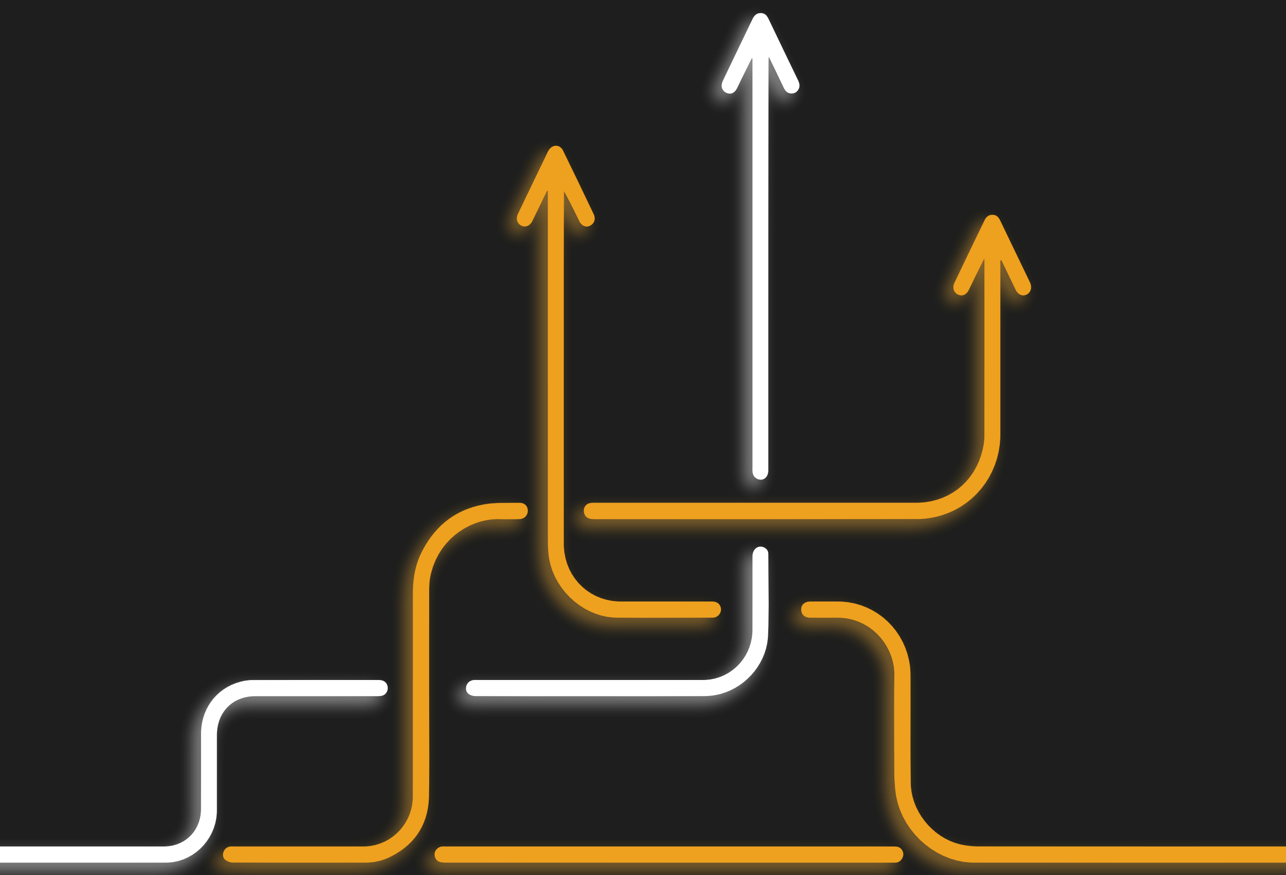
A MODELING APPROACH TO EVIDENCE-BASED TREATMENT DECISIONS

Surgical management of the neck in patients with oral cancer

Tim Govers



Cover
Wingens Webdesign

design/lay-out
Theo Hafmans Fotografie, Berg en Dal

Print
Ipskamp Drukkers, Enschede

A modeling approach to evidence-based treatment decisions

Surgical management of the neck in patients with oral cancer

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 11 maart 2016
om 16.30 uur precies

door
Tim Martin Govers
geboren op 25 juli 1989
te Heesch

Promotoren:	Prof. dr. M.M. Rovers
	Prof. dr. M.A.W. Merkx
Co-promotoren:	Dr. J.P.C. Grutters
	Dr. R.P. Takes
Manuscriptcommissie:	Prof. dr. W.J.J. Assendelft
	Prof. dr. J.H.A.M. Kaanders
	Prof. dr. J.L. Severens (Erasmus Universiteit Rotterdam)

Contents

Chapter 1	General introduction	7
Chapter 2	Lower value surgical procedures create a significant waste in healthcare: a modeling approach *Submitted*	17
Chapter 3	An international comparison of the management of the neck in early squamous cell carcinoma in the Netherlands, UK, and USA *Published: J Craniomaxillofac Surg 2016, Jan; 44 (1): 62-9*	37
Chapter 4	Innovating head and neck cancer care: an interactive evaluation *Submitted*	57
Chapter 5	Sentinel node biopsy for squamous cell carcinoma of the oral cavity and oropharynx: A diagnostic meta-analysis *Published: Oral Oncology 2013, Aug; 49 (8): 726-32*	69
Chapter 6	Management of the cN0 neck in early stage oral squamous cell cancer: A modeling study of the cost-effectiveness *Published: Oral Oncology 2013, Aug; 49 (8): 771-7*	85
Chapter 7	Quality of life after different procedures for regional control in oral cancer patients: Cross-sectional survey *Published: Clinical Otolaryngology 2015, July*	101
Chapter 8	Cost-effectiveness of selective neck dissection versus modified radical neck dissection for treating metastases in oral cavity cancer patients; a modeling study *Published: Head & Neck 2014, July*	113
Chapter 9	Decision modeling to inform evidence based personalized decisions: Using neck management in oral cavity cancer patients as an example cancer patients *Submitted*	129
Chapter 10	General discussion	147
Summary		160
Samenvatting		165
Bibliography		171
Publicaties		185
Curriculum Vitae		186
Dankwoord		187

