Surg. Endosc.* **31**, 1045–1060 (2017).
39. Nederlandse Vereniging van Ziekenhuizen (NVZ). *Salarisschalen Cao Ziekenhuizen (2018) (Dutch collective labor agreement)*. (2018).
40. Lindenberg, M. A. *et al.* Understanding the costs of surgery: a bottom-up cost analysis of both a hybrid and conventional operating room (manuscript has been accepted beginning of July 2020). *Int. J. Heal. Policy Manag.* (2020).
41. Dutch Healthcare Authority (NZa). DBC-zorgproduct (DRG) 182199024. *2018* Available at: https://www.opendisdata.nl/msz/zorgproduct/182199024? (Accessed: 15th March 2020)

SUPPLEMENTARY MATERIAL

Supplement 1 - Schematic overview of the model



Supplement 2 - Observed SHIM scores

SHIM scores	RARP (n=840) Missing: 67 (7.4%)	LRP (n=407) Missing n=56 (12.1%)
Severe erectile dysfunction (1-7)	578 (68.8%)	317 (77.9%)
Moderate erectile dysfunction (8-11)	92 (11.0%)	46 (11.3%)
Mild to moderate erectile dysfunction (12-16)	38 (4.5%)	19 (4.7%)
Mild erectile dysfunction (17-21)	73 (8.7%)	12 (2.9%)
No erectile dysfunction (22-25)	59 (7.0%)	13 (3.2%)

Supplement 3 - Surgeon questionnaire

Participating hospitals CERA-PRO
Operating clinicians between 2010-2012

You work in: *(name of hospital)* _____

Gender: Man / Woman

Age: _____ Years

Part 1: Experience

1. When was your first robot-assisted prostatectomy (approximately)? _____

2. How much experience do you have with the laparoscopic removal of the prostate? _____ Years / Months*

   _____ Average amount of surgery that you perform(ed) per year / month *

   * delete what does not apply

3. How many robot-assisted procedures do you perform currently on average per year? _____ Per year

4. How many robot-assisted procedures did you perform before 2010? _____ Number of procedures independently before 2010

   _____ Number of procedures while being educated / as a fellow before 2010

Part 2: Ergonomics

5. What proportion of your total work consists of surgical removal of the prostate? ☐ %

6. Do you experience chronical neck or back pain?
   - ◯ No
   - ◯ Yes, pain in the neck
   - ◯ Yes, pain in the back
   - ◯ Yes pain in both: neck and back

7. If yes, do you think that your surgical activities caused the pain or worsen the pain?
   - ◯ No
   - ◯ Yes these activities caused the pain
   - ◯ Yes, these activities worsened the pain

8. Do you or did you experience back or neck pain after a laparoscopic prostatectomy?
   - ◯ No
   - ◯ Yes, neck pain
   - ◯ Yes, back pain
   - ◯ Yes, neck and back pain
   - ◯ No, but other pain namely:

9. Do you or did you experience back or neck pain after a robot-assisted prostatectomy?
   - ◯ No
   - ◯ Yes, neck pain
   - ◯ Yes, back pain
   - ◯ Yes, neck and back pain
   - ◯ No, but other pain namely:

10. Is your preference for the type of OR (laparoscopic or robot-assisted) influenced by physical complaints? For example by neck and or back pain?
    - ◯ No this doesn't matter
    - ◯ Yes a preference because of physical complaints for laparoscopic removal
    - ◯ Yes a preference because of physical complaints for robot-assisted removal

11. If you indicated that you experience pain after performing a laparoscopic prostatectomy, which might have caused chronical complaints or worsened existing complaints:

    A. Did you use care for this?
    - ◯ No
    - ◯ Yes

    If yes, could you specify below what type of care, the frequency and the duration of this type of care? (medication, physiotherapy, occupational therapy etc.)

Frequency: Duration:

B. Did this result in higher sick leave?
- ◯ No
- ◯ Yes

If yes, could you specify (in days or weeks) the extend that the sick leave was worsened in days or weeks?

12. If you indicated that you experience pain after performing a robot-assisted prostatectomy, which might have caused chronical complaints or worsened existing complaints:

A. Did you use care for this?
- ◯ No
- ◯ Yes

If yes, could you specify below what type of care, the frequency and the duration of this type of care? (medication, physiotherapy, occupational therapy etc.)

Frequency: Duration:

B. Did this result in higher sick leave?
- ◯ No
- ◯ Yes

If yes, could you specify (in days or weeks) the extend that the sick leave was worsened in days or weeks?

13. According to you what are other advantages for you or for the patient in using the robot in a prostatectomy?

- For the surgeon:

- For the patient:

- I don't see specific advantages but I see the following disadvantages for me / patient / organization:

You completed the questionnaire. Thank you for your participation!

If you have any other remarks regarding the use of the robot for surgical removal of the prostate you can use the space below.

Supplement 4 - Results of the surgeon questionnaire

Six of the fourteen surgeons that completed the questionnaire agreed that their surgical activities worsened (n=3) or caused (n=3) chronical neck and/or back pain. After LRP, 9 of 13 (69.2%) responding surgeons described to experience back and/or neck pain, 3 surgeons (23.1%) used additional care for their complaints and for one surgeon it resulted in sick leave. After RARP, 3 of 14 surgeons experienced back or neck pain (21.4%), one surgeon (7.1%) used additional care and for no surgeon it resulted in sick leave. In the table below the characteristics or questions and answers are listed.

Supplement 4 Table - Results of the surgeon questionnaire.

Question or characteristic	Numbers of surgeons (%)
Sex	
Male	13
Female	1
Age median (y)	50.5 (40-62)
Year first RALP performed	
2005	1
2007	1
2008	1
2009	1
2010	3
2011	5
2018	1
Missing	1
Years of experience with LRP median (range)	7 (0.5-20)
Number of LRP procedures per year median (range)	40 (5-50)
Number of RALP procedures per year median (range)	80 (0-250)
Current percentage of performing surgeries of total work activities (median, range)	17.5% (0%-40%)
Do you experience chronical neck or back pain?	
No	7 (23.1%)
Yes, pain in the neck	3 (23.1%)
Yes, pain in the back	2 (15.4%)
Yes pain in both: neck and back	1 (7.7%)
If yes, do you think that your surgical activities caused the pain or worsen the pain?	
Yes these activities caused the pain	3 (50%)
Yes, these activities worsened the pain	3 (50%)
Is your preference for the type of OR (laparoscopic or robot-assisted) influenced by physical complaints? For example by neck and or back pain?	
No this doesn't matter	5 (38.5%)
Yes a preference because of physical complaints for laparoscopic removal	1 (7.7%)
Yes a preference because of physical complaints for robot-assisted removal	7 (53.8%)

(Continued on next page)

Supplement 4 Table (continued) - Results of the surgeon questionnaire.

Question or characteristic	LRP	RARP
Do you or did you experience back or neck pain?		
No	4 (30.8%)	11 (78.6%)
Yes, pain in the neck	1 (7.7%)	2 (14.3%)
Yes, pain in the back	2 (15.4%)	0
Yes pain in both: neck and back	6 (46.2%)	1 (7.1%)
No, but other pain namely:	0	0
If you indicated that you experience pain after performing a prostatectomy, did you use care for this?		
Yes	3 (37.5%)	1 (33.3%)
No	5 (62.5%)	2 (66.7%)
If you indicated that you experience pain after performing a prostatectomy did this result in higher sick leave?		
Yes	1 (12.5%)	0
No	7 (87.5%)	3 (100%)

Supplement 5 - More detailed description of the calculated costs

Surgeon effects

To calculate the costs related to complaints or sick leave of surgeons in the model, the percentage of urologists having sick leave (5%) and using care (20%) stratified for after RARP and LRP were linked to costs of additional care and productivity losses by using the friction cost method[14]. These costs were multiplied with the number of main operating urologists in the 12 hospitals between 2010-2012 (n=36) and the percentage of patients receiving LRP (34.8%) and RARP (65.2%) to incorporate the costs for both interventions.

The intervention costs

To calculate the personnel costs, the required personnel was linked to annual loaded salaries based on the collective labor agreements and the Dutch costing manual[1-3]. For the operating surgeon, the costs were multiplied with the average skin-to-skin duration of the procedure. The costs for other staff was multiplied to the average procedure time. For the material costs all materials used, including the specific instruments used for RARP, were identified and linked to internal costs. When the cost-analysis was performed, only one of the two LRP hospitals still performed LRP, therefore the material costs were based on literature[4]. All material costs were corrected for inflation[5]. For the medical device costs, the yearly costs of interest and amortization were calculated by using the purchasing costs of the Da Vinci robot, a depreciation period of 10 years, and an interest rate of 4.2%[1]. These yearly costs were divided by the number of prostatectomies per year in each hospital. Overhead expenses, 38% recommended by the Dutch guideline[1], were only calculated over the material costs to avoid double counting. The costs for using the OR were based on a previous study from a Dutch perspective[6] showing that one hour of OR use costs €238.20 without personnel. These costs were multiplied with the average total procedure time.

Complication costs

Costs for Grade 1 and 2 consisted of 1 and 2 additional nursing days respectively. For Grade 3, 3 additional nursing days and a general surgery to treat e.g. urine leakage were charged[7]. For a Grade 4 complication, we charged 5 ICU days including diagnostic activities[14]. These numbers were multiplied with the probability of having a Grade 1 to 4 complication after RARP and LRP.

Home care costs

For homecare costs, a weighted average of the unit costs for personal care, and nursing care was calculated[1]. These costs were multiplied to the number of patients not receiving homecare before surgery, the average duration, and average frequency[8].

Health state costs

To calculate the costs related to being in a certain health state, the evaluated activities were linked to unit costs and the time horizon. For patients being incontinent and impotent, the number of pads used per day at follow-up was evaluated and multiplied with the unit costs and the time horizon. Pad use was assumed to be stable over time. For patients having erectile dysfunction complaints, the percentage of patients initially using specific pharmaceuticals was multiplied with their success rates based on data from the survey and literature, and with the unit cost[9-11]. Based on guidelines, it was assumed that a PDE-5 inhibitor was used 3 times a week and Intra-urethral and Intra-cavernous injections twice a week when successful (Table 1).

Supplement 6 - Detailed information the input parameters for the scenario analysis

In the table below the parameters that were adjusted compared to the original Table 2 are marked in red. In addition, the transition probabilities for RARP were adjusted by evaluating the 458 number of patients that were operated in the two hospitals having a throughput above 150 cases per year. This resulted in: 18.9% in the continent and potent health state, 57.7% incontinent and impotent health state and 21.4% in incontinent and impotent health state. Compared to 13.0%, 55.6% and 31.3% respectively in the base case.

Supplement 6 Table - Intervention costs as used in the scenario analysis.

Intervention costs input (for centralization scenario)			
	LRP (95% CI)	**RARP (95% CI)**	**Source**
Input for RP without LND			
Total procedure time (mean hours)	3.61 (3.53-3.69)	2.65 (2.57-2.72)	[8]
Skin-to-skin procedure time (mean hours)	3.06 (2.99-3.12)	2.04 (1.97-2.11)	[8]
Length of stay (mean days)	2.99 (2.86-3.13)	2.69 (2.55-2.84)	[8] RARP based on 252 patients
Input for RP with LND			
Total procedure time (mean hours)	4.25 (4.07-4.42)	3.05 (2.96-3.14)	[8]
Skin-to-skin procedure time (mean hours)	3.74 (3.60-3.88)	2.44 (2.36-2.52)	[8]
Length of stay (mean days)	4.59 (4.03-5.14)	2.57 (2.41-2.72)	[8] RARP based on 212 patients
Input regardless of with or without LND			
Percentage of patients receiving LND	26.8% (23%-31%)	37.9% (35%-41%)	[8]
Costs of OR usage per hour	€238.20	€238.20	[6]
Personnel costs per hour: Anaesthetist (0.5), Surgeon (1-2), OR assistant (2.2), Medical assistant (1) on average per hour	€366.60	€323.66	Real time observation per hospital; Collective labour agreement
Hospitalization costs per day	€505.32	€505.32	[1]
Intervention costs results			
	LRP	**RARP**	**Source / calculation**
Intervention costs without LND			
Personnel per procedure	€1,225.25	€782.19	Collective labour agreement / internal registry data RARP
OR usage per procedure	€859.88	€630.24	[6]
Hospitalization per procedure	€1,512.97	€1,361.56	[1]
Material costs (e.g. surgical tools, suture material, Da Vinci materials)	€2,417.67	€2,786.85	LRP: [4]*; RARP: based on internal costs per hospital

(Continued on next page)

Intervention costs without LND results (continued)	LRP	RARP	Source / calculation
Medical devices costs (equipment costs and service costs)	-	€1,805.08	Interviews / internal cost information of 3 hospitals
Overhead	€918.71	€1,059.00	[1,4]
Intervention costs with LND			
Personnel per procedure	€1,459.54	€911.49	Collective labour agreements / internal registry data
OR usage per procedure	€1,011.32	€725.95	[6]
Hospitalization per procedure	€2,317.08	€1,296.67	[1]
Material costs (e.g. surgical tools, suture material, Da Vinci materials)	€2,417.67	€2,786.85	For LRP: [4]*; RARP based on internal costs
Medical device costs (equipment costs and service costs)	-	€1,805,08	Internal cost information
Overhead costs	€918.71	€1,059.00	[1,4]
Total costs without LND	€6,934.48	€8,424.92	
Total costs with LND	€8,124.32	€8,585.05	
Total costs per intervention (used in the CUA)	**€7,253.36**	**€8,485.61**	

Intervention costs in detail, as used in the scenario analysis (similar Table as Table 2). * exchange rate from pound to euro of 1.23 EUR (average rate of 2012) costs were corrected for inflation (1.105 from 2012 to 2019).

References used in supplementary material

1. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. Manual for cost research: methods and standard cost prices for economic evaluations in health care. (2015).
2. Nederlandse Vereniging van Ziekenhuizen (NVZ). Salarisschalen Cao Ziekenhuizen (2018) (Dutch collective labor agreement). (2018).
3. Dutch Federation of Academic Medical Centers. Collective labor agreement 2018-2020 for academic medical centers. (2018).
4. Ramsay, C., Pickard, R., Robertson, C., Close, A. & Vale, L. Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. Health Technol. Assess. (Rockv). 16, 313 (2012).
5. StatLine. Dutch price index numbers. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?ts=1534249924434 (2020).
6. Patel, S. et al. Understanding the Costs of Surgery: A Bottom-Up Cost Analysis of Both a Hybrid Operating Room and Conventional Operating Room. Int. J. Heal. Policy Manag. (2020) doi:10.34172/ijhpm.2020.119.
7. Dutch Healthcare Authority (NZa). DBC-zorgproduct (DRG) 182199024. 2018 https://www.opendisdata.nl/msz/zorgproduct/182199024?
8. Lindenberg, M. A. et al. Long-term functional outcomes after robot-assisted prostatectomy compared to laparoscopic prostatectomy: results from a national retrospective cluster study (submitted). (2020).
9. Zorginstituut Nederland (Dutch institute of healthcare). Medicijnkosten (costs of pharmaceuticals). http://www.medicijnkosten.nl/ (2019).
10. Coombs, P. G., Heck, M., Guhring, P., Narus, J. & Mulhall, J. P. A review of outcomes of an intracavernosal injection therapy programme. BJU Int. 110, 1787–1791 (2012).
11. Guay, A. T., Perez, J. B., Velásquez, E., Newton, R. A. & Jacobson, J. P. Clinical Experience with Intraurethral Alprostadil (MUSE®) in the Treatment of Men with Erectile Dysfunction. Eur. Urol. 38, 671–676 (2000).


ΔE
ΔE
TIL
ΔE
ΔE

11

Discussion

DISCUSSION

This dissertation aimed to contribute to the knowledge of applying HTA alongside the translational pathway in oncology by providing real-world examples. In each chapter, we aimed to inform clinical decision making, to steer further research and development (R&D), and/or to inform policy advisors in decisions related to reimbursement. Reflecting at the use of HTA methods during the technical development, we aimed to deliver input for positioning especially early HTA in the comprehensive evaluation of medical technologies and complex innovations during the translational pathway. In this final chapter, the main findings from each chapter are summarized, methodological issues faced and potential directions for future research are discussed, policy recommendations are proposed, and we conclude with final remarks.