Chapter 1

General introduction

Evidence-based surgery

Surgical innovations comprise new techniques, modified strategies, or innovative instruments. The evidence base for many of these innovations—and therefore for much of current surgical practice—is vastly weaker than for most modern drug treatments. It has been shown that only half of all new experimental treatments will prove to be better than established treatments when tested in randomized trials.[1] At the same time technological innovation is an important driver of the cost growth in healthcare.[2] Because new treatments are not always evaluated in trials or abandoned after negative outcomes a part of innovations increases healthcare costs without having added value for patients. The latter is especially true in surgery. Surgical trials are difficult and rare; some interventions have been widely adopted without rigorous evaluation.[3] Besides this, the introduction of new surgical procedures and devices is not yet regulated in most countries. Hence, these procedures and devices may become widely used with little evidence of their efficacy and cost-effectiveness. This is different for drugs, where both the Food and Drug Administration (United States) and the European Medicines Agency require new treatments to undergo rigorous premarket clinical trials.[4] The result of the current drug approval system, however, is a 8-15-year wait for new treatments with an estimated average cost to market between $500 million a nd $2 billion.[5] To come to more evidence based surgery in terms of effectiveness and efficiency, without jeopardizing innovation, we need creative and effective scientific approaches to reduce healthcare expenses, while still improving patient outcomes by innovation.

The IDEAL (Innovation, Development, Exploration, Assessment, Long-term study) collaboration developed a framework for the stages in surgical innovation including a set of recommendations on how evaluations should be conducted at each stage.[6] The IDEAL collaboration did not, however, take into account cost-effectiveness, norms and values of its social context, and its transdisciplinary aspects. Medical professionals, patients, health care insurers, manufacturers of medical devices, and policy makers would benefit from accurate information on costs and benefits. An efficient and evidence based health care system should enhance the ability of these stakeholders to make informed choices about the implementation and use of new procedures and devices. Therefore, there is a need to improve the scientific tools and methods to inform evidence based decisions in surgery.

Decision modeling

Decision-analytic modeling is a method, which synthesizes evidence to inform decisions. In a decision analytical model, two or more alternatives can be compared in terms of costs and/or effects. This analysis can be used to examine whether new procedures/ technologies are more effective, to which extent the use of new technologies influences the costs of a healthcare pathway and to determine whether the new technology represents value for money. Based on the available evidence combined in a model, a decision can be made whether a new technology should be implemented, if further evidence should be obtained or that the technology should be abandoned. Regardless of the availability of evidence, there will always be, to a greater or lesser extent, uncertainty regarding the costs and consequences of the technology. Value of information analyses can be performed to examine the value of acquiring additional information, and reducing uncertainty, through further research. These analyses show on which parameters further research is most valuable. With these analyses we can make informed decisions about new technologies, and about how to make efficient use of available resources for (further) research. Therewith, a decision analytical model can be an ideal starting point for clinical decision problems and the development of technologies. It can provide a fundament for further research and an evaluative framework for decision making.

From population to the individual patient

Evidence based treatment decisions based on decision models could improve outcome for a population of patients. However, a physician does not see an 'average patient' from a certain population. Optimal treatment could differ between individuals within a specified population, for instance a population that is specified within a randomized controlled trial. Research should help physicians making an optimal decision for their patients. Up to now, individual patient characteristics are often not taken into account when making evidence based decision models. Since real evidence based decision making has the care for individual patients as its top priority, it is crucial to bridge the gap between evidence based decision making and personalized healthcare.[7]

The case of evidence based decision making for neck management in oral cancer

In this thesis we study the value of several management strategies for patients with early oral squamous cell carcinoma (OSCC). This empirical example offers the unique opportunity to assess the opportunities and challenges of an evidence based approach for the evaluation of (surgical) innovation. OSCC is an important health issue in numerous countries. Because of important trade-offs underlying the surgical treatment of OSCC many decision problems exist in this field. An evidence based approach could therefore fulfill a need for stakeholders in this field.

Etiology

Each year, more than half a million patients worldwide are diagnosed with squamous-cell carcinoma (SCC) of the head and neck.[8] An important part of the SCCs in this region arise in the oral cavity. Tobacco use, alcohol consumption and betel quid chewing are the main risk factors in the etiology of intraoral cancer.[9,10] The incidence of OSCC is higher in men, which is mainly due to the greater exposure to the risk factors.[11] Men aged over 50 are, therefore the main risk group, although the ratio of males to females has decreased.[12]
Unfortunately, the epidemiological data concerning 'oral cancer' often includes both oral cavity and oropharyngeal cancers, which actually have different etiology, are diagnosed at different stages, and treated in different ways. Patients with oral cavity squamous cell carcinoma (OCSCC) generally present with early stage disease and the primary treatment is surgery with or without (chemo-radiation). Oropharyngeal cancers are mostly more advanced at the time of diagnosis and primary treatment for these patients is more likely to be (chemo-)radiation.[13] This thesis focus on early stage OCSCC.

Diagnosis of neck metastases

The most important factor in early stage OCSCC is spread of cancer to regional lymph nodes in the neck. Widely used preoperative neck staging techniques include palpation, ultrasound (US) with or without fine needle aspiration cytology (+/- FNAC), computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) scanning.

Surgical management of the neck

The presence of regional metastases decreases survival by approximately 50%.[14]
Therefore, adequate treatment of metastatic lymph nodes in the neck is considered crucial to improve oncologic outcomes for these patients.
Neck dissection is the standard surgical treatment for cancer in the regional lymph nodes of the neck. The purpose of a neck dissection is to remove those lymph nodes containing metastatic cancer.[15] Depending on the extend, neck dissections are generally classified into four categories:

1. Radical, consisting of the dissection of all the lymphatic tissue in level I-V including removal of the spinal accessory nerve, sternocleidomastoid muscle and internal jugular vein.
2. Modified radical, or any alteration of the radical procedure that involves preservation of one or more non-lymphatic structures of the neck.
3. Selective, including supramohyoid, posterolateral, lateral and anterior, each representing a specific procedure that preserves one or more lymph node groups routinely removed in radical neck dissection
4. Extended, or any alteration involving removal of additional lymph node groups or non-lymphatic structures relative to the radical procedure.[16]

In addition, neck dissections can be therapeutic, if there are metastases detected (cN_+) or elective or "prophylactic" if no metastases are detected (cN_0) during clinical diagnoses. To further clarify the nomenclature a more rational nomenclature has been proposed more recently.[17]

Neck dissection is invasive and shoulder dysfunction, cosmetic deformity, cutaneous paresthesia, and chronic neck and shoulder pain syndrome are well-recognized side effects of neck dissection that influence a patient's quality of life.[18-20]. Preserving structures (i.e. performing less radical types of dissections) result in less shoulder complaints and in a higher quality of life.[21-25]

<u>Trade-off</u>

The surgical management of regional lymph nodes includes a difficult trade-off between increasing prognoses by removing those lymph nodes involved by or at risk for involvement by metastatic cancer. On the other hand, quality of life could be reduced by shoulder complaints or complications as a result of removing those lymph nodes.

<u>Management decisions in cN_+</u>

The trade-off between survival prognoses and quality of life plays a role when metastases are detected. In this case the radical and modified radical neck dissection are considered as standard treatment. However, in many cases this results in overtreatment because metastases are not present in all levels of the neck. Some surgeons are therefore exploring the efficacy of selective neck dissection in cases of a clinically positive neck (cN_+) with limited disease, in an effort to reduce morbidity without reducing oncologic safety.[26,27]

<u>Management decisions in cN_0</u>

The trade-off between prognoses and quality of life is even more present for patients with a clinically uninvolved neck (cN_0). This is caused by the lack of sensitivity of current used diagnostic modalities, which results in a high incidence of occult, i.e. undetected, metastases. In patients with a cN_0 neck occult metastases have been reported in 20% up to 44% of the patients.[26]

For many years, two strategies have been used to treat the neck of these patients: elective neck dissection (END) or watchful waiting (WW). Proponents of END consider it to be the current gold-standard procedure for the cN_0 neck, as it not only provides valuable prognostic information regarding nodal status, but it is also therapeutic for patients who have occult metastases in the regional lymph nodes of the neck. A disease-survival advantage has been demonstrated for elective neck dissection in a meta-analysis comparing elective neck dissection versus observation in cN_0 patients.[26,28] Proponents of watchful waiting argue that most patients who develop metastases somewhere in the future, will be diagnosed early, if closely observed, and most can be salvaged successfully with a therapeutic neck dissection at time of relapse.[29,30] This treatment strategy, thereby, avoids the unnecessary morbidity of neck dissection in the majority of node-negative patients in whom the neck dissection was not needed. The trade-off between prognosis and quality of life could be influenced by new diagnostic modalities that could make a better distinction between patients with or without neck metastases.

Objectives/aims:

The overall aim of this thesis is to inform decisions in the management of OCSCC patients in an evidence based manner.
More specific aims are:

1. To evaluate the need for evidence based surgery in general, and more particularly regarding diagnostic and surgical innovations in OCSCC
2. To assess various management strategies for the neck in early stage OCSCC patients on their cost-effectiveness
3. To bridge the gap between evidence based decision making and personalized healthcare in the management of the neck in OCSCC patients

Thesis outline

Part 1 of this thesis starts with the costs of lower value surgery, which indicates the general need for evidence based surgery in monetary terms (*Chapter 2*). *Chapter 3* consists of a description of (inter)national practice variations in the management of the neck in OCSCC patients. Variation in practice could be an indicator of the absence of evidence based recommendations or guidelines. *Chapter 4* describes the results of an interactive evaluation in which we interviewed all relevant stakeholders regarding their needs of surgical innovations and research in head and neck cancer.

Part 2 is focused on evidence based management in OCSCC patients. Decision modeling was used to synthesize the relevant evidence regarding the various management options to provide estimates of effects and/or costs. *Chapter 5* comprises a diagnostic meta-analysis of the accuracy of a sentinel lymph node biopsy (SLNB) for the prediction of neck node involvement in OCSCC. The diagnostic accuracy of the SLNB was also an important parameter for a decision model of the management of the N_0 neck in OCSCC. In this model (*Chapter 6)*, five strategies for the management of the N_0 neck in early stage OCSCC are compared regarding their cost-effectiveness. Besides this value of information analysis for missing evidence was performed. This value of information analysis showed that more evidence regarding quality of life after different procedures was of high value. Therefore, we also studied the quality of life after different procedures (*Chapter 7)*. *Chapter 8* describes a decision model of the management of a clinical positive neck. This model evaluates survival, quality of life and costs associated with SND in patients with early stage OCSCC with singular nodal disease, compared to MRND.

Part 3 is about personalized decision making. Decision modeling provides results for an 'average patient in an a priori defined population. Physicians therefore still need to make subjective decisions for the individual patient. In *chapter 9* we explored the value of individualized care by constructing a personalized decision model for patients with a cN_0 neck.

In *chapter 10* we will discuss the overall implications of our evidence based approach for surgical treatment and innovations in general and for surgical management of the neck in OCSCC patients in particular.

Part I

Chapter 2

Lower value surgical procedures create a significant waste in healthcare: a modeling approach

Tim M Govers

Janneke PC Grutters

Ben VM Crul

Jacques van Limbeek

Gert P Westert

Maroeska M Rovers

Submitted

Abstract

Background: Several initiatives have focused on defining healthcare services that provide little or no benefit. No longer performing these lower value procedures might save considerable costs. In contrast to medicines, surgical procedures can be applied without robust studies, so particularly in surgery lower value procedures might be identified. It is important to translate lists of lower value services into meaningful metrics of avoidable costs to inform policy makers and providers how to bend the cost curve.

Methods and findings: We calculated avoidable costs associated with surgical procedures included in the 'do not do' recommendations of the National Institute for Health and Clinical Excellence, using both indication specific Dutch utilization data and more general utilization data from England, and actual costs from cost(-effectiveness) studies and reimbursement tariffs. We also extrapolated the results to all potential lower value surgical procedures, varying the assumption of lower value surgery to be between 5 and 33%.
The estimated costs that could be saved by not performing proven lower value surgery in the Netherlands were €11 million and €8 million per year using actual costs and reimbursement tariffs, respectively. An extrapolation of this result to all surgical procedures showed a potential cost saving that varied between €63 million and €419 million per year. Data from England showed cost savings of about 48.5 million using actual costs. Extrapolation of the data from England showed a potential cost saving between €382 million and €2.5 billion.