Main findings and implications

Part I - Very early HTA

The main aim of using HTA in a very early stage of product development is to steer further R&D), for example by prioritization of potential innovations for further research.

In every phase of the technical development a systematic review can be used. In this dissertation, a systematic review was performed very early in the development stage and was therefore incorporated in the first part of this thesis. Based on the analysis in Chapter 3, we showed that current evidence on the most accurate imaging modality to measure intermediate response on NACT stratified to breast cancer subtypes is of low quality. Furthermore, we identified that before performing subsequent studies that could deliver higher levels of evidence, consensus should be reached on definitions of pathologic complete response, cut-off values, the timing of monitoring, and specific imaging settings. By performing a systematic literature review in the early stage of response-guided NACT we could identify crucial knowledge gaps to inform further research activities and may prevent investing time and assets in setting-up (clinical) studies that will not bring the evidence needed.

Chapter 2 showed the results from the multiple criteria decision analysis (MCDA) focused on the prioritization of research activities regarding image-guided surgical technologies. This application of MCDA is not very common in literature as it is mostly used to inform decisions related to research funding allocations, authorization issues, investment choices, or to identify patient preferences regarding a new technology[1,2]. In our analysis, the evaluation of the importance of the various attributes of surgical success provided valuable insights for R&D purposes as it identified the most important attributes per indication, which are the ones the technology should fulfill to become successful. Besides, the performance evaluation

showed that in general the intraoperative navigation system had the highest expectation. For the removal of tongue and rectal tumors specifically, optical imaging was also preferred. These preferences can support the prioritization of clinical research activities and could increase the chance of receiving research funding.

Based on the results of especially the MCDA, we may conclude that very early HTA could deliver valuable information for R&D purposes and may contribute to more effective innovation. Furthermore, it could even have accelerated patient access by making an informed decision to further investigate the most promising technique, which potentially results in a higher chance of receiving research funding. This part illustrates the benefit that researchers, and potentially also policy-makers, could gain from applying very early HTA.

Part II - Early HTA up to and including the first clinical studies (phase I)

The use of HTA in this stage aims to answer the question of whether a company or clinical research group should continue with further development and (clinical) testing, and aims to guide the potential follow-up steps. Furthermore, it could identify the expected required investments for further R&D activities and inform the medical community, policy advisors, and/or decision-makers on the application of innovations.

In the second part of this dissertation, early HTA was applied during the early development or first clinical tests. The early budget impact analysis (BIA) in Chapter 4 showed that due to lower implementation costs compared to its alternatives, gradual adoption of magnetic seed localization (MSL) in non-palpable breast cancer lesions has the potential to be cost-efficient. As it is an early BIA, some of the parameters were uncertain, e.g. the price of the magnetic seed, showing that when the price of the magnetic seed is substantially higher than the price of a radioactive seed, the adoption of MSL may not be preferred. The development team could use this information in pricing decisions. Furthermore, this analysis identified aspects that are crucial for MSL to become a successful innovation: (1) the retail price of the magnetic seed, (2) enabling response monitoring for patients receiving NACT when using MSL, and/or (3) MSL should show improved clinical outcomes compared to its alternatives.

In Chapter 5 the results from a comprehensive bottom-up cost-analysis of a conventional and hybrid operating room (OR) were presented. Although the hybrid OR has been implemented in multiple Dutch hospitals, we categorized this analysis as early HTA because clinical studies evaluating its (clinical) benefits are still limited. The rapid diffusion of surgical technologies that are expected to contribute to clinical performance but without having firm clinical evidence, is a phenomenon that is seen more often in surgical innovation (e.g. the Da Vinci robot)[3].

The analysis showed that personnel costs for the conventional OR, and the utilization rate for the hybrid OR were the most prominent cost drivers of these ORs. Based on this information

hospital managers could make an informed decision whether or not to adopt a hybrid OR. Besides, it could inform scheduling activities within a hospital. Furthermore, the analysis showed that using the hybrid OR was substantially more expensive than a conventional OR, indicating that an intervention performed in the hybrid OR should be substantially more effective. The additional costs of the hybrid OR can be used to evaluate the expected value for money of performing specific interventions in the hybrid OR, potentially resulting in reimbursement agreements and therefor facilitating patient access.
The final chapter in this part (Chapter 6) showed that the navigation system for removing rectal tumors has the potential to become cost-effective, especially when the system is optimally used. Therefore, this study stresses the importance of identification of additional indications where the navigation system could add value. The results from the MCDA (Chapter 2) could inform this quest. Furthermore, the value of information (VOI) analysis showed that the continuation of research activities is advised. The results form the sensitivity analysis, could inform future clinical study designs as it identified the most crucial factors for the expected cost-effectiveness. Finally, as the analysis was based on the first clinical data, the results may be used to inform policy-makers or investors on the expected added value of navigated surgery, potentially accelerating patient access.

In summary, these three chapters show that the use of early HTA at the beginning of the clinical evaluation process could guide the further development and clinical testing process of new health technologies. Besides, its use is expected to increase the chance to develop a successful and cost-effective innovation, and it could inform pricing decisions. Although showing large uncertainties, early CEAs could inform policy-makers or investors on the potential of new technologies, which may result in faster adoption and therefore it may improve patient outcomes.

Part III - Early HTA: having multiple phase I/II studies

The use of HTA methods after the first clinical studies could identify potential implementation barriers. Additionally, HTA in this phase can be used by clinicians, pharmaceutical companies or medical device companies to evaluate whether the results are promising enough to continue further evaluation.

In this third part, two chapters applied early HTA during a clinical trial evaluating the (cost-) effectiveness of TIL-therapy for treating advanced melanoma patients. Chapter 7 showed the results of a constructive technology assessment (CTA) that identified the most important barriers and facilitators in the clinical implementation of TIL-therapy. Those barriers and facilitators were similar to general barriers and facilitators well-known for the implementation of advanced therapeutic medicinal products (ATMP). TIL-therapy is currently implemented as a treatment option in the Netherlands (in a clinical trial), but if TIL-therapy was not included

in a coverage with evidence development (CED) program, it is unlikely that patients would have access to this highly promising treatment. This chapter illustrates the complexity of the clinical implementation of ATMPs and shows the potential positive effect of using CED programs. These results are important to policy-makers and may fuel the debate on using financial risk-sharing programs for ATMPs. The analysis provides also useful information for other academic centers aiming to implement TIL-therapy, facilitating the diffusion of TIL-therapy, which is expected to improve clinical outcomes for advanced melanoma patients.
In Chapter 8, future adoption scenarios were drafted based on the findings presented in Chapter 7 and these were incorporated in a CEA to identify the most crucial contextual factors. This study showed that while the future is uncertain, TIL-therapy is expected to be cost-effective in the majority of likely scenarios. However, when TIL-therapy is commercialized or provided in a combination with ipilimumab, TIL-therapy would have a low probability to become cost-effective. As current European Union (EU) regulation complicates the provision of ATMPs to patients by academic institutes, action from governmental bodies is required to prevent commercialization. These results, together with the previously published early CEA[4], may stimulate the (preparation of) implementation of TIL-therapy in other countries as well. As the implementation of TIL-therapy is time-consuming and complex, we may conclude that this analysis could support timely patient access when the results from the phase III clinical trial become available.

Based on our experiences we evaluated that applying HTA alongside a clinical trial could identify crucial contextual factors that request anticipation upon when aiming to diffuse an innovation, which is useful information for various stakeholders (e.g. hospitals, policy-makers, governmental bodies, and patient organizations). Finally, it is expected that especially patients benefit from such analyses as it would facilitate implementation and therefore patient access to innovations enhancing clinical outcomes.

Part IV - Mainstream HTA

The use of HTA after showing the effectiveness and safety of a technology aims to inform pricing and reimbursement decisions, and clinical guideline development.

The two chapters included in this part describe the mainstream HTA analysis on the (cost-) effectiveness of robot-assisted prostatectomy (RARP) and laparoscopic prostatectomy (LRP). Chapter 9, shows that long-term urinary incontinence is improved in patients that underwent RARP compared to patients undergoing LRP. As this is one of the first long-term clinical follow-up studies comparing RARP to LRP (instead of open surgery) in a real-world setting, and it is, to our knowledge, the first study showing a clear benefit for RARP over LRP, it suggests that the experience of the surgeon played a crucial role in earlier analyses. Based on this analysis we may conclude that HTA researchers and policy-makers should be careful with drawing

conclusions based on early (HTA) analyses in surgery because of potential learning effects. Using the results from the clinical study, Chapter 10 showed that RARP seems to be cost-effective compared to LRP. Furthermore, the CEA showed the potential effect of centralization of care and a higher utilization rate of the Da Vinci robot, strengthening the first conclusion as RARP showed a higher chance to be cost-effective. For the Dutch situation, these results are a clear incentive to reimburse RARP. Furthermore, the analysis stressed that in calculating the actual costs, an agreement should be reached with insurance companies or the national healthcare institutes on the indications that may be performed with the Da Vinci robot. This may request additional CEAs opting for a similar format: evaluating the cost-effectiveness of the Da Vinci robot solely for the indication of interest, together with scenarios evaluating the increased used and therefore, reduced costs of the Da Vinci robot.

This comprehensive (mainstream) HTA analysis could inform policy-makers, clinicians, and insurance companies in reimbursement and clinical guideline decisions. These results are expected to improve the quality of care in general when RARP would be reimbursed, and improve patient outcomes by granting access to RARP. Furthermore, it showed that the evaluation of late effects could result in unexpected outcomes. This is expected to be mainly applicable in complex technologies where it is hard to oversee all potential effects.

Iterative use of HTA alongside the technical development

In addition to the benefits of using HTA per development phase, it has been suggested to be most effective when HTA is used iteratively throughout the full development process of healthcare technology[5–7]. Two of the described HTA analyses evaluated the navigation technology used in oncologic surgeries which could be seen as iterative HTA. Based on the results from the MCDA (chapter 2), development continued for the navigated surgery some specific indications (e.g. lymph node removal and colorectal cancer). Approximately 1.5 years later the first clinical results were available of using navigated surgery in colorectal cancer which were used in the early CEA (chapter 6), showing that the navigation system has the potential to be cost-effective but stresses optimal utilization of the system. For this purpose, the results of the MCDA could again be used to identify alternative indications where navigation is promising to continue clinical evaluation. Based on this example we may conclude that the separate analyses provided helpful information per development phase, but together they seem to complement each other and guide the development of medical technology more effectively towards a successful innovation.

Methodological considerations and future research

In performing the HTA analyses we encountered several challenges that should be discussed and could point at future research directions to optimize the application of HTA alongside the

translational pathway.

General reflection on HTA methods used

In the previous section, we shortly discussed the assets of the use of HTA per phase. As stated by Miquel-Cases et al. 2017 and Markiewicz et al 2014[8,9], several HTA methods are used alongside the translational pathway. However, there is currently no guideline describing which methods would be most appropriate to use in a certain development stage or for a specific type of technology. In this dissertation, specific HTA methods were chosen to evaluate a technology in a certain development stage, the chosen methods are discussed and we reflect on the appropriateness of the methods used.

In the **very early HTA phase**, we chose to perform an **MCDA** to support a research group in outlining their further research and development steps. In performing the MCDA some of the challenges described by Garcia-Hernandez (2015) were faced, e.g. the completeness of the criteria framework, and potential overlapping criteria[10]. Additionally, as the research was conducted in a very early phase of technical development no clinical data was available, due to which the performance assessment was based on expectations from end-users. As the level of experience with the technologies of interest differed per technology, this likely affected the results of the expected value. When using the Analytical Hierarchical Process (AHP) in such a very early phase, this should be taken into account.

Regarding the choice of MCDA in the early development stage, we question whether the amount of time invested (multiple interview and feedback rounds) in performing MCDA weighted up to alternative methods such as semi-structured interviews or panel discussions. One advantage of MCDA over the alternative methods is the possibility to quantify the importance of attributes and the expected value. Regarding the amount of time, in future MCDA analyses, this could be reduced by organizing group sessions instead of interviews[11], which was not deemed feasible in our set-up, and by reducing the number of attributes included in the analysis[12,13]. Although qualitative methods may also reveal expectations regarding new technologies, the quantified results of the MCDA enabled a transparent comparison of the technologies over the chosen indications.

Based on our experiences, we conclude that the relative importance evaluation is recommended to incorporate in very early HTA analyses. However, the valuation part of AHP (performance measurement) is potentially more applicable to include in a later development stage, when end-users are more experienced with the technology of interest. For this part, the applicability of semi-structured interviews or panel discussions could be similar to a structured method as AHP, especially when the information retrieved is only used within a medical device company or research group.

In the **early HTA phase**, multiple methods were used and all were evaluated as applicable and of added value. One of the used methods, **scenario drafting,** has not been applied widely in HTA research. Our research group published some examples of using scenario drafting to better inform future adoption and identify crucial potential barriers, based on the methods described by Shell international bv[14–16]. Based on the experiences of our research group, we suggest that this method is especially useful in complex and/or disruptive innovations as their diffusion is often influenced by a large range of (crucial) aspects that require attention. For example, our analysis showed that outsourcing TIL production was a likely scenario that would result in a 0% chance to become cost-effective compared to ipilimumab. At the beginning of 2020, this scenario became reality, namely, a biotech company (AdaptImmune) wants to bring TIL-therapy to the market and was reaching out to our research group for collaboration. It is likely that due to the scenario analysis, the research group was better prepared for such decisions, where even cost-effectiveness considerations were taken into account. As the future is uncertain – "Who would have foreseen a situation as the COVID-19 pandemic?" – the main challenge of applying this method in HTA is to capture the most relevant aspects that may happen in the future. Inspired by the methods described by Shell and Enserink and Hermans[14,17], we involved medical experts, besides scanning clinical trials and literature, to identify the most relevant themes to incorporate in the survey and piloted the first survey in a larger group of experts. Currently, we are one of the few research groups incorporating scenario drafting in HTA, and therefore currently no methodological guidance is available for drafting scenarios for HTA purposes. The methodology would benefit from increased application in HTA, to result in more evidence-based recommendations.
Based on our experiences, we would recommend incorporating scenario drafting in HTA performed in an early phase, before clinical implementation. This would also result in the possibility to learn from each other, ideally resulting in a general methodological guideline.

Stakeholder involvement (e.g. clinicians and patients) is important for the validity and applicability of HTA analyses[18]. When using HTA in an early or very early stage, the involvement of stakeholders is even more important as clinical outcomes are often uncertain or unknown. In multiple analyses that we presented in this dissertation, we observed that stakeholders evaluate the involvement of HTA researchers often as judgmental; hampering them for example in sharing clinical estimates of certain parameters. This could be explained by the fact that the field of early HTA is relatively young; hence most stakeholders are, if at all, only familiar with mainstream HTA. This observation calls for action by HTA researches, to collaborate more actively with clinicians and/or companies to increase the understanding and awareness of the use of early HTA. Specific attention should be given to the various aims that HTA could have when introduced during the translational pathway. It is expected that such activities improve the results from expert elicitation and the validity of the model structure, resulting in more accurate early HTA results.

Although the **iterative use of HTA** has been recommended for years[5], guidance on timing and the most appropriate methods is not yet available. In this dissertation, navigated surgery was incorporated in both the MCDA (chapter 2) and early CEA (chapter 6), which both provided useful information for the development of navigated surgery in oncology. The MCDA delivered useful information in addition to the early CEA and therefore it seems helpful to incorporate MCDA and/or stakeholder interviews in the toolbox for very early HTA analyses in iterative HTA to obtain a first insight into the expected (clinical) benefit. Besides, we advise incorporating quantitative methods such as CEA, BIA, and/or cost analysis at a somewhat later stage, to guide the further development process by identification the crucial factors for obtaining a cost-effective technology, and inform pricing decisions. In complex innovations, in this stage also more comprehensive methods are advised to use in iterative HTA to most effectively guide clinical implementation. This is further explained in the section "Duration of the translational pathway".
Based on available literature and a qualitative study by Fasterholdt on the early evaluation of innovative technologies, we may conclude that the iterative evaluation of promising technologies is not common, especially not in hospitals[19]. This is also indicated by the rapid introduction of surgical technologies based on scarce or no clinical evidence. To improve this, specific committees in hospitals may be created, focusing on prioritizing medical technologies and making implementation decisions using a transparent decision-support tool[19]. Such activities may stimulate the use of early HTA and improve stakeholder involvement.