Conclusions: Lower value surgical procedures are creating a significant waste in healthcare. Evidence-based surgery could reduce this waste and thereby result in considerable savings.

2

Introduction

The current financial climate requires all healthcare systems to find ways of controlling costs without cutting quality of care. One way to achieve this is to no longer perform costly activities that provide little or no health benefit to patients. Several recent initiatives, including the Choosing Wisely campaign by the American Board of Internal Medicine Foundation and the 'do not do' recommendations of the National Institute for Health and Clinical Excellence (NICE), have focused on directly defining these (lower value) activities.[31,32] Although challenging, it is important to translate these evidence-based lists of lower value services into meaningful metrics of avoidable costs to inform policy makers and providers how to bend the cost curve.

It has been estimated that for 30-50% of all surgical procedures the effectiveness has not been studied.[33] This implies that the procedures that are proven to be of lower value according to recent initiatives are only part of all existing lower value activities.[31,32] Others have shown that about half of all new therapeutic interventions are not better than their best alternative.[1] In contrast to medicines, surgical procedures and devices can be applied without robust clinical studies. Hence, particularly in surgery many lower value procedures might be identified, and large potential cost savings may exist.

We used existing evidence-based lists to calculate the avoidable costs of lower value surgery.

Methods

Our approach consisted of five interrelated steps. First, we searched for existing evidence regarding lower value surgical procedures, where lower value was defined as having little or no health gain compared to their best alternative. Second, the yearly volumes of these procedures were obtained. Third, costs for the procedures and their alternatives were determined. Fourth, a calculation model was constructed to calculate avoidable costs by not performing these specific procedures. Fifth, we extrapolated our findings to estimate the total potential cost savings at the national level by not performing all lower value surgical procedures, including those for which no evidence exists, assuming that this is a fixed percentage of all surgical procedures.

Step 1: Selection of lower value surgical procedures

Several organizations within different countries started initiatives to identify and reduce the use of lower value care of which the Choosing Wisely campaign, an Australian initative and the NICE 'do not do' campaign are probably best known. [31,34,32]

We used NICE's 'do not do' recommendations to select surgical procedures that are deemed to deliver little or no health gain. NICE provides national guidance and advice to improve health and social care in England, but many of their results are adopted in other countries in Western Europe. Based on both available data and expert consultation, we selected those surgical procedures that are performed in the Netherlands.

Step 2: Volumes

Volumes of the selected procedures in step 1 were based on indication specific utilization data from the Dutch Hospital Data (DHD), an institute that manages the national data of all performed procedures in Dutch hospitals.[35] Some of the do not do recommendations were more specific than as presented in the DHD database. For these procedures, we assumed that 50% of the obtained volume actually related to the specific indication. In a sensitivity analysis we varied this percentage from 33% up to 66%. We also used more general utilization data from the Hospital Episode Statistics of the National Health Service (England).[36] Where the Dutch volume data could be linked to specific indications, this was not possible for the data from England. We therefore applied the Dutch indication-specific percentage of performed procedures to the utilization data from England. Ethical approval was not required, as this modeling study only used anonymous data from databases and previously performed studies.

Step 3: Costs

Unit costs of the procedures and their alternatives were derived using both actual costs and reimbursement tariffs. The alternative procedures were drawn from the 'do not do' recommendations when mentioned and otherwise from medical literature and expert opinions.

To obtain actual costs we used the reported cost-effectiveness studies in the NICE guidelines. If such studies were not available, we searched the literature for relevant cost-effectiveness studies. If possible, we took into account both direct cost (e.g. costs of the procedure) and indirect cost (e.g. complication costs and costs of retreatment). When no cost-effectiveness study was available, we searched for studies that reported on the costs of the (single) procedure, or the alternative. In such cost studies only direct healthcare cost are reported. If no data could be retrieved, we used the cost data of our own University medical centre. Costs in currencies other than Euros were converted to Euros by using the exchange rate of the year for which the costs were presented. Costs were inflated to the year 2012 with price index rates.[37] Furthermore, we used Dutch procedure and indication specific reimbursement tariffs from Achmea, which is with approximately 5 million insured individuals one of the largest health care insurance companies in the Netherlands.[38]

Step 4: Calculation Model

To calculate the potential cost savings of the selected surgical procedures, we multiplied the volumes by the differences in both actual costs and reimbursement tariffs of these surgical procedures and their alternatives. To enable comparisons between countries and health care systems, we additionally calculated avoidable costs per 1000 inhabitants. For this purpose we used a number of 16,730,348 inhabitants in the Netherlands and 53,500,000 in England in 2012.[39,40] The English volumes were only multiplied by differences in actual costs since reimbursement tariffs were only available for the Dutch situation.

Step 5: Extrapolation model

Since effectiveness data are still lacking for many surgical procedures, it is expected that the identified lower value procedures represent only a percentage of all lower value surgical procedures. Therefore, we also extrapolated our results assuming that 20% of all surgical procedures are of lower value. This number was used since the effectiveness has not been studied for about 40% of all interventions and about half of all (new) interventions are not more effective than their best alternatives.[33,1] In a sensitivity analysis we varied this percentage between 5% and 33%.

Results

Procedures

The ‘do not do’ recommendations of NICE[32] consist of 949 recommendations of which 52 were related to surgical procedures. Twenty-three of these procedures (44%) are not performed in the Netherlands. For the 29 remaining lower value surgical procedures, indication-specific and overall volumes were requested from Dutch Hospital Database (DHD). For four procedures no data could be obtained because of the absence of a specific code for the procedure (see appendix). Three procedures for Barret-Oesofagus, that were mentioned separately in NICE, were combined in the DHD. We therefore included 23 lower value surgical procedures in our further analyses (see also Table 1). Since indication specific utilization data were not available in England, we used the same 23 surgical procedures so that we could apply the Dutch indication specific percentages to the overall utilization data from the Hospital Episode Statistics of the NHS.[35]

Volumes

Indication-specific volumes and assumptions based on the Dutch data regarding the lower value surgical procedures are presented in Table 2. The frequency of all 23 lower value surgical procedures was 11,802 operations per year in the Netherlands. The two most often performed lower value procedures were facet denervation for non-specific low back pain and adenoidectomy for otitis media with 5,423 and 1,718 procedures yearly, respectively (Table 3). Table 3 provides an overview of the volumes of the 23 procedures in England. When we applied the Dutch indication-specific percentages per procedure to the data from England, the frequency of all 23 lower value surgical procedures was 29,196 operations per year.