In the **mainstream HTA** section, a large cohort study was performed as a second-best alternative for a Randomized Controlled Trial (RCT) to evaluate the efficacy of RARP compared to LRP. A **cost-utility analysis (CUA)** based on the clinical results can be seen as the best practice for applying HTA to inform reimbursement decisions.

Our experiences with several HTA methods alongside the translational pathway affirmed that the different development stages of a technology indeed bring specific challenges in performing HTA and therefore should be incorporated when choosing the most appropriate HTA method to use.

Relevant concepts faced in performing HTA

Three main methodological concepts were faced in the various HTA analyses performed in this dissertation which are described and discussed below.

HTA to inform adoption and reimbursement decisions; the definition of value
Although most HTA definitions[20,21] and HTA models[22] describe a full range of attributes (costs, effects, organization, legal, ethical, etc.) that should be incorporated in the evaluation of

technology, the majority of adoption and reimbursement decisions are based solely on cost and effectiveness data. This neglects other relevant perspectives such as specific value for patients and/or providers[23]. In especially complex innovations such decisions may result in acceptance of a technology that is eventually not used due to issues in one of the other aspects.

In chapters 4 and 10, we aimed to incorporate the full range of effects in an analysis focusing on costs and effects. Although we encountered that monetization of some aspects, for example having an easier scheduling procedure for surgery with MSL than with wire-guided localization (WGL), and easier application of MSL outside the breast cancer indication compared to Radio-active Seed Localization (RSL) is challenging. Such benefits are valuable for planners and hospitals, but due to a lack of data and difficulties in monetization, we were unable to incorporate these benefits in our analysis and therefore in our conclusion aimed at decision-makers. A similar issue was faced in chapter 10 with incorporating the potential ergonomic advantages of RARP compared to LRP. To incorporate these effects in a health economic model, we needed to relate the ergonomic complaints of a surgeon to patient effects. This can be seen as a pragmatic solution to incorporate effects on different levels, although concepts of surgeon satisfaction, increasing waiting times in case of sick leave, and the health effects of the surgeon were still not incorporated.

To enable more comprehensive healthcare decision making, multiple research groups aimed to reach consensus about a more complete definition of value[24,25] and aimed to identify alternative methods to incorporate all relevant effects enabling comprehensive decision making[23,26], e.g. multiple criteria analyses. Embracing the multiple criteria approaches would increase the transparency of decision making and potentially facilitating consensus on a decision problem among stakeholders[23]. We think such additional effects are important to incorporate and we therefore would suggest incorporating such approaches in “mainstream” HTA analyses, to evaluate its applicability in formal decision-making.

Variability in cost data used in HTA

In this dissertation, several issues were encountered related to the evaluation of costs.
One challenge was to obtain accurate cost information of complex innovations or usual care. For the Dutch situation, this is mostly explained by the current payment system, due to negotiated tariffs for a Diagnosis-Related Group (DRG). As in (complex) innovations often a small change is expected in the DRG, the exact expected additional cost of the innovation is hard to estimate. The current Dutch manual describes that costs could either be evaluated top-down or bottom-up, but no recommendation is provided[27]. Also internationally, no consensus has been reached on methods to estimate the costs[28,29], although it has been described that costs from DRGs and costs based on reimbursement received by hospitals

differ substantially from the actual costs[30]. Bottom-up costing analyses have been proposed for their transparency and accurateness[31,32], though these analyses are time-consuming and request intensive collaborations with the financial department of hospitals and clinicians. Additionally, the method would benefit from further guidance on estimating overhead and personnel costs as it is likely to underestimate those elements when following the current guideline[27], as discussed in Chapter 5. Due to the lack of consensus on methods to estimate costs for HTA purposes, it is difficult to compare our results with existing data, limiting the generalizability. We, therefore, advocate that national, and preferably international, guidelines should be developed on the most appropriate methodology for cost analyses (e.g. bottom-up or top-down cost analysis).

Another issue faced was the evaluation of medical device costs, as list prices may differ – similar to pharmaceuticals – per hospital and country[33,34] from the actual (negotiated) prices. Most cost-effectiveness analyses include however list prices which may result in a rather negative outcome, potentially resulting in withholding patients from promising and potentially cost-effective technologies.
For example in chapter 10 we found that the costs for RARP (bottom-up) ranged between €9,670 and €10,250 (3 hospitals), which showed a large influence on the ICER (~€29,000 to ~€39,000). As most of the comparable studies evaluated the costs based on list prices or even neglected the costs of the robot, as they interpreted the purchase of the robot as an investment[35], we were limited to compare our findings with literature. Based on the discrepancies we observed, we recommend that the actual prices at least should be addressed in a sensitivity analysis to present a fair analysis of the cost-effectiveness of the novel technology. We stress that international guidance on this aspect is needed to increase the generalizability of results from HTA analyses (inter)nationally.

The final challenge that we faced relates to the assessment of the per-patient costs of (surgical) medical technology as it often requires a large investment (e.g. Da Vinci robot)[36]. Hence the costs should be linked to the utilization rate and the expected lifetime of the technology, similar to the price-setting of pharmaceuticals which also incorporates R&D costs. However, as currently no guideline is available, one may also propose that the investment is on account of a hospital and should not be incorporated in the per-patient costs, requiring hospitals to evaluate the return on investment.
When the investment costs are incorporated in the per-patient costs, the utilization rate challenges the evaluation of the (expected) cost-effectiveness, as the utilization rate could increase over the years (e.g. due to usage in multiple indications). In chapters 6 and 10 we identified that the utilization rate of a medical device could have a significant influence on the cost-effectiveness results. Therefore, we recommend incorporating the utilization rate parameter in a sensitivity analysis, to evaluate the potential effect of increased utilization.

Furthermore, to incorporate multiple utilization rates, we strongly advise using bottom-up cost analyses.

Based on our experiences regarding the evaluation of costs, we thus conclude to preferably use bottom-up costing approaches to accurately evaluate the costs of standard of care and the proposed innovative technologies. The application of this method would also facilitate to evaluate the impact of utilization rates of medical devices on the (expected) cost-effectiveness. We advise to incorporate the utilization rate at least in a sensitivity analysis to inform decision-makers about the effect of for example centralization of care. Finally, as actual prices may differ from list prices we recommend to incorporate the actual (negotiated) prices in a sensitivity analysis.

Duration of the translational pathway

As described by Rogers and Cain & Mittman, the speed of diffusion is depending on roughly twelve attributes, of which the relative advantage of a novel technology is only one[37,38], the other factors are mainly social and contextual. In this dissertation, we observed that the adoption of surgical devices tends to be quite quick, potentially due to key opinion leaders and its compatibility, while it does not always show a convincing relative advantage. On the other hand, a long translational pathway in TIL-therapy or ATMPs was observed, while TIL-therapy and/or ATMPs are expected to substantially improve patient outcomes. This is likely to be explained by the complexity of the treatment, the rapidly evolving field, and weaker market power and position of academic institutions compared to the pharmaceutical industry.

As HTA research aims to contribute to the reduction of healthcare costs and the improvement of health outcomes, it is relevant to identify differences in the (expected) duration of the translational pathway of innovations to guide and support life-saving technologies towards the market and being critical on technologies that come to the market because for example, technology push. In technologies showing a long translational pathway, based on our experiences in chapters 7 and 8, we would recommend to use comprehensive methods such as CTA and scenario drafting, as those could identify crucial contextual factors, on which actions can be taken upon. For technologies that are expected to be diffused quickly, we would recommend using quantitative methods (e.g. cost analyses, cost-effectiveness analyses, and value of information analyses), to evaluate whether they bring the required benefits for the expected additional costs. In choosing the most appropriate HTA method in a certain phase of development, it may thus be beneficial to incorporate the (expected) duration of the translational pathway.

One of the main barriers observed in chapter 7 for the implementation of TIL-therapy was inadequate financial support, which was resolved in our setting by obtaining a coverage with

evidence development program (CED). In translating personalized technologies it seems that generic regulatory assessments may be unsuitable[39,40]. Based on the TIL-therapy case, we suggest that innovative assessment procedures or risk-sharing programs such as a CED should be further developed and applied by policy-makers internationally, as such technologies are often less costly and highly promising for patients. Besides the suggested support, as has been described in the main results section, we think a role exists for policy advisors on a regulatory level as currently it is nearly impossible for academic-based therapies to remain in the academic setting after proving its effectiveness and safety[41]. However, transferring a technology to a biotech company will likely result in higher prices that would increase the pressure on healthcare budgets, or even may result in withholding patients from this promising technology. Therefore, in our view, supporting ATMPs to remain in an academic setting is also of interest to the national health authorities.

Recommendations for policy

Based on the findings presented in this dissertation and the methodological considerations discussed, we present some recommendations that might improve effective innovation and timely patient access to promising innovations, which is expected to result in improved health outcomes.

The most recent definition of HTA describes that it is a systematic and transparent evaluation of technologies at different points in its lifecycle[20]. Based on our experiences in very early and early HTA (e.g. MCDA, early BIA, early CEA), we think it is indeed beneficial to use HTA also at the beginning of the development process. We would especially recommend that researchers and developers use (very) early HTA analyses more often to facilitate their development process and guide them towards a successful product or strategy. HTA may for example be introduced as a "reality check", where an HTA researcher evaluates independently whether or not it is expected to reach the required added value for the additional costs. In case it is not (yet) expected to be cost-effective, the analysis can point at directions for further development that may result in a successful innovation. Introduction of (very) early HTA is then expected to reduce the failure rate of innovations after a full development process. Furthermore, HTA researchers are recommended to seek collaborations with clinicians, surgeons, and medical device companies to stimulate the use of (very) early HTA.

To further stimulate the use of (very) early HTA by researchers and/or developers, we think it is wise to request early HTA data (e.g. early value assessment, evaluate the expected budget impact, and potentially perform a risk assessment evaluating potential adoption barriers) in application rounds for research grants. This will likely stimulate effective innovation. Furthermore, the availability of early HTA data is expected to support the process of choosing

the best research to fund, similarly to for example the NICE (national institute for health and clinical excellence) initiative where several prioritization criteria are used focusing on e.g. having an impact on decision making, having an impact on patient care, and the level of uncertainty surrounding the topic, in deciding which research to fund[42]. Also, the use of VOI analyses that evaluate the level of uncertainty could be informative in choosing the best research to fund as described by the ISPOR (International Society for Pharmacoeconomics and Outcomes Research) taskforce[43].

Besides the use of (very) early HTA to prioritize research funding, we would also recommend engaging early HTA results more often in adoption decisions of promising innovations, instead of waiting for the best evidence (RCT results), especially when the innovation targets a high unmet clinical need.

As suggested by Bindels and colleagues, VOI analyses could be informative in adoption decisions. However, they identified some crucial barriers for its implementation, e.g. that VOI analyses do often not cover all uncertainties and not all research designs indicated by VOI analyses may be feasible[44]. Such hurdles seem hard to overcome and therefore it has been suggested that relevant uncertainties not incorporated in the VOI analysis should be clearly communicated to the decision maker[45]. To our best knowledge, the use of VOI data to inform reimbursement decisions has not been piloted in the Netherlands yet. Based on our experiences with the TIL-therapy case we would suggest starting a pilot in which reimbursement is granted to technologies that shown to be (1) safe and (2) effective in a phase II study, and, based on the expected cost-effectiveness (3) show a low level of uncertainty in the VOI analysis. This would enable us to evaluate the benefit/risk ratio of granting access to promising innovations on early data instead of waiting – while potentially life-years or improved quality of life is missed – for higher-level evidence.
Such a pilot would also provide valuable input for HTA researchers on the applicability of VOI analyses for decision making, and guide potential improvements of the method(s) used and/or presentation of such analyses[46].

To HTA researchers we would recommend based on our results in chapters 4, 5, 6, and 10, to use bottom-up cost analyses to evaluate the costs of innovative medical devices and its comparators. As previously discussed, these analyses are often time-consuming and request collaboration with the financial department of a hospital and clinicians. To facilitate the use of bottom-up costing analyses, initiatives should be started to register actual costs per activity per patient within a hospital and preferably provide these costs in national databases (big data). These initiatives are also likely to ease the implementation of value-based healthcare (VBHC) programs aiming to improve patient outcomes at lower costs[47,48].

Furthermore, our results suggest that in choosing the most appropriate HTA method alongside the translational pathway, the development stage of the technology (e.g. quantitative vs qualitative), the expected duration of the translational pathway, the expected clinical benefit (e.g. comprehensive HTA/implementation research vs cost-benefit analysis), and the question at hand should be taken into account. Further research may be focused on the evaluation of potential other aspects that influence this choice. Subsequently, all aspects may be incorporated in a decision matrix aiming to support choosing of the best HTA method per development phase.

Finally, in this era where medical devices and new treatment strategies evolve quickly, HTA researchers are recommended to take the potential adoption of new technologies into account by for example making the analyses (CEA/BIA) adaptive.

Concluding remarks

The aim of this dissertation was to inform multiple stakeholders and to contribute to the knowledge on the application of early HTA by providing real-world examples of using HTA alongside the translational pathway of innovations in oncology. Based on our findings we conclude that the use of HTA alongside the development of a technology could be valuable to guide stakeholders in making decisions on development, on the continuation of R&D, and on reimbursement of innovative technologies. The use of iterative HTA seems especially beneficial because it continuously informs the development process, which increases the likelihood of developing an innovation that will be accepted by end-users and is expected to increase the chances to receive reimbursement.

Based on our experiences, we suggest that HTA researchers should involve clinical researchers and other relevant stakeholders in performing HTA, and should perform HTA as early as possible to decrease the risk of falling in one of the "Valleys of death". We observed that this can be challenging, therefore we suggest that funders or investors request (very) early HTA data in research proposals which could increase the chance of funding the innovations that are most likely to obtain market access in the end. Additionally, the introduction of the new law regarding medical technologies is likely to support and increase the use of early HTA and involve the relevant stakeholders.

By performing HTA in various phases of technical development we evaluated that per phase a different stakeholder is served. For example, very early HTA is especially informative for researchers and developers, where mainstream HTA is most valuable to policy-makers. Regarding the decision on choosing the most appropriate HTA method per development phase, based on our results we suggest that this depends on the development phase

(availability of data), the expected duration of the translational pathway, and the expected benefit of the technology.

This dissertation identified that to introduce HTA alongside the translational pathway in complex innovations and/or medical devices effectively, (inter)national guidance is needed on the most appropriate cost methodology to accurately, transparently, and consistently evaluate and report the costs of innovative interventions. This would improve the generalizability of health economic evaluations. Furthermore, as next to costs and effects other aspects (e.g. legal and/or ethical issues) could be crucial for the adoption of a technology, new methods to incorporate additional attributes in evaluations that could be used in formal decision-making should be further developed and introduced. Such innovative methods are likely to decrease the number of promising technologies that will not be implemented in the clinic.

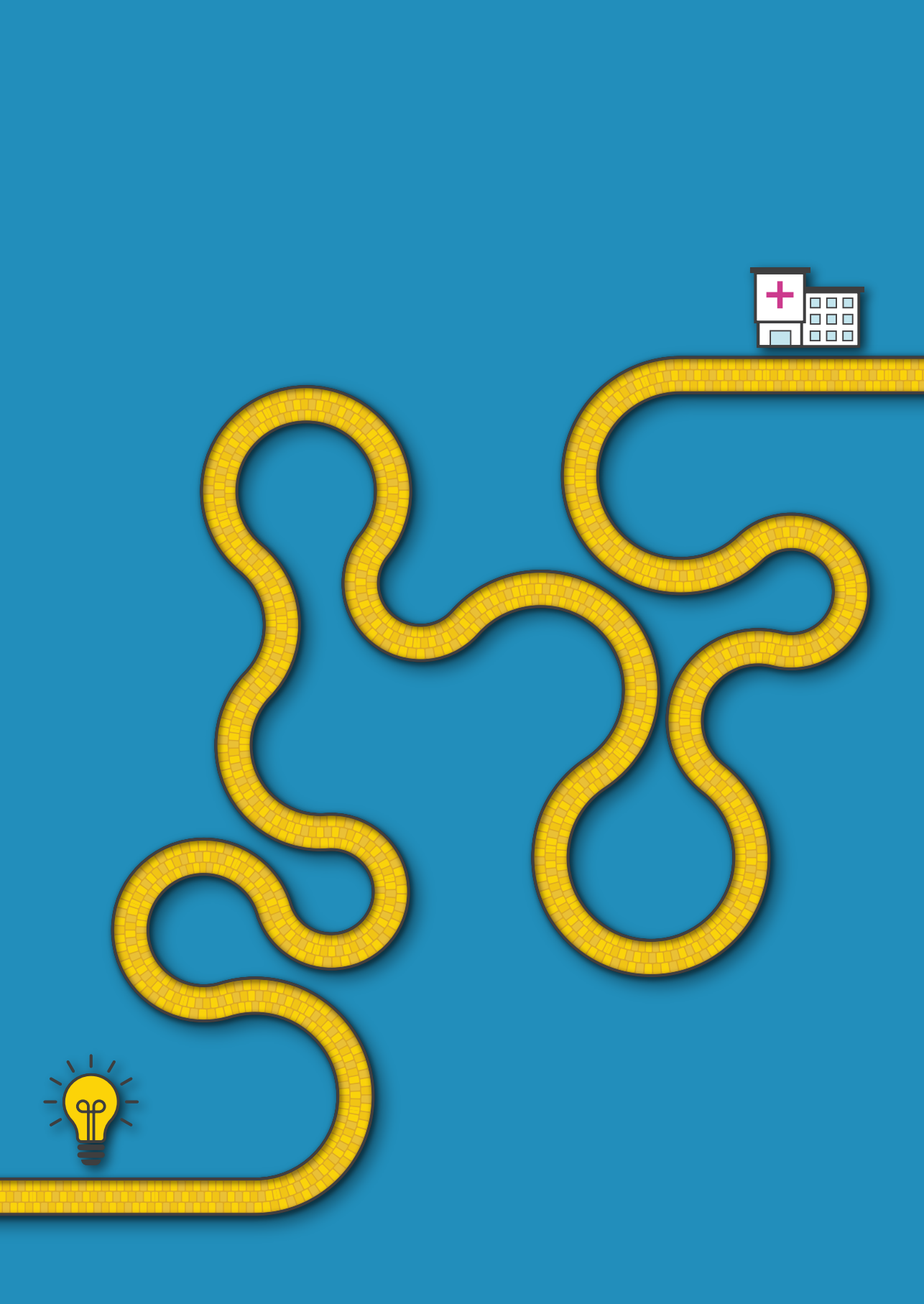
Based on our experiences with the TIL-therapy case, a final remark is that policy-makers are strongly advised to evaluate alternative regulatory assessment programs or expand the use of CED programs to improve the adoption of very promising technologies for patient populations with a high unmet clinical need. It could be beneficial to introduce methods such as VOI in decision-making processes and/or start a pilot on the early introduction of highly promising technologies based on early HTA results. Furthermore, we think the allocation process of research funding could also benefit from the application of HTA methods (e.g. MCDA) to choose the best technologies.

We believe that the multiple HTA analyses included in this dissertation provided valuable information to various stakeholders (clinicians, researchers, developers, staff members, and policy-makers) to guide the next steps of the translational pathway. The conclusions and recommendations from this dissertation are expected to improve the applicability of HTA alongside the translational pathway and stimulate stakeholders in applying HTA in various development phases. This is expected to result in effective healthcare innovation and accelerating patient access to promising technologies.

REFERENCES

1. Marsh, K., Lanitis, T., Neasham, D., Orfanos, P. & Caro, J. Assessing the value of healthcare interventions using multi-criteria decision analysis: A review of the literature. *PharmacoEconomics* vol. 32 345–365 (2014).
2. 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 Heal.* **19**, 1–13 (2016).
3. Abrishami, P., Boer, A. & Horstman, K. Understanding the adoption dynamics of medical innovations: Affordances of the da Vinci robot in the Netherlands. *Soc. Sci. Med.* **117**, 125–133 (2014).
4. Retèl, V. P. *et al.* Early cost-effectiveness of tumor infiltrating lymphocytes (TIL) for second line treatment in advanced melanoma: a model-based economic evaluation. *BMC Cancer* **18**, 895 (2018).
5. Sculpher, M., Drummond, M. & Buxton, M. The iterative use of economic evaluation as part of the process of health technology assessment. *J. Health Serv. Res. Policy* **2**, 26–30 (1997).
6. 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* **24**, 459–64 (2008).
7. IJzerman, M. J., Koffijberg, H., Fenwick, E. & Krahn, M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *Pharmacoeconomics* **35**, 727–740 (2017).
8. Miquel-Cases, A. *et al.* (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. *Cancer Treat. Rev.* **52**, 117–127 (2017).
9. Markiewicz, K., van Til, J. A. & IJzerman, M. J. MEDICAL DEVICES EARLY ASSESSMENT METHODS: SYSTEMATIC LITERATURE REVIEW. *Int. J. Technol. Assess. Health Care* **30**, 137–146 (2014).
10. Garcia-Hernandez, A. A Note on the Validity and Reliability of Multi-Criteria Decision Analysis for the Benefit–Risk Assessment of Medicines. *Drug Saf.* **38**, 1049–1057 (2015).
11. Hummel, J. M., Van Rossum, W., Verkerke, G. J. & Rakhorst, G. The effects of team expert choice on group decision-making in collaborative new product development: a pilot study. *J. Multi-Criteria Decis. Anal.* **9**, 90–98 (2000).
12. 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 Heal.* **19**, 125–137 (2016).
13. Hummel, J. M., Steuten, L. G. M., Groothuis-Oudshoorn, C. J. M., Mulder, N. & IJzerman, M. J. Preferences for Colorectal Cancer Screening Techniques and Intention to Attend: a Multi-Criteria Decision Analysis. *Appl. Health Econ. Health Policy* **11**, 499–507 (2013).
14. Shell international BV. *Scenarios: An Explorer's Guide*. https://www.shell.com/energy-and-innovation/the-energy-future/scenarios/new-lenses-on-the-future/earlier-scenarios.html (2008).
15. Joosten, S. E. P., Retèl, V. P., Coupé, V. M. H., van den Heuvel, M. M. & van Harten, W. H. Scenario drafting for early technology assessment of next generation sequencing in clinical oncology. *BMC Cancer* **16**, 66 (2016).
16. Retèl, V. P., Joore, M. A., Linn, S. C., Rutgers, E. J. T. & van Harten, W. H. Scenario drafting to anticipate future developments in technology assessment. *BMC Res. Notes* **5**, 442 (2012).
17. Enserink, B. & Hermans, L. *Policy Analysis of Multi-Actor Systems chapter 5: Exploring the Future*. (Eleven International Publishing, 2010).
18. Nielsen, C. P. *et al.* Involving stakeholders and developing a policy for stakeholder involvement in the European network for Health Technology Assessment, EUnetHTA. *Int. J. Technol. Assess. Health Care* **25**, 84–91 (2009).
19. Fasterholdt, I., Lee, A., Kidholm, K., Yderstræde, K. B. & Pedersen, K. M. A qualitative exploration of early assessment of innovative medical technologies. *BMC Health Serv. Res.* **18**, 837 (2018).
20. O'Rourke, B., Oortwijn, W. & Schuller, T. The new definition of health technology assessment: A milestone in international collaboration. *Int. J. Technol. Assess. Health Care* **36**, 187–190 (2020).
21. World Health Organization. Health Technology Assessment. http://www.who.int/medical_devices/assessment/en/.
22. EUnetHTA. HTA Core Model Handbook,. *EUnetHTA Jt. Action 2, Work Packag. 8. HTA Core Model ® version 3.0 (Pdf); Available from www.htacoremodel.info/BrowseModel.aspx* **2**, 408 (2016).
23. Phelps, C. E. & Madhavan, G. Using Multicriteria Approaches to Assess the Value of Health Care. *Value Heal.* **20**, 251–255 (2017).

24. Lakdawalla, D. N. *et al.* Defining Elements of Value in Health Care—A Health Economics Approach: An ISPOR Special Task Force Report [3]. *Value Heal.* **21**, 131–139 (2018).
25. Garrison, L. P., Pauly, M. V., Willke, R. J. & Neumann, P. J. An Overview of Value, Perspective, and Decision Context—A Health Economics Approach: An ISPOR Special Task Force Report [2]. *Value Heal.* **21**, 124–130 (2018).
26. Angelis, A. & Kanavos, P. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework. *Soc. Sci. Med.* **188**, 137–156 (2017).
27. Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T. & Swan Tan, S. *Manual for cost research: methods and standard cost prices for economic evaluations in health care*. (2015).
28. Adam, T., Evans, D. B. & Koopmanschap, M. A. COST-EFFECTIVENESS ANALYSIS: CAN WE REDUCE VARIABILITY IN COSTING METHODS? *Int. J. Technol. Assess. Health Care* **19**, 407–420 (2003).
29. Franklin, M., Lomas, J., Walker, S. & Young, T. An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis. *Pharmacoeconomics* **37**, 631–643 (2019).
30. Skeie, B., Mishra, V., Vaaler, S. & Amlie, E. A comparison of actual cost, DRG-based cost, and hospital reimbursement for liver transplant patients. *Transpl. Int.* **15**, 439–445 (2002).
31. Kaplan, R. S. & Anderson, S. R. Time-driven activity-based costing. *Harv. Bus. Rev.* **82**, 131-138,150 (2004).
32. Lievens, Y., van den Bogaert, W. & Kesteloot, K. Activity-based costing: a practical model for cost calculation in radiotherapy. *Int. J. Radiat. Oncol.* **57**, 522–535 (2003).
33. van Harten, W. H., Wind, A., de Paoli, P., Saghatchian, M. & Oberst, S. Actual costs of cancer drugs in 15 European countries. *Lancet Oncol.* **17**, 18–20 (2016).
34. Pauly, M. V. & Burns, L. R. Price Transparency For Medical Devices. *Health Aff.* **27**, 1544–1553 (2008).
35. Schroeck, F. R. *et al.* Cost of New Technologies in Prostate Cancer Treatment: Systematic Review of Costs and Cost Effectiveness of Robotic-assisted Laparoscopic Prostatectomy, Intensity-modulated Radiotherapy, and Proton Beam Therapy. *Eur. Urol.* **72**, 712–735 (2017).
36. Ventola, C. L. Challenges in evaluating and standardizing medical devices in health care facilities. *P T* **33**, 348–59 (2008).
37. Cain, M. & Mittman, R. Diffusion of Innovation in Health Care. *Ihealthreports* (2002).
38. Rogers, E. M. The Diffusion of Innovations. *Diffus. Innov.* 576 (2003).
39. Corbett, M. S., Webster, A., Hawkins, R. & Woolacott, N. Innovative regenerative medicines in the EU: a better future in evidence? *BMC Med.* **15**, 49 (2017).
40. Faulkner, E. *et al.* Challenges in the development and reimbursement of personalized medicine-payer and manufacturer perspectives and implications for health economics and outcomes research: a report of the ISPOR personalized medicine special interest group. *Value Heal. J. Int. Soc. Pharmacoeconomics Outcomes Res.* **15**, 1162–1171 (2012).
41. Pearce, K. F. *et al.* Regulation of advanced therapy medicinal products in Europe and the role of academia. *Cytotherapy* **16**, 289–297 (2014).
42. Sharma, T., Choudhury, M., Rejón-Parrilla, J. C., Jonsson, P. & Garner, S. Using HTA and guideline development as a tool for research priority setting the NICE way: reducing research waste by identifying the right research to fund. *BMJ Open* **8**, e019777 (2018).
43. Fenwick, E. *et al.* Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Heal.* **23**, 139–150 (2020).
44. Bindels, J. *et al.* Use of Value of Information in Healthcare Decision Making: Exploring Multiple Perspectives. *Pharmacoeconomics* **34**, 315–322 (2016).
45. Pouwels, X. G. L. V., Grutters, J. P. C., Bindels, J., Ramaekers, B. L. T. & Joore, M. A. Uncertainty and Coverage With Evidence Development: Does Practice Meet Theory? *Value Heal.* **22**, 799–807 (2019).
46. Grimm, S. E. *et al.* Development and Validation of the TRansparent Uncertainty ASsessmenT (TRUST) Tool for Assessing Uncertainties in Health Economic Decision Models. *Pharmacoeconomics* **38**, 205–216 (2020).
47. Porter, M. E. What is value in health care? *N. Engl. J. Med.* **363**, 2477–2481 (2010).
48. Etges, A. P. B. da S., Ruschel, K. B., Polanczyk, C. A. & Urman, R. D. Advances in Value-Based Healthcare by the Application of Time-Driven Activity-Based Costing for Inpatient Management: A Systematic Review. *Value Health* **23**, 812–823 (2020).



Annex

A

SUMMARY

This dissertation aims to contribute to the knowledge on the application of health technology assessment (HTA) methodologies alongside the translational pathway of medical innovations in oncology. By doing so, this dissertation aims to serve as a start to position early HTA in the comprehensive evaluation of medical technologies.

Healthcare budgets are under pressure because of increasing healthcare costs. This results in an increased awareness that money has to be spent wisely and therefore health innovations need to prove their value for money. Which is often evaluated by a cost-effectiveness analysis. These analyses are mostly performed after the full development process, just before market entry. At this stage, it could be that a technology is not deemed cost-effective and would not be adopted in a healthcare system. Such late failures could be prevented by earlier introduction of comprehensive HTA, and at that moment it is still possible to act upon undesired developments. This dissertation is divided into four parts according to the moment of using HTA in the development process of a new technology (e.g. very early HTA). Each chapter aims to inform the relevant stakeholder(s) on topics such as further research & development (R&D), implementation, and/or reimbursement decisions of the technology of interest.

Chapter 1 provides background information on the translational pathway of innovations and the barriers that may be encountered on the path towards implementation. Furthermore, the use and application of HTA is explained, and the various HTA methods that could be used alongside the translational pathway are presented.

Part I - Very early HTA

In **Chapter 2**, three image-guided surgical tools that are developed to be used in oncology were prioritized by means of a multiple criteria decision analysis (MCDA). The analysis aimed to guide further R&D activities. In this MCDA, the analytic hierarchical process (AHP) was used to evaluate the expected added value of navigation, optical imaging, and augmented reality in five oncologic indications compared to usual care. Sixteen decision criteria were selected after which their relative importance was evaluated. Subsequently, the expected performance of the technologies on the decision criteria were evaluated among surgeons. The combination of these scores provided the expected value per technology. Differences were seen on the importance of specific criteria. For example, improvement of patient satisfaction was thought more important in the removal of breast cancer tumors compared to the other indications. Combining the relative importance with the expected performance

showed that; navigation use is expected to be most valuable in the removal of lymph nodes, liver tumors, and rectal tumors; the use of optical imaging is expected to be most valuable in the removal of tongue and rectal tumors; and the addition of augmented reality was judged most promising in tongue tumor resections. None of the technologies showed a convincing benefit for breast cancer. In selecting and prioritizing these technologies based on these results, especially optical and navigation technologies are expected to add value in addition to usual care. Further development and evaluation of those technologies for the preferred indications seems therefore worthwhile. Based on this analysis we may conclude that performing multi-attribute analysis is useful in prioritization of clinical studies and to steer R&D initiatives.

In **chapter 3**, a systematic literature search was conducted focused on the performance of imaging technologies to detect early response on neoadjuvant chemotherapy (NACT) in patients treated for breast cancer. As previous studies showed that the performance varies per breast cancer subtype, we stratified our analysis to breast cancer subtypes (ER status and HER2 status). In total 229 articles were identified of which 30 were selected for full reading and finally 15 observational studies were included. In ER-positive/HER2-negative breast cancer patients, ^{18}F-FDG-PET/CT showed the best results (highest observed sensitivity of 89%). In triple-negative breast cancer patients, the best results were found for ^{18}F-FDG-PET/CT after 2 NAC cycles (highest observed sensitivity of 79%), but the performance of MRI seemed also promising. Among HER2-positive patients, ^{18}F-FDG-PET/CT showed also the best performance, and MRI seemed promising. Further investigation towards the use of MRI in HER2-positive and triple-negative patients seemed therefore relevant. These results should be interpreted with caution, as the included studies were underpowered, had heterogeneous study designs and various outcome measures. Therefore, this chapter concludes that the present evidence is insufficient to recommend on using specific imaging technologies per breast cancer subtype.