Table 1: Overview of the included surgical procedures as reported in the NICE do not do recommendations

Surgical procedure	Diagnosis/indication	Recommendation in do not do
Hysterectomy	Heavy menstrual bleeding	Hysterectomy should not be used as a first-line treatment solely for heavy menstrual bleeding
Ovariectomy/ Oophorectomy1	Heavy menstrual bleeding	Removal of healthy ovaries at the time of hysterectomy should not be undertaken
Arthroscopy of the knee	Knee osteoarthritis	Referral for arthroscopic lavage and debridement should not be offered as part of treatment for osteoarthritis unless the person has knee osteoarthritis with a clear history of mechanical locking (not gelling, giving way or X-ray evidence of loose bodies
Adenoidectomy	Otitis media	Once a decision has been taken to offer surgical intervention for otitis media with effusion in children, insertion of ventilation tubes is recommended. Adjuvant adenoidectomy is not recommended in the absence of persistent and/or frequent upper respiration tract symptoms
Sentinel node biopsy	Breast cancer	Do not perform sentinel node biopsy routinely in patients with a preoperative diagnosis of ductal carcinoma in situ who are having breast conserving surgery, unless they are considered to be at high risk of invasive disease.
Axillary lymph node dissection	Breast cancer	Do not offer further axillary treatment to patients fount do have only isolated tumour cells in their sentinel lymph nodes. These patients should be regarded as lymph-node negative
Ovariectomy/ Oophorectomy2	Breast cancer	Do not offer adjuvant ovarian ablation/suppression to premenopausal women with estrogen receptor (ER)-positive early invasive breast cancer who are being treated with tamoxifen and, if indicated chemotherapy
Glaucoma surgery1	Glaucoma	Treatment should not be routinely offered to people over the age threshold which varies according to intraocular pressure and central corneal thickness (Above 80 no treatment is recommended)
Glaucoma surgery2	Glaucoma	Do not treat people with suspected chronic open angle glaucoma and normal intraocular pressure
Facet denervation	Low back pain (non-specific)	Do not refer people for radiofrequency facet denervation
Minimal invasive prostate surgery	Benign prostate enlargement	If offering surgery for managing voiding lower urinary tract symptoms presumed secondary to benign prostate enlargement do not offer minimally invasive treatments as an alternative to transurethral resection of the prostate, transurethral vaporization of the prostate or holium laser enucleation of the prostate
Ablative therapy	Barretts Oesophagus	Do not use argon plasma coagulation, laser ablation or multipolar electrocoagulation alone, or in combination with each other.
Fundoplication	Gastro-Oesophageal reflux disease	Surgery cannot be recommended for the routine management of persistent gastro-oesophageal reflux disease
Retroperitoneal lymphadenectomy	Ovarian cancer	Do not include systematic retroperitoneal lymhadenectomy (block dissection of lymph nodes from the pelvic side walls to the level of the renal veins) as part of standard surgical treatment in women with suspected ovarian cancer whose disease appears to be confined to the ovaries (that is, who appear to have stage I disease)

Caesarean section1	Twin pregnancy	In otherwise uncomplicated twin pregnancies at term where the presentation of the first twin cephalic, perinatal morbidity and mortality is increased for the second twin. However, the effect of planned Caesarian section in improving outcome for the second twin remains uncertain and therefore caesarian section should not routinely be offered outside a research context
Caesarean section2	Small for gestational age	The risk of neonatal morbidity is higher with 'small for gestational age' babies. However, the effect of planned caesarian section in improving these outcomes remains uncertain and therefore caesarian section should not routinely be offered outside a research context
Caesarean section3	HIV	Do not offer a caesarian section on the grounds of HIV status to prevent mother-to child transmission of HIV to a women on highly active anti-retroviral with a viral load of less than 400 copies per ml or a women on any anti-retroviral therapy with a viral load of less than 50 copies per ml. Inform women that in these circumstances the risk of HIV transmission is the same for a caesarian section and a vaginal birth.
Caesarean section4	Obesitas	Do not use a body mass index (BMI) of over 50 alone as an indication for planned caesarian section
Stent placement	Lower Limb peripheral arterial disease	Do not offer primary stent placement for treating people with intermittent claudication or critical limb ischemia caused by aorto-iliac disease (except complete occlusion) or femeropopliteal disease
Varicoceles surgery	Infertility	Men should not be offered surgery for varicoceles as a form of fertility treatment because it does not improve pregnancy
Diagnostic laparoscopy	Tubal Occlusion	Women who are not known to have comorbidities (such as pelvic inflammatory disease, previous ectopic pregnancy or endometriosis) should be offered hysterosalpingography to screen for tubal occlusion because this is a reliable test for ruling out tubal occlusion and it is less invasive and makes more efficient use of resources than laparoscopy
Thoracic sympathectomy	Hyperhidrosis or blushing	Do not offer endoscopic thoracic sympathectomy to treat hyperhidrosis or facial blushing in people with social anxiety disorder. This is because there is no good-quality evidence showing benefit from endoscopic thoracic sympathectomy in the treatment of social anxiety disorder and it may be harmful
Mechanical thrombus extraction	ST elevated myocardial infarction	Do not routinely use mechanical thrombus extraction during primary PCI for people with acute STEMI

Table 2: Dutch utilization volumes (obtained from the DHD database) of the surgical procedures reported in the NICE do not do recommendations

Surgical Procedure	Corresponding Diagnosis DHD database	Volume in DHD database
Hysterectomy	Heavy menstrual bleeding	488*
Ovariectomy/ Oophorectomy1	Heavy menstrual bleeding	51
Arthroscopy of the knee	Osteo-artrose/arthritis of the knee	964*
Adenoidectomy	Otitis media	3,435*
Sentinel node biopsy	Carcinoma in situ mamma	579
Axillary lymph node dissection	Malignant neoplasm mamma or axilla	823*
Ovariectomy/ Oophorectomy2	Malignant neoplasm mamma	16*
Glaucoma surgery1	Open angle glaucoma (>80 years)	61
Glaucoma surgery2	Open angle glaucoma	627*
Facet denervation (lumbar region)	Lumbago or back pain (non-specific)	5,423
Cryo-surgery	Prostatic hyperplasia	11
Ablative therapy	Benign neoplasm of oesophagus	4
Fundoplication	Diaphragmatic hernia	286*
Retroperitoneal lymphadenectomy	Malignant neoplasm Ovary	8*
Caesarean section1	Twin pregnancy	525*
Caesarean section2	Poor growth	570
Caesarean section3	Aids	4
Caesarean section4	Obesity	3
Stent placement	Peripheral arterial disease	2,189*
Varicoceles surgery	Scrotal varices	187*
Diagnostic laparoscopy	Tubal occlusion	144*
Thoracic sympathectomy	Hyperhidrosis	82*
Mechanical thrombus extraction	Acute myocardial infartcion	417*

* For these procedures the assumption was made that 50% of the obtained volume was actually performed for the specific indication reported in the NICE do not do recommendations. In sensitivity analyses these percentage were varied from 33% up to 66%.

Table 3: Utilization volumes from both the Netherlands and England.

Surgical procedure	Total volume Netherlands	Indication	Volume for indication Netherlands (% of total)	Total volume England	Volume for indication England using Dutch % of total volume
Hysterectomy	7,484	Heavy menstrual bleeding	244 (3.3)	46,253	1,508
Ovariectomy/ Oophorectomy1	6,222	Heavy menstrual bleeding	51 (0.8)	44,757	367
Arthroscopy of the knee	24,260	Knee osteoarthritis	482 (2)	92,791	1,844
Adenoidectomy	14,946	Otitis media	1,718 (11.5)	20,296	2,333
Sentinel node biopsy	6,399	Breast cancer	579 (9.0)	25,499	2,307
Axillary lymph node dissection	984	Breast cancer	413 (42.0)	11,841	4,970
Ovariectomy/ Oophorectomy2	6,766	Breast cancer	8 (0.1)	48,021	57
Glaucoma surgery (1 and 2 together)	1,994	Glaucoma	375 (18.8)	7,463	1,404
Facet denervation	12,333	Low back pain (non-specific)	5423 (44.0)	7,599	3,341
Minimal invasive prostate surgery	50	Benign prostate enlargement	11 (22.0)	28	6
Ablative therapy	427	Barretts Oesophagus	4 (0.9)	117	1
fundoplication	543	Gastro-Oesophageal reflux disease	143 (26.3)	3,862	1,017
Retroperitoneal lymphadenectomy	565	Ovarian cancer	4 (0.7)	562	4
Caesarean section (1,2,3 and 4 together)	19,025	Twin pregnancy, poor growth, aids, obesity	840 (4.4)	168,575	7,443
Stent placement	3,257	Lower Limb peripheral arterial disease	1095 (33.6)	5,875	1,975
Varicoceles surgery	263	Infertility	94 (35.7)	1,484	530
Diagnostic laparoscopy	7,503	Tubal Occlusion	72 (1.0)	1,276	12
Thoracic sympathectomy	109	Hyperhidrosis or blushing	41 (37.6)	204	77
Mechanical thrombus extraction	1,021	ST elevated myocardial infarction	209 (20.4)	0	0

Costs

Table 4 provides an overview of the actual costs and reimbursement tariffs of the different procedures and their alternatives. For 13 procedures (57%) the alternative was either watchful waiting or doing nothing, whereas for the other 10 procedures (43%) an alternative less invasive procedure (e.g. medicines) was reported as the best alternative. For 15 of the 23 procedures (65%) both direct and indirect costs were obtained; for the other 8 (35%) only direct cost were reported.

Calculation model

Table 5 shows the avoidable costs for each lower value surgical procedure using both actual costs and reimbursement tariffs for the Dutch situation. The two procedures that are responsible for the highest potential cost saving when using actual costs were arthroscopic lavage and debridement for osteoarthritis of the knee, and facet denervation for non-specific low back pain, with potential cost savings of about €2.3 million and €2.0 million, respectively. When using reimbursement tariffs, facet denervation for non-specific low back pain and adenoidectomy as an adjuvant treatment for otitis media were responsible for the highest potential cost saving with €3.5 million and €1.1 million respectively.

The total potential cost savings associated with lower value surgical procedures were €10.6 million and €8.2 million per year using cost prices and reimbursement tariffs, respectively. Taking into account the total number of inhabitants, the avoidable costs per 1,000 inhabitants were €633 and €489 using actual costs and reimbursement tariffs, respectively. The sensitivity analyses (Table 6) showed that using a variation of 33% to 66% regarding the volume indication, the potential avoidable costs varied from about €8.4 million to €12.6 million (€504 to €754 per 1000 inhabitants) per year using actual costs and from about €7.0 to €9.3 million (€419 to €554 per 1000 inhabitants) using reimbursement tariffs.

Table 7 shows the avoidable costs for each lower value surgical procedure using the utilization data from England together with the Dutch indication specific percentages per procedure. Axillary lymph node dissection in breast cancer patients and arthroscopic lavage and debridement for osteoarthritis of the knee were the two procedures with the highest potential cost savings, with potential cost savings of €17.8 and €8.9 million respectively. The total potential cost savings were about €48 million. Taking into account the number of inhabitants, the avoidable costs were €906 per 1000 inhabitants in England.

Table 4: Costs prices and reimbursement tariffs of surgical procedures and their alternatives