Part II - Early HTA: up to and including the first clinical studies (phase I)

In the Netherlands Cancer Institute (NKI) a novel technology was developed: magnetic seed localization (MSL) to resect non-palpable breast cancer tumors effectively. The effectiveness of this technology is expected to be similar to localization techniques that are currently available: wire-guided localization (WGL) and radioactive seed localization (RSL), but it is expected to overcome several disadvantages of those techniques. Therefore, in **chapter 4**, the budget impact of gradually adopting MSL in the Netherlands next to RSL and WGL is evaluated. The intervention and implementation costs of the technologies were evaluated bottom-up in multiple Dutch hospitals. As the price of the magnetic seed was still to be determined, this was included as a range. The intervention costs of MSL were expected to

be slightly higher to RSL and WGL. The implementation costs of MSL were expected to be substantially lower compared to RSL. When the magnetic seed would be priced at €100 per seed, the budget impact was in favor of adopting MSL. When the magnetic seed was above €200 the budget impact was in favor of standard of care. Based on this early analysis we conclude that MSL could be a cost-efficient technique in resecting non-palpable tumors in the Netherlands.

Over the past decade, many hospitals have adopted hybrid operating rooms (OR). As resources are limited, these ORs have to prove themselves in adding value. Current estimations on standard OR costs show great variety while cost analyses of hybrid ORs are lacking. Therefore, in **chapter 5** the costs and cost drivers of a conventional and hybrid OR were evaluated by a comprehensive bottom-up cost analysis in five Dutch hospitals that have one or more hybrid ORs. The analysis can serve as a first step in evaluating the added value of the hybrid OR. The cost categories that were evaluated were: construction, inventory, personnel, and overhead. The per-minute costs of the conventional and hybrid OR were €9.45 (€8.60 - €10.23) and €19.88 (€16.10 - €23.07), respectively. For the conventional OR, the personnel costs was the biggest driver. For the hybrid OR, the utilization rate of the OR had the largest impact, followed by the total inventory costs. The analysis showed that it might be challenging for the hybrid OR to become cost-effective due to its higher costs compared to a conventional OR.

Based on chapter 2, especially the addition of the navigation system seemed promising in multiple indications. The research group developing the navigation system at the NKI continued their research and showed, based on the first series of Locally Recurrent Rectal Cancer (LRRC) and Locally Advanced Rectal Cancer (LARC) patients, that navigated surgery improved surgical margin rates compared to standard surgery in a historical control group[1]. Based on these results we evaluated, the potential of using the navigation system in LARC and LRRC to become cost-effective in **Chapter 6**. An early cost-effectiveness analysis (CEA) was performed based on the clinical data of Kok et al.[1], and survey data on quality of life that was sampled in a prospective cohort study. To estimate the expected costs and outcomes a Markov decision model was constructed. We found that navigated surgery showed incremental costs and QALYs in both indications. The scenario analysis showed that optimal utilization of the navigation technology results in incremental cost-effectiveness ratios (ICER) below the accepted willingness to pay thresholds in the Netherlands. Based on this early evaluation we conclude that navigated surgery is expected to be cost-effective in LRRC and has the potential to become cost-effective in LARC patients. To decrease the level of uncertainty in this analysis, it is recommended to continue with prospective clinical evaluation of navigated and standard surgery in these patients. Based on this data the cost-effectiveness analysis should be updated which could inform decisions on reimbursement of navigated surgery in LRRC and LARC.

Part III - Early HTA: both phase I/II studies

Clinical implementation of advanced therapy medicinal products (ATMP) is known to be challenging. In **chapter 7**, we evaluate the early clinical implementation of a promising ATMP: TIL-therapy in advanced melanoma, by means of a constructive technology assessment (CTA). Since 2011 TIL-therapy has been introduced as an experimental therapy in the NKI and is included in a coverage with evidence development (CED) program since 2014. In this chapter, first literature was summarized on barriers and facilitators known for the clinical implementation of ATMPs. Subsequently, semi-structured interviews were held with 26 stakeholders covering six CTA domains: clinical, economic, patient-related, organizational, technical, and future. Additionally, for the economic domain, a bottom-up costing analysis was performed to identify the treatment costs of TIL-therapy. The barriers identified for TIL-therapy corresponded to the barriers found for ATMPs in literature. The analysis identified the following facilitators for implementation of TIL-therapy: financial support from the CED program, availability of an in-house pharmacist, quality assurance expertise, and a TIL-skilled technician. This chapter concludes that institutional and national implementation of TIL-therapy remains complex, but is feasible. The results of the ongoing RCT comparing TIL-therapy with its current standard ipilimumab are expected to facilitate the adoption of TIL-therapy. Finally, as financial support was one of the main barriers reported in the literature for clinical implementation of ATMPs, we recommend that national healthcare institutes evaluate possibilities such as innovative and conditional reimbursement procedures to support similar promising therapies.

Based on the identified barriers and described future perspectives in chapter 7, future adoption scenarios for TIL-therapy in advanced melanoma were drafted in **chapter 8**. These scenarios were evaluated among international experts and finally incorporated in an existing cost-effectiveness model (Retèl et al[2]). The scenarios incorporated multiple contextual factors such as competitive therapies coming to the market, research developments surrounding TIL-therapy, the influence of pharmaceutical companies, and the attitude of patients and clinicians. This chapter aimed to inform adoption decisions of TIL-therapy for which in total 14 scenarios were drafted. The likelihood that these scenarios would happen within the coming 5 years was surveyed among 29 experts from 12 countries. The average likelihood of the scenarios ranged between 29% and 58% indicating that future developments of TIL-therapy were uncertain. Using the estimated likelihood, recent literature, and expert elicitation, eight scenarios were labeled as likely. Six of those were incorporated in the cost-effectiveness analysis. In most of the scenarios, TIL-therapy was still considered cost-effective compared to ipilimumab. These scenarios may therefore function as facilitators. TIL-therapy was however not likely to be cost-effective when the production of TILs would be outsourced or TIL-therapy would be provided in a combination with ipilimumab. These scenarios should

be considered in the adoption decision as these may act as crucial barriers.

Part IV - Mainstream HTA

Despite multiple studies on the efficacy of robot-assisted radical prostatectomy (RARP), there is no definitive conclusion about the added value of RARP. In **Chapter 9**, we present the results from a national retrospective cluster study evaluating long-term sexual and urinary functioning (6-9 years) after RARP and laparoscopic radical prostatectomy (LRP). This evaluation was based on real-world data from 12 Dutch hospitals. Data (clinical characteristics, patient-reported outcomes, and peri- and post-operative outcomes) was collected from patients who underwent surgery between 2010 and 2012 in one of the selected hospitals. Primary outcomes were urinary and sexual functioning measured by the EPIC-26. In total 1370 patients participated of whom 907 underwent RARP and 463 LRP. After a median follow-up time of 7 years, we found a clinically relevant and a significantly better score for urinary functioning after RARP compared to LRP. We found no significant difference in sexual functioning. The regression analysis showed that better functional outcomes were associated with having a nerve-sparing procedure, not receiving adjuvant radiotherapy, and being younger at the surgery. Furthermore, this study showed a preference towards RARP in perioperative outcomes such as procedure times, the chance of receiving a nerve-sparing procedure, and blood loss. This chapter concludes that RARP showed better urinary function compared to LRP and is preferred on peri-operative outcomes. Therefore, guidelines regarding radical prostatectomy may change and decision-makers are advised to reconsider their position on coverage, especially when RARP proves to be cost-effective.

Based on the results described in chapter 9, **chapter 10** presents a cost-effectiveness analysis evaluating RARP compared to LRP from a Dutch perspective. As oncologic outcomes between the groups were similar, the analysis focused on functional outcomes and the use of additional care for complaints of urinary incontinence and erectile dysfunction. The intervention costs were based on a bottom-up cost analysis in five hospitals (2 for LRP, 3 for RARP). The costs for additional care were based on patient-reported information (e.g. type of care received and frequency of care) and national reference prices. A decision-tree was built to evaluate the costs and effects in QALYs over a time horizon of approximately 7 years. RARP showed higher intervention costs (€9,964) compared to LRP (€7,253), but the follow-up costs were higher for LRP. The analysis showed that RARP was more costly but also more effective within the informal willingness to pay range of €80,000, showing an ICER of €34,206. As a best-case scenario, when RARP is being centralized (>150 cases/year), total trajectory costs decreased due to higher throughput, shorter procedure time, and shorter length of stay. This showed an ICUR of €3,495 per QALY gained. RARP showed to be cost-effective compared to LRP based on data from the population-based, large-scale study. This is a clear incentive to fully reimburse

RARP, especially when hospitals provide RARP centralized.

In **chapter 11**, the main findings and the implications of this dissertation are discussed. The findings from the studies performed in a very early stage (chapters 2 and 3) are informative to researchers and developers to decide on their future research steps and could contribute to more effective innovation and acceleration of patient access. The results from the HTA studies performed up to and including the first clinical studies (chapters 4, 5, 6) are useful to guide the further development and clinical testing process of new technologies, inform the optimal use of innovations, and to guide pricing decisions. This part also suggests that early CEAs – although showing large uncertainties – could inform policy-makers or investors on the potential of new technologies, which might result in faster adoption and therefore improved patient outcomes. Using HTA during a clinical phase III trial (chapters 7 and 8) showed that it could reveal crucial contextual factors that request anticipation upon when aiming to diffuse a new technology. These insights are expected to improve the adoption and implementation of promising new technologies and are thus expected to be beneficial to patients. In the final chapters (chapters 9 and 10), HTA was used in an innovation that has already been used for multiple years but is still not reimbursed. The results could inform policy-makers, clinicians, and insurance companies in reimbursement and clinical guideline decisions. It became apparent that performing HTA in various phases of technical development serves a different type of stakeholder.

Based on the methodological issues faced in the analyses, we listed recommendations and areas for further research. First, methods should be developed on how to incorporate the full range of relevant effects in a CEA to inform adoption and reimbursement decisions more effectively. Second, we advocate that guidelines should be developed regarding the most appropriate methodology for cost analyses to increase the generalizability and uniformity of health economic evaluations and its results. Third, we advise using bottom-up costing analyses when possible as this allows us to assess the impact of utilization rates on the cost-effectiveness results. Fourth, to choose the most appropriate HTA method in a specific development phase, it is advised to incorporate the (expected) duration of the translational pathway as a factor, next to the development phase itself, and the expected benefit of the technology.

For policy-makers, we would recommend to use HTA as early as possible and to use it iteratively to decrease the risk of failure of promosing innovations and guide the development process towards patient access most optimally. To stimulate the use of early HTA, HTA researchers are advised to seek collaborations with clinicians and medical device companies. It is recommended to request early HTA data in application rounds for research grants to further stimulate its use. This data is also expected to support the grant selection process;

choosing the best research to fund. We also see potential in using early HTA data (e.g. VOI analyses) in adoption decisions of promising innovations and suggest to start a pilot to evaluate the benefit/risk ratio of granting early access to innovations (based on early data) instead of waiting for higher-level evidence. Finally, we recommend to include uniform cost data collection in (large) national registries to improve the generalizability of health economic outcomes and increase the use of bottom-up costing methods.

In this dissertation, we informed multiple stakeholders on development, implementation, and reimbursement decisions of promising innovations in oncology, throughout the translational research process. We provided (methodological) areas for future research and provided policy recommendations that will improve the use of HTA alongside the translational pathway.

REFERENCES

1. Kok, E. N. D. et al. Association of image-guided navigation with complete resection in patients with locally advanced primary and recurrent rectal cancer: a nonrandomized trial. JAMA Netw. Open 3, e208522–e208522 (2020).
2. Retèl, V. P., Steuten, L. M. G., Mewes, J. C. & van Harten, W. H. Early Cost-Effectiveness Modeling for Tumor Infiltrating Lymphocytes (TIL)-Treatment Versus Ipilimumab in Metastatic Melanoma Patients. *Value Heal. J. Int. Soc. Pharmacoeconomics Outcomes Res.* **17**, A640 (2014).

A

SAMENVATTING

Het doel van dit proefschrift is om bij te dragen aan de kennis over de toepassing van gezondheids-technologische evaluaties (HTA) en de mogelijke methodes die hiervoor gebruikt kunnen worden tijdens het translationele proces van een medische innovatie in de oncologie. Hiermee beogen we om een begin te maken met positioneren van de vroege inzet van deze methoden (vroege HTA) voor de evaluatie van medische innovaties.

De budgetten van de gezondheidszorg staan onder druk door de stijgende zorgkosten. Mede door deze stijging is het van essentieel belang dat dit op de juiste manier wordt gespendeerd. Nieuwe technologieën moeten daarom hun meerwaarde bewijzen ten opzichte van de mogelijke additionele kosten. Dit wordt geëvalueerd met een kosten-effectiviteitsanalyse (KEA), welke vaak pas aan het eind van ontwikkeling van een nieuwe technologie wordt uitgevoerd. Echter, wanneer in deze fase de technologie niet kosten-effectief blijkt te zijn, wordt de kans dat de technologie op de markt komt klein, wat kan worden gezien als verspilling van kostbare tijd en onderzoeksgelden. Door een KEA eerder in het proces op te nemen zouden deze situaties voorkomen kunnen worden. Daarnaast is er op dat moment nog ruimte om te anticiperen op factoren die van invloed kunnen zijn op de kosten-effectiviteit.
Dit proefschrift is ingedeeld op basis van het moment dat een innovatie geëvalueerd wordt (bijvoorbeeld bij een zeer vroege HTA). Elk hoofdstuk beoogt de relevante belanghebbende te informeren over bijvoorbeeld mogelijke vervolgstappen in zowel het ontwerp als de opzet van toekomstig onderzoek of om bijvoorbeeld informatie te leveren voor keuzes met betrekking tot de implementatie en vergoeding.

In **Hoofdstuk 1** wordt achtergrondinformatie gegeven over het translationale proces van innovaties, de weg vanaf idee tot aan het gebruik van een innovatie in het ziekenhuis, en de mogelijke barrières die men kan tegenkomen in dit proces. Verder wordt het gebruik en de toepassing van HTA uitgelegd en worden verschillende methodes gepresenteerd die per fase van het translationele proces te gebruiken zijn.

Deel I - Zeer vroege HTA

In **hoofdstuk 2**, zijn drie innovatieve chirurgische instrumenten geëvalueerd door middel van interviews (multiple criteria decision analysis (MCDA)) om de toekomstige onderzoek en ontwerp stappen te begeleiden. Hierbij werd gebruik gemaakt van een analytisch hiërarchisch proces (AHP) om de verwachte meerwaarde van navigatie technologie, optische beeldvorming (spectroscopie) en augmented reality te evalueren ten opzichte van de standaard zorg voor vijf oncologische indicaties. Hiervoor werden 16 beslissingscriteria gekozen welke bijdragen

aan het succes van een operatie. De invloed van deze criteria en de verwachte prestatie van de technologieën op de criteria zijn geëvalueerd door chirurgen. De combinatie van deze uitkomsten gaf uiteindelijk de verwachte waarde van elke technologie. De verwachte invloed van de criteria bleek te verschillen per indicatie, bijvoorbeeld patiënten tevredenheid was belangrijker voor borstkanker operaties dan voor de andere 4 indicaties. De combinatie van de zwaarte van de criteria en de verwachte prestatie van de technologieën op deze criteria leerde ons dat de chirurgen het meest verwachten van het gebruik van navigatie technologie bij het verwijderen van lymfeklieren, levertumoren en rectumtumoren. Voor het gebruik van optische beeldvorming wordt verwacht dat deze waarde toevoegt bij het verwijderen van tong- en rectum tumoren. De toevoeging van augmented reality werd met name als veelbelovend gezien bij tong tumor resecties. Geen van de technologieën toonde een overtuigend voordeel voor de inzet bij borstkanker operaties. Voor met name de navigatie technologie en optische beeldvorming wordt meerwaarde verwacht ten opzichte van de huidige zorg. Het lijkt daarom de moeite waard om de technologieën verder te ontwikkelen voor deze indicaties. Op basis van deze analyse lijkt het nuttig te zijn om dit soort analyses (MCDA/AHP) in te zetten om vervolgonderzoek te prioriteren en technische ontwikkelingen te sturen.