Procedure	Alternative	Actual costs Procedure (€)	Actual costs alternative (€)	Type of costs (time interval)	Source	Reimbursement tariff procedure (€)	Reimbursement tariff alternative (€)
Hysterectomy	Endometrial ablation	2,864	2,647	Direct and indirect (4 years)	(41)	2,751	750
Ovariectomy/ Oophorectomy1	Do nothing	3,609	0	Direct (costs for diagnosis mamma carcinoma)	(42)	2,541	0
Arthroscopy of the knee	Do nothing	4,804	0	Direct	(43)	1,171	0
Adenoidectomy	Do nothing (only ventilation tubes)	2,223	1,983	Direct and indirect (unclear)	NICE guideline	653	0
Sentinel node biopsy	Watchful waiting	2,702	0	Direct	(44)	2,444	0
Axillary lymph node dissection	Watchful waiting	3,583	0	Direct	(44)	4,140	4087
Ovariectomy/ Oophorectomy1	Watchful waiting	3,609	0	Direct	(42)	2,541	0
Glaucoma surgery1	Do nothing	11,895	10,252	Direct and indirect (life-time)	NICE guideline	980	0
Glaucoma surgery2	Do nothing	11,895	10,252	Direct and indirect (life-time)	NICE guideline	980	0
Facet denervation (lumbar region)	Do nothing	362	0	Direct	Cost price Radboud University medical centre	1,228	0
Cryo-surgery	Transurethral resection of the prostate	5,622	4,483	Direct and indirect (2 years, costs for transurethral needle ablation)	(45)	1,590	1,590
Ablative therapy	Watchful waiting	24,047	22,233	Direct and indirect (50 years)	NICE guideline	957	358
Fundoplication	Pharmaceutical treatment (proton-pump inhibitors)	1,920*	0*	Direct and indirect (5 years)	(46)	3,067	358
Retroperitoneal lymphadenectomy	Watchful waiting	10,709	0	Direct (costs for diagnosis testis cancer)	(47)	4,316	4,316
Caesarean section1	Vaginal delivery	2,792	2,055	Direct and indirect (unclear)	NICE guideline	2,500	1,287
Caesarean section2	Vaginal delivery	2,792	2,055	Direct and indirect (unclear)	NICE guideline	2,500	1,287
Caesarean section3	Vaginal delivery	2,792	2,055	Direct and indirect (unclear)	NICE guideline	2,500	1,287
Caesarean section4	Vaginal delivery	2,792	2,055	Direct and indirect (unclear)	NICE guideline	2,500	1,287
Stent placement	No stent placement during percutaneous transluminal angioplasty	18,371	17,890	Direct and indirect (unclear)	(48)	6,068	6,068
Varicoceles surgery	Do nothing	2,031	0	Direct and indirect	(49)	1,098	0
Diagnostic laparoscopy	Hysterosalpingography	6,570	5,767	Direct and indirect	(50)	1,233	273
Thoracic sympathectomy	Do nothing	1,160	0	Direct	(51)	906	0
Mechanical thrombus extraction	No mechanical thrombus extraction during primary percutaneous coronary intervention	968 (is difference of procedure without mechanical thrombus extraction)	0	Direct	NICE guideline	1,404	1,404

* is difference with proton-pump inhibitors

Table 5: Calculation model: Cost differences using obtained utilization volumes from the Netherlands and actual cost prices and reimbursement tariffs

Procedure	Volume used for indication do not do recommendations	Difference in actual costs procedure and alternative (€)	Total cost difference using actual costs(€)	Difference in reimbursement tariffs procedure and alternative	Total cost difference using reimbursement tariffs (€)
Hysterectomy	244	217	52,948	2,001	488,244
Ovariectomy/ Oophorectomy1	51	3,609	184,059	2,541	129,591
Arthroscopy of the knee	482	4,804	2,315,528	1,171	564,422
Adenoidectomy	1,718	240	412,200	653	1,121,528
Sentinel node biopsy	579	2,702	1,564,458	653	378,087
Axillary lymph node dissection	412	3,583	1,474,405	53	21,810
Ovariectomy/ Oophorectomy2	8	3,609	28,872	2,541	20,328
Glaucoma surgery1	61	1,643	10,023	980	59,780
Glaucoma surgery2	314	1,643	515,081	980	307,230
Facet denervation (lumbar region)	5,423	362	1,963,126	639	3,465,297
Cryo-surgery	11	1,139	12,529	0	0
Ablative therapy	4	1,814	7,256	599	2,396
Fundoplication	143	1,920	274,560	2,709	387,387
Retroperitoneal lymhadenectomy	4	10,709	42,836	0	0
Caesarean section1	263	737	193,463	1,213	318,413
Caesarean section2	570	737	420,090	1,213	691,410
Caesarean section3	4	737	2,948	1,213	4,852
Caesarean section4	3	737	2,211	1,213	3,639
Stent placement	1,095	481	526,455	0	0
Varicoceles surgery	94	2,031	189,899	1,098	102,663
Diagnostic laparoscopy	72	803	57,816	960	69,120
Thoracic sympathectomy	41	1,160	47,560	906	37,146
Mechanical thrombus extraction	209	968	201,828	0	0
Total	**11,802**	**x**	**10,590,349**	**x**	**8,173,342**

Table 6: Sensitivity analysis calculation model

Calculation model assumptions	Percentage of obtained volume included in analysis*	Total actual cost savings (€)	Cost savings per 1000 inhabitants using actual costs (€)	Total reimbursement savings(€)	Cost savings per 1000 inhabitants using reimbursement tariffs (€)
Base assumptions	50%	10,590,349	633	8,173,342	489
Low assumptions	33%	8,436,976	504	7,004,323	419
High assumptions	66%	12,617,052	754	9,273,594	554

* For the procedures for which specific volume data was not available because the reported indication in the 'do not do' recommendation was narrower than as presented in the DHD data

Table 7: Sub analysis: Cost differences using utilization volumes from England and actual cost prices

Surgical procedure	Total volume NHS (England)	Volume for indication England	Difference in actual costs procedure and alternative (€)	Total cost difference using actual costs(€)
Hysterectomy	46,253	1,508	217	327,232
Ovariectomy/ Oophorectomy1	44,757	367	3,609	1,324,000
Arthroscopy of the knee	92,791	1,844	4,804	8,856,561
Adenoidectomy	20,296	2,333	240	559,912
Sentinel node biopsy	25,499	2,307	2,702	6,234,117
Axillary lymph node dissection	11,841	4,970	3,583	17,806,975
Ovariectomy/ Oophorectomy2	48,021	57	3,609	204,916
Glaucoma surgery (1 and 2 together)	7,463	1,404	1,643	2,305,988
Facet denervation	7,599	3,341	362	1,209,584
Minimal invasive prostate surgery	28	6	1,139	7,016
Ablative therapy	117	1	1,814	1,988
fundoplication	3,862	1,017	1,920	1,952,764
Retroperitoneal lymphadenectomy	562	4	10,709	42,609
Caesarean section (1,2,3 and 4 together)	168,575	7,443	737	5,485,488
Stent placement	5,875	1,975	481	950,056
Varicoceles surgery	1,484	530	2,031	1,077,249
Diagnostic laparoscopy	1,276	12	803	9,832
Thoracic sympathectomy	204	77	1,160	89,011
Mechanical thrombus extraction	0	0	968	0
Total	**486,503**	**29,196**	**x**	**48,445,298**