In **hoofdstuk 3** is een systematisch literatuuronderzoek uitgevoerd gericht op de verschillende beeldvormende technologieën die gebruikt kunnen worden om vroege respons op neoadjuvante chemotherapie (NACT) te evalueren bij borstkanker patiënten. Omdat eerdere onderzoeken hebben laten zien dat de betrouwbaarheid van deze beeldvormende technologieën afhangt van het type borstkanker is er onderscheid gemaakt tussen borstkankersubtypen (ER-status en HER2-status). Er zijn in totaal 229 artikelen geïdentificeerd, waarvan er 30 volledig zijn gelezen en uiteindelijk 15 observationele studies zijn geïncludeerd. In ER-positieve en HER2-negatieve borstkanker patiënten laat ^{18}F-FDG-PET/CT de beste resultaten zien (hoogst waargenomen sensitiviteit van 89%). Bij triple-negatieve borstkankerpatiënten zijn de beste resultaten gevonden voor ^{18}F-FDG-PET/CT na 2 NAC cycli (hoogst waargenomen sensitiviteit van 79%), maar ook de resultaten van MRI zijn veelbelovend. Bij HER2-positieve patiënten scoort ook ^{18}F-FDG-PET/CT het best, maar ook MRI blijkt veelbelovend. We adviseren om vervolgonderzoek te starten naar het gebruik van MRI bij HER2-positieve en triple-negatieve borstkankerpatiënten. Op basis van deze resultaten kunnen echter alleen voorzichtige conclusies worden getrokken omdat de onderzoeken erg verschillend zijn uitgevoerd, er nog geen consensus lijkt over de uitkomstmaten die gebruikt worden en de bewijskracht van de meeste studies laag zijn. Dit hoofdstuk concludeert daarom dat het huidige bewijs onvoldoende is om specifieke aanbevelingen te doen met betrekking tot welke beeldvormende technologie gebruikt zou moeten worden per borstkankersubtype.

Deel II - Vroege HTA: tot en tijdens de eerste klinische onderzoeken (fase I)

In het Nederlands Kanker Instituut (NKI) is een nieuwe technologie ontwikkeld voor het lokaliseren van niet-voelbare borsttumoren tijdens de operatie: lokalisatie met een magnetisch zaadje (MSL). Naar verwachting is MSL net zo effectief als technologieën die momenteel gebruikt worden: draadgeleide lokalisatie (WGL) en lokalisatie door het gebruik van een radioactief zaadje (RSL), maar is MSL in staat de nadelen van RSL en WGL te overwinnen. In **hoofdstuk 4** is de invloed op de Nederlandse zorgkosten geëvalueerd wanneer MSL geïmplementeerd zou worden. De interventie- en implementatiekosten van de drie technologieën zijn bottom-up geëvalueerd in enkele Nederlandse ziekenhuizen. Omdat de prijs van het magnetische zaadje nog bepaald moet worden, is er met een range gerekend. De verwachte interventiekosten van MSL liggen hoger dan van RSL en WGL. De verwachte implementatiekosten van MSL zijn echter substantieel lager dan van RSL. Wanneer MSL wordt geïmplementeerd en het magnetische zaadje zou €100 per zaadje kosten, dan zijn de totale zorgkosten lager dan wanneer MSL niet wordt geïmplementeerd (huidige situatie). Als het magnetische zaadje €200 of meer kost, is de standaardzorg in het voordeel. Op basis van deze vroege analyse concluderen we dat MSL een kosten efficiënte technologie zou kunnen zijn voor niet-voelbare borstkanker tumoren.

Veel ziekenhuizen hebben in de afgelopen jaren hybride operatiekamers (OK) laten bouwen. Aangezien middelen beperkt zijn, moeten deze OKs aantonen dat ze van toegevoegde waarde zijn voordat deze zorg kan worden vergoed. Huidige inschattingen voor de kosten van een OK laten grote variatie zien, terwijl de kosten van een hybride OK tot op heden nog niet zijn onderzocht. Daarom zijn in **hoofdstuk 5** de kosten en de drijfveren voor een conventionele en hybride OK geëvalueerd door middel van een uitgebreide bottom-up kostenanalyse in vijf Nederlandse ziekenhuizen, waarbij de volgende kostencategorieën werden geëvalueerd: bouw, inventaris, personeel en overhead. Deze resultaten zijn een eerste stap naar het evalueren van de meerwaarde van de hybride OK. De gevonden kosten per minuut zijn respectievelijk €9.45 (€8.60 - €10.23) en €19.88 (€16.10 - €23.07) voor de conventionele en hybride OK. De personeelskosten hebben het meeste effect op de kosten voor de conventionele OK, terwijl voor de hybride OK de bezettingsgraad van de OK de grootste impact heeft, gevolgd door de kosten van de inventaris. Op basis van deze resultaten zal de hybride OK een substantieel voordeel moeten laten zien om beoordeeld te kunnen worden als kosten-effectief.

Gebaseerd op de resultaten van hoofdstuk 2 bleek met name het gebruik van het navigatiesysteem bij verschillende indicaties veelbelovend. De onderzoeksgroep in het NKI zette het onderzoek voort en toonde aan dat het risico op het hebben van positieve snijranden lager was in patiënten die werden geopereerd met navigatie ten opzichte van

een historische controlegroep. Dit werd onderzocht bij patiënten met lokaal recidief en lokaal gevorderd rectum kanker (LRRC en LARC)[1]. Op basis van deze eerste resultaten is in **hoofdstuk 6** bekeken of het gebruik van het navigatie systeem bij LARC en LRRC patiënten de potentie heeft om als kosten-effectief beoordeeld te worden. Op basis van de klinische gegevens van Kok et al. en kwaliteit van leven, data uit een lopende prospectieve studie in het NKI, is een vroege KEA uitgevoerd. Om de verwachte kosten en effecten te evalueren werd een Markov beslissingsmodel opgesteld. Dit model toonde dat het gebruik van navigatie voor beide indicaties resulteert in een hoger aantal levensjaren in perfecte gezondheid (QALYs) en hogere kosten ten opzichte van de standaard operatie. De scenario analyse geeft aan dat optimaal gebruik van het navigatie systeem kan leiden tot een acceptabele meerprijs ten aanzien van de gewonnen effecten voor de Nederlandse situatie. Op basis van deze vroege evaluatie concluderen we dat de inzet van het navigatie systeem waarschijnlijk kosten-effectief is bij LRRC patiënten en dat het de kans heeft om kosten-effectief te worden bij LARC patiënten. Om de onzekerheid rondom de resultaten weg te nemen raden we aan om een klinische studie te starten om het opereren met het gebruik van navigatie te vergelijken met de huidige situatie in dezelfde patiëntenpopulatie. Deze data zou vervolgens gebruikt kunnen worden in dit kosten-effectiviteits model om te besluiten over de implementatie en vergoeding van deze innovatie.

Deel III - Vroege HTA: zowel fase I/II studies

We weten dat de klinische implementatie van geavanceerde celtherapieën (advanced therapeutic medicinal product (ATMP)) complex kan zijn. In **hoofdstuk 7** is de vroege klinische implementatie van een veelbelovende ATMP: TIL-therapie bij gevorderd melanoom, geëvalueerd middels een constructieve technologie analyse (CTA) . Sinds 2011 is in het NKI TIL-therapie geïntroduceerd als een experimentele therapie en is deze in 2014 opgenomen in een voorwaardelijk toelatingstraject (coverage with evidence development (CED)). In dit hoofdstuk is eerst literatuur samengevat over barrières en bevorderende factoren die bekend zijn bij klinische implementatie van ATMPs. Vervolgens zijn er met 26 stakeholders semi-gestructureerde interviews gehouden waarbij vragen werden gesteld gerelateerd aan de 6 CTA domeinen: klinisch, economisch, patiënt-gerelateerd, organisatorisch, technisch en toekomst. Voor het economische domein is ook een bottom-up kostprijs analyse gedaan om de behandelkosten van TIL-therapie in kaart te brengen. De barrières die voor TIL-therapie zijn geïdentificeerd, kwamen grotendeels overeen met de barrières die in de literatuur zijn gevonden. De analyse identificeerde de volgende bevorderende factoren voor de implementatie van TIL-therapie: financiële steun vanuit het voorwaardelijke toelatingstraject, de beschikbaarheid van een apotheker in het ziekenhuis, deskundigheid rondom kwaliteit van ATMPs en een lab medewerker met ervaring in het opkweken van TILs. Op basis van deze resultaten is geconcludeerd dat de implementatie van TIL-therapie complex blijft, maar

wel haalbaar is. De resultaten van de klinische studie (welke momenteel nog patiënten rekruteert) waarin TIL-therapie vergeleken wordt met de huidige standaardzorg (ipilimumab), zal naar verwachting de adoptie van TIL-therapie vergemakkelijken. Omdat het verkrijgen van voldoende financiële ondersteuning één van de belangrijkste in literatuur beschreven barrières is voor het implementeren van ATMPs, raden we nationale zorgautoriteiten aan om innovatieve en voorwaardelijke vergoedingsprogramma's te evalueren en in te zetten voor de ondersteuning van dit soort veelbelovende therapieën.

Op basis van de geïdentificeerde barrières en de genoemde verwachte veranderingen met betrekking tot TIL-therapie in hoofdstuk 7, zijn in **hoofdstuk 8** adoptiescenario's voor TIL-therapie bij gevorderd melanoom opgesteld. De waarschijnlijkheid dat deze scenario's tussen nu en 5 jaar kunnen voorkomen is middels een vragenlijst onder internationale experts geëvalueerd en de meest waarschijnlijke scenario's zijn opgenomen in een bestaand kosten-effectiviteit model[2]. De scenario's bevatten verschillende contextuele factoren, zoals concurrerende therapieën die op de markt kunnen komen, ontwikkelingen rondom TIL-therapie, invloed van farmaceutische bedrijven en de houding van patiënten en clinici ten aanzien van de behandeling. Dit hoofdstuk beoogde om de beslissing om TIL-therapie te implementeren te informeren. In totaal werden 14 scenario's opgesteld waarvan de waarschijnlijkheid of deze scenario's kunnen optreden is geëvalueerd bij 29 experts uit 12 landen. De gemiddelde waarschijnlijkheid varieerde tussen 29% en 58%, waardoor toekomstige ontwikkelingen omtrent TIL-therapie onzeker lijken. Op basis van de waarschijnlijkheid, recente literatuur en gesprekken met interne experts worden 8 scenario's als waarschijnlijk geacht. Zes hiervan zijn opgenomen in het kosten-effectiviteit model. In de meeste scenario's blijkt TIL-therapie, vergeleken met de huidige standaard, nog steeds kosteneffectief te zijn. De factoren in deze scenario's kunnen bevorderend werken voor de implementatie van TIL-therapie. Echter, wanneer de TIL productie wordt uitbesteed of TIL-therapie wordt gegeven in combinatie met ipilimumab is het onwaarschijnlijk dat TIL-therapie kosteneffectief blijft. Deze factoren moeten dus in overweging worden genomen in de beslissing om TIL-therapie te implementeren.

Deel IV - Reguliere HTA

Ondanks vele studies naar de effectiviteit van robot- geassisteerde verwijdering van de prostaat (RARP), is er geen definitieve conclusie over de toegevoegde waarde van RARP. In **hoofdstuk 9**, presenteren we de resultaten van een nationale retrospectieve clusterstudie welke de lange termijn effecten (6-9 jaar) op zowel incontinentie als erectiele dysfunctie evalueerde na RARP en laparoscopische verwijdering van de prostaat (LRP). Deze evaluatie is gebaseerd op data van 12 Nederlandse ziekenhuizen. Klinische kenmerken, patiënten ervaringen en per- en postoperatieve uitkomsten zijn verzameld van patiënten die geopereerd zijn tussen 2010 en

2012 in één van de geselecteerde ziekenhuizen. Primaire uitkomsten zijn seksueel en urinair functioneren gemeten door de EPIC-26 vragenlijst. In totaal namen 1370 patiënten deel aan de studie waarvan 907 RARP ontvingen en 463 LRP. Ongeveer 7 jaar na de operatie zien we een klinisch relevant en significant betere score voor continentie na RARP vergeleken met LRP. Er is geen significant verschil gevonden in seksueel functioneren. De regressieanalyse toont aan dat betere functionele resultaten samenhangt met het ontvangen van een zenuwsparende procedure, het niet ontvangen van adjuvante radiotherapie en een jongere leeftijd bij het ondergaan van de procedure. Op factoren zoals de duur van de procedure, bloedverlies en kans op een zenuwsparende operatie scoorde de RARP groep ook beter dan de LRP groep. Deze resultaten kunnen de huidige richtlijnen veranderen en we adviseren daarom de beleidsmakers hun standpunt ten aanzien van vergoeding te heroverwegen, in het bijzonder wanneer RARP kosten-effectief blijkt te zijn.

Op basis van de resultaten beschreven in hoofdstuk 9, presenteert **hoofdstuk 10** een KEA waarin RARP vergeleken wordt met LRP vanuit een Nederlands perspectief. Omdat de oncologische uitkomsten vergelijkbaar waren, richtte de analyse zich op functionele uitkomsten en het gebruik van aanvullende zorg bij klachten van incontinentie en erectiele dysfunctie. Met een bottom-up kostenanalyse in vijf ziekenhuizen (2 voor LRP, 3 voor RARP) zijn de interventiekosten voor RARP en LRP geëvalueerd. De kosten voor aanvullende zorg zijn gebaseerd op informatie aangeleverd door de patiënten (bijv. het type zorg dat werd ontvangen en de frequentie van deze zorg) en Nederlandse referentieprijzen. Om de kosten en effecten in QALYs over een periode van 7 jaar te evalueren is een beslisboom gemaakt. De interventiekosten voor RARP (€9,964) zijn hoger dan de kosten van LRP (€7,253). Deze analyse laat zien dat RARP duurder is, maar ook effectiever vergeleken met LRP wat resulteerde in een ICUR van €34.206 per gewonnen QALY. Dit valt binnen de Nederlandse grens voor maximale meerkosten per QALY van €80,000. Het meest optimistische scenario, wanneer RARP wordt gecentraliseerd (>150 patiënten/jaar), verlaagde de totale kosten vanwege optimaler gebruik van de technologie, de kortere procedure en een kortere ligduur wat resulteerde in een ICUR van €3,495 per gewonnen QALY. Deze studie concludeert dat RARP kosten-effectief is vergeleken met LRP, wat een duidelijke stimulans is om RARP volledig te vergoeden, zeker wanneer RARP gecentraliseerd wordt aangeboden.

In **hoofdstuk 11** worden de belangrijkste bevindingen van dit proefschrift besproken. De resultaten van het gebruik van HTA in een zeer vroege fase (hoofdstukken 2 en 3) waren informatief voor onderzoekers en ontwikkelaars in het sturen van toekomstige onderzoeksactiviteiten en kunnen bijdragen aan effectievere ontwikkeling en daarmee versnelling van de toegang van innovatieve zorg aan de patiënten. De resultaten van de HTA-onderzoeken die vóór of tijdens de eerste klinische studie werden uitgevoerd (hoofdstukken 4, 5, 6) konden worden gebruikt om de verdere (technologische) ontwikkeling

te sturen, om de klinische studies op te zetten en de prijs van de technologie te bepalen. Deze hoofdstukken gaven ook aan dat vroege KEAs – hoewel omringt door onzekerheid – beleidsmakers of investeerders kunnen informeren over het te verwachten potentieel, wat vervolgens kan bijdragen aan een snellere acceptatie van de desbetreffende technologie en betere uitkomsten voor patiënten. Door het toepassen van HTA tijdens een klinische fase III studie (hoofdstukken 7 en 8) zijn cruciale factoren gevonden, welke actie vereisen wanneer beoogd wordt de technologie te adopteren. De identificatie van deze factoren kan de adoptie en implementatie van nieuwe technologieën verbeteren, wat kan leiden tot betere uitkomsten voor de patiënten. In de laatste twee hoofdstukken (hoofdstukken 9 en 10) is HTA toegepast bij een innovatie die al meerdere jaren gebruikt wordt, maar nog niet vergoed wordt. De resultaten kunnen beleidsmakers, clinici en verzekeringsmaatschappijen ondersteunen in besluitvorming over klinische richtlijnen en vergoedingen. De resultaten uit deze hoofdstukken geven aan dat het uitvoeren van HTA in verschillende fasen van technische ontwikkeling verschillende type stakeholders kan dienen.

Op basis van verschillende methodologische aspecten die we tegenkwamen in de uitvoering van onze analyses zijn aanbevelingen en richtingen voor vervolgonderzoek beschreven. Ten eerste zullen er methodes moeten worden ontwikkeld waarmee het volledige scala aan relevante effecten kan worden geëvalueerd en daarmee kunnen worden meegenomen in adoptie- en vergoedingsbesluiten. Ten tweede zouden er richtlijnen moeten worden opgezet voor het doen van kostenanalyses waardoor de resultaten van gezondheidseconomische analyses vanuit verschillende landen beter vergeleken kunnen worden. Ten derde adviseren we om bottom-up kostenanalyses uit te voeren, wanneer mogelijk, omdat dit het mogelijk maakt om het effect van de bezettingsgraad van de technologie op de kosten-effectiviteit te evalueren. Tot slot, adviseren we om in de keuze voor de meest geschikte HTA-methode, naast het moment in de ontwikkeling en de verwachte voordelen van de technologie, de verwachte snelheid van het adoptieproces mee te nemen. Wanneer een innovatie langzamer wordt opgenomen kan het namelijk zinvol zijn om een meer kwalitatieve onderzoeksmethode te gebruiken.

Voor beleidsmakers raden we aan om HTA zo vroeg mogelijk te gebruiken en het iteratief in te zetten om het ontwerp en ontwikkelproces zo optimaal mogelijk te begeleiden. Om dit te stimuleren is het aan te raden dat HTA-onderzoekers samenwerking zoeken met clinici en bedrijven welke medische hulpmiddelen ontwikkelen. Daarnaast stellen wij voor om HTA gerelateerde gegevens op te vragen in aanvraagrondes voor subsidies. Deze informatie kan de subsidieverstrekkers ook helpen om de beste onderzoeken te kiezen voor het ontvangen van subsidie. We zien ook potentie in het gebruik van vroege HTA- gegevens (bijv. VOI-analyses) bij beslissingen over adoptie van veelbelovende innovaties waar nog geen fase III resultaten beschikbaar zijn. We stellen daarbij voor om een pilot te starten om de verhouding

tussen baten en risico's te evalueren voor het verlenen van vroege toegang tot innovaties (op basis van vroege gegevens) in plaats van te wachten op sterker bewijs. Ten slotte wordt geadviseerd om gegevens over kosten uniform te verzamelen in (grote) nationale registraties om de vergelijkbaarheid van gezondheids-economische evaluaties te verbeteren en het gebruik van bottom-up kostenanalyses te verhogen.

In dit proefschrift hebben we verschillende partijen geïnformeerd over de verdere ontwikkeling, implementatie en vergoeding van een innovatie in de oncologie. Daarnaast hebben we (methodologische) gebieden voor vervolgonderzoek aangegeven en aanbevelingen gedaan voor beleidsmakers waarmee het gebruik van HTA langs het translationele proces kan worden verbeterd.

REFERENTIES

1. Kok, E. N. D. *et al.* Association of image-guided navigation with complete resection in patients with locally advanced primary and recurrent rectal cancer: a nonrandomized trial. *JAMA Netw. Open* **3**, e208522–e208522 (2020).
2. Retèl, V. P. *et al.* Early cost-effectiveness of tumor infiltrating lymphocytes (TIL) for second line treatment in advanced melanoma: a model-based economic evaluation. *BMC Cancer* **18**, 895 (2018).

A

DANKWOORD

En toen was het ineens af?! Ondanks dat dit boekje nu bijna voor me ligt, vind ik het maar moeilijk te geloven dat mijn tijd in het Antoni van Leeuwenhoek ziekenhuis (AVL) en het samenwerken met alle fijne collega's echt klaar is. In dit hoofdstuk wil ik graag een aantal personen persoonlijk bedanken voor hun rol bij het realiseren van het proefschrift.

Prof. dr. W.H. van Harten, beste Wim, toen ik in 2015 vanuit jouw vak Quality and Safety op de Universiteit Twente een dag binnen mocht kijken in het AVL was ik meteen enthousiast. Dat ik mijn afstudeeropdracht mocht uitvoeren onder jouw begeleiding (toen nog lid van de RvB) in het AVL vond ik dan ook heel bijzonder. Ondertussen zijn we 5 jaar verder en wil ik graag mijn bewondering en dank uitspreken. Allereerst, bedankt dat je me mogelijkheid gaf om dit PhD project uit te voeren en dat je me hierin wilde begeleiden. Dit zijn voor mij vijf leuke en leerzame jaren geweest. Daarnaast heb ik een grote bewondering voor jouw manier van werken; hoe snel en scherp je op mijn (maar ook van mijn mede-promovendi) vragen of stukken kon reageren ondanks je vele andere activiteiten. Ondanks dat je meestal in Rijnstate was, voelde het nooit alsof je ver weg of niet betrokken was. Je scherpe en overstijgende commentaren hebben niet alleen bijgedragen aan de kwaliteit van de stukken, maar zeker ook mijn eigen ontwikkeling. Daarnaast heb ik je directe en heldere communicatie altijd erg gewaardeerd, na een overleg kon ik (meestal) gelijk weer verder. Tot slot vond ik het heel prettig dat er veel vertrouwen was en ruimte geboden werd om mijn projecten op te zetten en uit te voeren. Bedankt voor alles!

Dr. V.P. Retèl, beste Valesca, je was samen met Wim al mijn begeleider bij mijn afstudeerproject. Ik had me bij dit PhD traject geen fijnere co-promotor kunnen wensen (wie kan nou zeggen dat ze met haar co-promotor heeft opgetreden?). Al snel merkte ik dat we aardig wat raakvlakken hadden op zowel persoonlijk als professioneel vlak wat het erg prettig maakte om samen te werken. Ik wil je heel erg bedanken voor je toegankelijkheid, bereidheid om altijd mee te denken en je persoonlijke aandacht. Als ik ergens tegen aan liep was jij degene die de vinger op de zere plek kon leggen waardoor het bespreekbaar werd. Vervolgens leek je het volste vertrouwen te hebben dat het daarna wel weer goed kwam, wat bij mij zorgde voor veel energie en mij het gevoel gaf dat ik op mezelf kon vertrouwen. Onze overleggen waren nuttig, gezellig en gingen geregeld ook over andere HTA gerelateerde zaken dan mijn PhD project an sich. In het AVL ben jij de spin in het web als het gaat om HTA, jouw activiteiten om HTA te vertalen in de praktijk en in te zetten om de beste zorg snel bij de patiënt te krijgen was en is heel inspirerend. Ik hoop dat we contact houden en wie weet komt er nog eens een gezamelijk project op ons pad. We kunnen in elk geval nog één paper afmaken.

A

Geachte prof. dr. A.J.M. van den Eertwegh, prof. dr. A. Klink, dr. ir. H. Koffijberg, prof. dr. J.P Ruurda, dr. N. van der Vaart en prof. dr. ir. R.M. Verdaasdonk, heel hartelijk bedankt dat u mijn proefschrift heeft beoordeeld en zitting heeft willen nemen in mijn promotiecommissie. Beste Erik, helaas is er geen gezamenlijk project gestart m.b.t. de hybride OK tussen de UT en het NKI, maar dit was wel de basis voor onze kennismaking. Bedankt voor jouw altijd oprechte interesse in de status van mijn projecten en je bereidheid om zijdelings mee te denken.

Voor mijn project gericht op het evalueren van de Da Vinci robot bij prostaatkanker wil ik in de eerste plaats dr. H. van der Poel heel hartelijk bedanken. Je was al vroegtijdig betrokken bij de opzet van het project en gaf toen al aan dat dit geen eenvoudig project zou worden. Toch is het gelukt om binnen 12 ziekenhuizen data verzamelen, mede dankzij jouw netwerk. Je bent een krachtige verbinder en denkt razendsnel met de ander mee. Bedankt voor jouw sterke analytische denkwijze, klinische inzichten en commentaren tijdens het project en het schrijven van de manuscripten, wat ze absoluut naar een hoger niveau getild heeft. Heel erg bedankt en ik hoop dat onze database in de toekomst nog van pas komt.

Ook gerelateerd aan het Da Vinci project wil ik drs. C. Wijburg bedanken. Toen eenmaal de studieopzet uitgedacht was, was jij al snel betrokken als externe partij (Rijnstate) waar we onze studie konden piloten. Ik kwam met plezier naar het Rijnstate en jouw enthousiaste houding t.a.v. mijn project heb ik zeer gewaardeerd. Dit maakte al die huisartsen bellen en brieven vouwen een stuk aagenamer. Je betrokkenheid en commentaren bij de manuscripten en gedurende het project heb ik zeer gewaardeerd. Ik wil daarnaast een shout-out doen naar drs. E. Collette, zonder jou had de studieopzet er waarschijnlijk anders uitgezien. Jammer dat we de studie uiteindelijk niet samen hebben uit kunnen voeren, maar gelukkig kwam je – wanneer ik in Rijnstate bezig was – geregeld even polshoogte nemen, wat ons gelijk de mogelijkheid gaf om even bij te praten.

Dankzij de samenwerking met de 12 Nederlandse ziekenhuizen is er uiteindelijk data verzameld van 1370 patiënten. Ik wil graag de urologen en de betrokken verpleegkundig specialisten en onderzoeksverpleegkundigen bedanken voor de participatie, gastvrijheid en de ondersteuning bij dit onderzoek. Bedankt drs. O.S. Klaver en het team in Maasstad (Maasstad ziekenhuis), drs. C. Wijburg en A. Aalbers (Rijnstate), drs. L.M.C.L. Fossion, drs. K. de Laet en J. de Rooij (Maxima Medisch Centrum), drs. J.B.W. Rietbergen en M. van Gelder (Sint Franciscus Gasthuis), drs. B.C. Knipscheer en J. Boertien (Treant zorggroep), drs. Joost Leijte (Amphia), drs. M.B. Busstra en V.J.E.C. Werdmuller von Elgg (Erasmus Medisch Centrum), dr. R.P. Meijer en M. van Elst (Universitair Medisch Centrum Utrecht), drs. Hillenius en M. Merks (Bravis), dr. A. Kooistra (Meander Medisch Centrum), dr R.F.M. Bevers en C. de Jong (Leids Universitair Medisch Centrum), dr. H.G. van der poel, C. Tillier en E. van Muilekom (Antoni van Leeuwenhoek). Ondanks dat het verzamelen van de data een hectische tijd was, heb ik ervan

genoten om met elk van jullie kennis te maken en in de verschillende ziekenhuizen een kijkje te mogen nemen (inclusief de meegelopen operaties), dank hiervoor!

Specifiek voor de kosten-analyse wil ik nog graag de betrokkenen van het Maasstad ziekenhuis, het Rijnstate, Maxima Medisch Centrum, Leids Universitair Medisch centrum en het Antoni van Leeuwenhoek ziekenhuis bedanken voor hun inzet.

Ik wil ook graag de leden van mijn OOA commissie: prof. dr. ir. F.E. van Leeuwen, prof. dr. L.V. van der Poll, prof. dr. T.J.M. Ruers en prof. dr. J.B.A.G. Haanen bedanken. Ondanks dat deze bijeenkomsten slechts jaarlijks plaatsvonden heb ik deze evaluaties als zeer prettig en motiverend ervaren.

Prof. dr. T.J.M. Ruers, beste Theo, bedankt dat je me zo warm hebt ontvangen in jouw onderzoeksgroep en me de kans gaf te connecten met alle onderzoekers. Ik heb altijd met veel plezier gewerkt aan de projecten die voortkwamen uit jullie ontwikkelingen, al was het soms een puzzel. Fijn dat we – als het nodig was – snel een afspraak in konden plannen en inhoudelijk over de analyses konden praten. Bedankt voor je interesse in het HTA gebied en je soms kritische en klinische vragen. Ik heb veelvuldig gebruik kunnen maken van je NKI netwerk wat absoluut heeft bijgedragen aan de kwaliteit van de analyses. Ik hoop dat de analyses bij hebben gedragen aan jullie R&D plannen. Ik blijf jullie groep in elk geval met veel interesse volgen.

De andere onderzoekers van de image-guided groep welke ook betrokken zijn geweest wil ik ook graag bedanken: dr. Kuhlmann, Nikie Hoetjes, Bram Schermers, Bas Pouw, Esther Kok, Lisanne de Boer, Lisanne Baltussen, Susan Brouwer de Koning, Esther Kho, Jasper Nijkamp, Roeland Eppenga, Jasper Smit en Ruben van Veen. Dank voor het meedenken en laagdrempelige beschikbaarheid. Het was fijn om een onderdeel te mogen zijn van jullie fijne, gezellige groep. Jasper Nijkamp, Bas, Nikie, Bram en Esther Kok, jullie wil ik nog in het bijzonder bedanken voor jullie nauwe betrokkenheid en de fijne overleggen en samenwerking rondom de OCR en TALENT studie.

Gerelateerd aan deze projecten wil ik ook graag alle chirurgen bedanken die mee hebben gedaan aan de vroege waarde evaluatie van deze nieuwe chirurgische tools. In het bijzonder wil ik prof. dr. G. Beets noemen. Bedankt voor jouw klinische input in de kosten-effectiviteits analyse gericht op de inzet van navigatie bij colorectale tumoren.

Sejal Patel, wij leerden elkaar kennen via onze copromotors op LoLa 2018. Samen hebben we een mooi project op kunnen zetten met een raakvlak naar onze beide promotie trajecten. Het was nuttig, maar vooral ook heel leuk om eens een project samen met iemand te doen en

daarnaast zeer leerzaam omdat we op een andere manier denken. De dagen in het Radboud, NKI of Utrecht waren altijd productief en gezellig. Het was fijn om met een 'outsider' PhD en HTA struggles te kunnen delen, heel erg bedankt daarvoor! Ik hoop dat we onze regelmatige etentjes er in kunnen houden!

Dr. J. Grutters en prof. dr. M.M Rovers wil ik bedanken voor de begeleiding en bijdrage in de kosten evaluatie. Het was zinvol om naast de begeleiding van Wim en Valesca andere input te krijgen! In dit rijtje zou ik ook graag dr. L. Poot willen noemen. Lieke, bedankt voor je inzet vanuit het Isala en je bijdrage aan het manuscript.

De kosten analyse hadden we niet uit kunnen voeren zonder deelnemende ziekenhuizen. Daarom wil ik graag de medewerkers die betrokken waren vanuit het Radboud, Isala, Jeroen Bosch Ziekenhuis, Rijnstate en Antoni van Leeuwenhoek ziekenhuis bedanken voor het verstrekken van de data en het meedenken. In particular I would like to thank prof. dr. E. Fosse for thinking along with our project and confirming the results from a Norwegian perspective.

De laatste klinische studie waar ik bij betrokken was is er één die me aan het hart gaat, de TIL studie. Het was heel bijzonder om hier deel van uit te maken. Prof. dr. J.B.A.G. Haanen, bedankt voor het meedenken in mijn hieraan gerelateerde onderzoeksprojecten en voor het betrekken van mij en Valesca bij de verschillende hurdles in het project. Dr. van den Berg, beste Joost, ook jou wil ik bedanken voor het meedenken in de artikelen en de kostprijsberekening. Ik wil ook zeker Marnix (dr. Geukes) bedanken, al was onze samenwerking maar een klein jaar, het was een fijne start om met jou laagdrempelig over de studie te sparren en mijn onderzoeksopzetten uit te denken. Drs. Rohaan, beste Maartje, jij vervulde na Marnix de arts-onderzoeker rol, dank voor de hele fijne samenwerking en altijd gezellig als je weer enkele vragenlijsten kwam langs brengen.

Ik wil ook graag de co-auteurs bedanken van mijn systematische literatuur studie welke onderdeel was van mijn afstudeerproject. Prof. dr. M.J.T.F.D. Vrancken Peters, prof. dr. J. Wesseling, prof. dr. G.S. Sonke en dr. M.P.M. Stokkel, bedankt. Dankzij jullie betrokkenheid is dit een mooi stuk geworden.

Martijn van der Schaik, bedankt voor de fijne gang ontmoetingen, het meedenken wanneer nodig en je interesse in mijn projecten.
Tevens wil ik de betrokken medewerkers van het IFC bedanken voor de hulp en betrokkenheid bij onze kosten vraagstukken. Het was voor ons, maar ook voor jullie vaak een puzzel die toch meermaals is opgelost. Bedankt voor jullie inzet!

Onderzoek doen naar de (kosten-)effectiviteit van een behandeling kan niet zonder patiënten

die mee willen doen aan dit soort studies. Ik wil hierbij dan ook alle patiënten die mee hebben gedaan aan één van de studies welke onderdeel waren van mijn werkzaamheden, CERA-PRO, OCR, TALENT en de TIL-studie, heel erg bedanken.

Naast al deze samenwerkingen was mijn standplaats 'gewoon' op de PSOE, een grote onderzoeksafdeling met veel verschillende achtergronden waardoor het een interessante en leerzame werkomgeving is. Ik zou een persoonlijk stukje voor iedereen kunnen schrijven, maar dan wordt het dankwoord nog langer, daarom noem ik een aantal van jullie bij naam. Bedankt Jacqueline, Daniëlle, Daniela, Vera, Emmie, Silvie, Kete, Barbara, Lisanne en Jacobien voor alle koffiemomentjes, de 'ik loop even langs'-momenten en momenten buiten werk om. Wat fijn dat jullie er altijd waren en we onze (PhD) ups en downs konden delen. Ik ben blij dat we samen de PSOE hebben kunnen vegroenen met alle pannenkoekplantjes. Karin en Daniëlle, bedankt dat jullie altijd klaar staan om onze vragen te beantwoorden.

In het bijzonder van de PSOE wil ik de Wim van Harten groep bedanken. Wim Groen, Anke, Laura, Willeke, Nora, Ann-Jean, Hester, Bruno, Inge, Danalyn en Joost, we zijn onderling best verschillend, maar we zijn een hele leuke en gezellige groep. Het was fijn om met jullie te kunnen sparren over mijn projecten en ik heb genoten van de gezellige borrels en etentjes! In het bijzonder wil ik de roomies (Willeke, Nora) en ex-roomies (Laura, Anke, Hester, Daniëlle) bedanken voor alles! Nora en Willeke, wat ga ik jullie missen. De uurtjes powerworking, de kwarkbollen, de Nederlandse (spreekwoorden)'les' en het feit dat we het over alles konden hebben. Het was zo fijn om met jullie de laatste 2 jaar van mijn PhD alle hoogtepunten en de wat mindere dagen te kunnen delen!
Marion en Jorrita, bedankt dat jullie er altijd waren om eventuele vragen te beantwoorden.

Ann-Jean, we waren dan wel geen kamergenoten maar wat was het fijn om jou als mijn HTA maatje te hebben! Bedankt voor alle leuke tripjes, de vele uren koffie drinken om over (onder andere) onze projecten te kunnen sparren, je waardevolle klinische input en onze bier en burger avondjes als we toch echt weer eens bij moesten kletsen. En juist dat moeten we snel maar weer eens doen!

Ik wil ook de masterstudenten bedanken die ik heb mogen begeleiden. Anne, Astrid, Ferdau en Melvin, bedankt voor jullie enthousiasme, gezelligheid en harde werk. Niet al jullie werk is uiteindelijk in mijn boekje gekomen, maar dat maakte het voor mij niet minder waardevol. Ik ben benieuwd of onze wegen wellicht nog eens kruisen. Astrid, heel leuk dat jij na ons project zo enthousiast geworden bent om ook te starten met een promotietraject! Ik hoop hier in de toekomst nog wat van mee te krijgen.

I want to thank Wilma, Heleen, Anna, Jacobien and Lisanne for the nice dinners and drinks

we had. It was great to have you close, even though you had left the NKI already. A special thanks to Lisanne and Jacobien for the open talks about everything during the many coffee breaks, borrels en dinners!

Anna, if I would not have done my internship with you, I don't think I would have even considered doing a PhD. You have been a great inspiration to me. Thanks for your guidance back then, for taking me to the borrel which really helped in feeling at home in Amsterdam, and thanks for being such an enthusiastic and loving person. I hope we can meet again soon! The Friday evening borrel in the NKI is a great opportunity to meet a lot of other researchers from the NKI. Specifically I would like to thank Nanne, Maarten and Max who had the courage to start an occasional band with me to play at my sister's wedding. I loved our band nights! Other names I want to mention are: Lorenzo, Giusi, Ewald, Andrea, Marcelo, Joao, Bruno, Christ, Tom, Renato, Georgi and Maria, thank you for the great parties, dinners and social talks! I hope I can still join you at a borrel when the Corona pandemic is over.

Dank aan ook mijn nieuwe collega's: Bert, Janne, Anam, Josepha en Isabelle voor jullie interesse in het afrondingsproces en de fijne eerste maanden bij Panaxea.

Gelukkig is er ook nog tijd om iets naast een PhD te doen. Véronique en Daphne, bedankt voor álles! Wat ben ik blij dat ik jullie de afgelopen jaren nog beter heb leren kennen! Het voelt altijd als thuis bij jullie. De vakantie tripjes, high-winen, high-tea-en, kerstmarkten en alle andere leuke uitjes die wij doen waren een fijne afleiding. Fijn dat ik altijd op jullie kan bouwen. Daph, wat was het gezellig om tijdens onze PhD regelmatig bij te kletsen en fijn dat we samen naar het einde toe konden werken! Binnenkort hebben we waarschijnlijk meer tijd voor leuke dingen, ik kijk er nu al naar uit! George en Bas, ook jullie bedankt voor de altijd gezellige avondjes als we een 'met de mannen erbij' avond doen.

De Fotonners, ondanks dat we elkaar soms maanden niet zien is het altijd als vanouds. Djurre, Robert-Jan, Mark, Joris en Janita, dank voor de gezellige en fijne wandelingen, fietsrondjes of gewoon het even bijkletsen via Skype! Prettig dat ik ook bij jullie aan kan kloppen voor tips en trics m.b.t. mijn PhD waar ik met name tegen het einde graag gebruik van maakte.

Frank, of ook wel Frank fiets, zo fijn dat jij toevallig ook net van Enschede naar Amsterdam was verhuisd! Van etentjes bij elkaar, Amsterdam ontdekken en rondjes fietsen, tot aan een nieuwjaarsavond met z'n driën (want Tessa mag hier ook zeker genoemd worden). En laten we de AMPA niet vergeten. Alhoewel we elkaar in de praktijk soms maanden niet zien is het altijd goed. Bedankt voor al die gezellige en fijne momenten!

Ondanks dat ik 2 jaar geleden ben gestopt met wedstrijddansen was en is het dansen toch

zeker één van mijn uitlaatkleppen. Ik wil hierbij de 4Happyfeeters en Amsterdancers van op en langs de vloer, de trainers, de fijne inspannende maar ontspannende danskampen in Tsjechië en de groep van het NTDS en ETDS benoemen. In het bijzonder wil ik Frank, Randy, Daniël en Mark bedanken voor ook de momenten naast het dansen. En ja Mark we gaan echt binnenkort een keer clubben! (Als het weer mag dan.)

Ik wil graag het AMCK en de AMCK-ers bedanken voor de fijne invulling van mijn dinsdagavond. Fijn samen zingen gaf me de energie die ik nodig had om weer enthousiast aan de slag te gaan. Sinds kort is hier DOS aan toegevoegd, nu vooral nog via Zoom, maar fijn hoe DOSsers nu al meeleven. Hopelijk kunnen we snel weer mooie dingen maken.

Anne, bedankt voor de fijne zanglessen. Jij was naast zangdocente ook een aardige mental coach. Heel fijn dat ik door jouw lessen gedwongen werd om regelmatig te zingen en er even alles uit te gooien. Ik waardeer je lessen enorm en kom er altijd met een bulk nieuwe energie vandaan, dank hiervoor!

Mijn paranimfen, Sus (Angela) jij bent er letterlijk altijd! Ups, downs of momenten dat ik zelf nog niet doorheb waar ik zit. Toen ik begon met mijn PhD zei ik al, “Jij wordt mijn paranimf”, ondanks dat je geen idee had wat het inhield zei je volmondig ja. In de afronding was jij het waarmee ik mijn boekje opmaakte en degene die er alle vertrouwen in had dat dit op tijd af zou komen, wat voor mij een fijne geruststelling was. Ik kon ook (zoals altijd) rekenen op je directe en ongezouten mening, maar gelukkig was er ook ruimte voor mijn eigenwijze ideeën. Bedankt voor de vele en vele telefoontjes, de uren bijslapen als ik op ‘visite’ kwam in Zeeland, je hulp bij het invullen van vragenlijsten, de relativerende en leuke momenten met Annemijn en eigenlijk gewoon voor alles. Ik ben heel blij dat jij bij dit moment achter mij staat. Frank (Frank huis), al toen we huisgenoten waren bespraken we alles rondom onze studies en ervaringen in de kliniek, maar ook zeker persoonlijke topics kwamen langs. Jij kent dus als geen ander mijn (goede en minder goede) karaktertrekjes en bij elke ontmoeting voelt het weer als thuis komen. Ondanks dat we elkaar soms maanden niet spreken was jij degene die me enorm kon inspireren en motiveren, of het nu in een hutje op de hei was of tijdens het kijken van de ETDS finales. Bedankt voor je altijd oprechte interesse, je ongevraagde maar terechte en waardevolle adviezen, je ongelimiteerde flow aan positiviteit en ideeën en juist ook alle momenten waar we het even niet over onze PhD hebben.

Johan, Eline, Dennis, Sanne en Tess ook jullie bedankt voor de interesse in mijn projecten, maar vooral voor de gezellige momenten zonder PhD onderwerpen.

Tjarco, ook jij verdient een plek hier. Ondanks dat je een promotietraject soms beschrijft als een papierfabriek konden wij regelmatig goed sparren over mijn projecten. Bedankt voor je

input en de fijne momenten in Dreister.

Dave, tijdens het ODK vertelde ik je al een klein beetje van mijn PhD, al snel daarna ontving je (voor velen wel bekend) trots van mij de meest recente versie van een plaatje of poster. Hier reageerde jij dan op met – nuchter zoals je bent – “Ik zou het niet aan mijn muur willen hangen”. Toch was jij het die me op een zondagmiddag, in ruil voor appeltaart, in het AVL kwam helpen om vragenlijsten in te voeren. En nu zo een 2 jaar later, ben je vanzelfsprekend bij mijn promotie aanwezig. Heel fijn dat ik steeds met jou over mijn projecten kon sparren en dat het altijd oké was als het me even teveel werd. Maar misschien nog wel veel fijner waren de momenten even zonder PhD. Dansen, naar het theater, een stuk fietsen, de fijne vakanties en alle andere leuke dingen die we samen doen. Dankzij Corona klinkt vreemd, maar het is wel zo dat we hierdoor in mijn laatste maanden veel samen waren wat me enorm heeft gemotiveerd. Bedankt voor je continue vertrouwen en support, je spontane ‘ik neem iets leuks mee’ en je efforts om de highlights ondanks corona te vieren. Ik kijk er naar uit om samen te gaan dansen en wellicht – als ik wat tijd over heb straks – ga ik zelfs het Magiccen of gamen wel proberen. Who knows…

Pap en mam, waar moet ik bij jullie beginnen? Dank voor ALLES. Jullie support en vertrouwen heeft er zoveel aan bijgedragen dat dit boekje er nu daadwerkelijk ligt. Jullie zijn altijd ontzettend betrokken geweest, gingen met me mee naar een open dag in het AVL, kennen al mijn collega’s bij naam en kwamen me zelfs in de weekenden helpen met vragenlijsten invoeren. Bedankt voor jullie motto’s zoals: “Als je iets wil, kan je het.”, “Voorbereiden, voorbereiden, voorbereiden.” en misschien zelfs wel “Zoek het uit en vooral zelf.” die vaak in mijn hoofd spookten als ik het even niet zo scherp had. Bedankt dat ik bij jullie (het vaakst bij mam) altijd terecht kon als er weer eens een artikel afgewezen werd of ik even niet wist welke kant ik op moest gaan. Pap, dank voor de vele momenten dat je even heel nuchter met me meedacht. In de universitaire en wetenschappelijke wereld zijn jullie niet zo thuis, maar ondanks dat konden jullie me altijd helpen, supporten en helpen denken. Dank ook voor de vele fijne momenten buiten het PhD leven. Ik kijk uit naar de uitjes, etentjes, gesprekken en vakanties zonder een PhD in het achterhoofd.

Bedankt ook aan iedereen die ik alsnog vergeten ben.

A

A

LIJST VAN PUBLICATIES

Lindenberg MA, Miquel-Cases A, Retèl VP, Sonke GB, Wesseling J, Stokkel MPM, van Harten WH. Imaging performance in guiding response to neoadjuvant therapy according to breast cancer subtypes: A systematic literature review. *Critical Reviews Oncology and Hematology.* 2017;112:198-207. doi:10.1016/j.critrevonc.2017.02.014

Retèl VP, Steuten LMG, Geukes Foppen MH, Mewes JC, **Lindenberg, MA**, Haanen JBAG, Harten, WH Early cost-effectiveness of tumor infiltrating lymphocytes (TIL) for second line treatment in advanced melanoma: a model-based economic evaluation. *BMC Cancer.* 2018;18(1):895. doi:10.1186/s12885-018-4788-5

Lindenberg MA, Retel VP, van den Berg JH, Geukes Foppen MH, Haanen JB, van Harten WH. Treatment with tumor-infiltrating lymphocytes in advanced melanoma: evaluation of early clinical implementation of an advanced therapy medicinal product. *Journal of Immunotherapy.* 2018;41(9):413-425. doi:10.1097/CJI.0000000000000245

Nijkamp J, Kuhlmann KFD, Ivashchenko O, Pouw, B, Hoetjes, N, **Lindenberg, MA**, Aalbers, AGJ, Beets, G, van Coevorden, F, Kok, N, Ruers, T. Prospective study on image-guided navigation surgery for pelvic malignancies. *Journal of Surgical Oncology.* 2019;119(4):510-517. doi:10.1002/jso.25351

Lindenberg M, van Beek A, Retèl V, van Duijnhoven F, van Harten W. Early budget impact analysis on magnetic seed localization for non-palpable breast cancer surgery. *PLoS ONE.* Published online 2020. doi:10.1371/journal.pone.0232690

Lindenberg M, Retèl V, Rohaan M, van den Berg J, Haanen J, van Harten W. Evaluating different adoption scenarios for TIL-therapy and the influence on its (early) cost-effectiveness. *BMC Cancer.* 2020;20(1):712. doi:10.1186/s12885-020-07166-9

Patel S*, **Lindenberg M***, Rovers MM, van Harten WH, Ruers TJM, Poot L, Retèl VP, Grutters JPC. Understanding the costs of surgery: A bottom-up cost analysis of both a hybrid operating room and conventional operating Room [published online ahead of print, 2020 Jul 27]. *International Journal of Health Policy Management.* 2020;10. doi:10.34172/ijhpm.2020.119

Lindenberg M, Retèl V, van Til J, Kuhlmann K, Ruers T, van Harten W. Selecting image-guided surgical technologies in oncology: A surgeon's perspective, *Journal of Surgical Research.* 2021; 257; 333-343; doi: 10.1016/j.jss.2020.08.003

A

ABOUT THE AUTHOR

Melanie Lindenberg was born on March 23, 1992 in Goes, the Netherlands. In 2010 she graduated from high school at the Ostrea Lyceum in Goes and started with the Healthcare and Technology educational program (study route from HBO-V) at Saxion Applied University in Enschede at which she obtained her Bachelor of Nursing in 2014. She subsequently obtained a Master's degree in Health Sciences at the University of Twente in August 2015 (supervisors: prof. dr. W.H. van Harten and dr. V.P. Retèl). From September 2015 onwards Melanie worked on her PhD project at the Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital (NKI- AVL). She performed multiple studies evaluating innovations in oncology in order to support the further development, implementation or optimal use of these innovations, in and outside the NKI-AVL. Her research activities were supervised by prof. dr. W.H. van Harten and dr. V.P. Retèl and have resulted in this thesis. Since July 2020 Melanie is employed as an HTA researcher/consultant at Panaxea, a consultancy company aiming to support the decision making about healthcare innovations to improve health outcomes.

A