Extrapolation model

Since 1,414,558 surgical procedures are performed per year in the Netherlands, the 11,802 identified lower value surgical procedures represent 0.83% of all surgical procedures.[52] Using the assumption that 20% of all procedures are of lower value, the potential costs (using actual costs) that could be saved amount to approximately €254 million per year, i.e. €15,175 per 1000 inhabitants. A sensitivity analyses using a range of 5% up to 33% lower value surgery suggests that the potential avoidable costs range between €63 million and €419 million per year, which means €3,794 to €25,038 per 1000 inhabitants (Table 8). In England about 4,600,000 surgical procedures are performed yearly. The 29,196 identified lower value surgical procedures represent 0.63% of all surgical procedures. Assuming that 20% of all procedures are of lower value, the potential costs savings (using actual costs) that could be saved amount to about € 1.5 billion per year, and € 28,534 per 1000 inhabitants. Sensitivity analyses assuming a range between 5 and 33% lower value procedures, showed a potential cost saving varying between €382 million and €2.5 billion, i.e. €7,133 and €47,081 per 1000 inhabitants (Table 9).

Table 8: Sensitivity analysis extrapolation model for the Netherlands: extrapolating results of calculation model using actual costs

Extrapolation Model	Assumption (% of all surgical procedures that will be of lower value)	Potential cost savings (€)	Potential cost savings per 1,000 inhabitants (€)
Base assumption	20	253,877,256	15,175
Low assumption	5	63,469,314	3,794
High assumption	33	418,897,473	25,038

Table 9: Subanalysis: Sensitivity analysis extrapolation model for England: extrapolating results of calculation model using actual costs

Extrapolation Model	Assumption (% of all surgical procedures that will be of lower value)	Potential cost savings (€)	Potential cost savings per 1,000 inhabitants (€)
Base assumption	20	1,526,567,685	28,534
Low assumption	5	381,641,921	7,133
High assumption	33	2,518,836,680	47,081

Discussion

Our results show that about half (23 out of 52) of the proven lower value surgical procedures are not performed anymore, at least not in the Netherlands. On the other hand, our results also show that still about €60 million per year can be saved if the other proven lower value surgical procedures would not be performed anymore, in England and the Netherlands alone. If we extrapolate this to all potential lower value surgical procedures, conservatively assuming that the percentage of lower value surgery will be between 5% and 33%, the potential avoidable costs in the Netherlands and England alone vary between €445 million and €2.9 billion per year.

These findings are consistent with the notion that avoidable costs are evident in healthcare and that overtreatment is one of the major contributors.(54) A claim-based study, focused on the US Medicare population, also showed a potential cost saving of billions using another subset of interventions.[55]

The major strength of this study is the combination of the NICE do not do recommendations with indication specific volumes and costs (both using cost-effectiveness studies and reimbursement tariffs). This provided the unique opportunity to calculate evidence-based potential cost savings of lower value surgical procedures.

Some potential limitations should also be discussed. First, we realize that in practice there will always be (subsets of) patients for whom a specific procedure does provide health benefit. However, this does not mean that the use of the procedure should be continued in all patients. Instead, routine use should be stopped, and future research should be aimed at identifying those patients who benefit from the procedure, for example by means of individual patient data meta-analyses and stratified decision analysis.[56,57] Second, in England no indication specific utilization data were available. We therefore had to assume that the indication specific percentages in England were comparable to those in the Netherlands, whereas the overall utilization data of both countries show large variations in volume (taking the number of inhabitants into account). For example, the number of caesarean sections was much higher in England. However, in the best possible way our results provide an estimation of the potential cost savings in England, instead of providing exact numbers. Third, due to the lack of indication specific utilization data in England, we also had to exclude the 23 procedures that were not performed in the Netherlands from all further analyses (see Appendix). However, since NICE provides national guidance and advice to improve healthcare in England, it is expected that at least part of these 23 procedures will still be performed in England. As a consequence, our estimations are probably an underestimation of the potential cost savings of proven lower value procedures in England. Fourth, the indication-based volume data from Dutch Hospital Data used in the main analyses did not always exactly match the do not do recommendations, i.e. sometimes the do not do recommendations were more specific than could be obtained from Dutch hospital data. To match the more detailed indications from the do not do recommendations we included only 50% of the obtained volumes in 14 procedures. This has probably resulted in an underestimation of the potential cost savings. We provided a range of potential cost savings by performing a sensitivity analysis in which we varied this percentage (using 33 and 66%). Fifth, for 15 procedures we could use data from cost-effectiveness studies, whereas for the other eight procedures only the initial costs of treatment could be taken into account, which might again have led to an underestimation of the potential cost savings. For two procedures, we had to make an estimation of costs based on the same surgical procedure for a different indication.

Despite these limitations the present study provides important insights in the extent of avoidable costs in surgical care. Considerable resources could be saved by not performing surgical care that is known to be of lower value for patients. Ultimately, this could lead to lower healthcare spending by society without jeopardizing patients' health. The extrapolation model suggests that withholding lower value surgical procedures may play an important role in sustaining an affordable healthcare system. The most relevant question therefore is how to withhold these lower value surgical procedures since it is known that with hindsight judging procedures undesirable is difficult.[31,58] Particularly in the surgical field, high quality research is often lacking, which is also reflected in the low percentage (i.e. 6%) of surgical procedures in the do not do recommendations.[59] Furthermore, abandoning procedures at a late stage is inefficient, given the considerable investments that have been made. The (early) assessment of the potential value of a surgical procedure as suggested by the IDEAL collaboration to steer its development and prevent ineffective surgical procedures from entering the market is needed and should be complemented by a system of governance to make decisions about whether to continue the development and use of new procedures and under what conditions.[4,6,60]

In conclusion, lower value surgical procedures are creating a significant waste in healthcare systems. Even our conservative estimations show a waste that amounts to billions of Euros in the Netherlands and England alone. Evidence-based surgery could reduce this waste and thereby result in considerable savings.

Acknowledgements

We thank Wieteke Nieuwboer for her help in obtaining the reimbursement data from the Achmea Health database.

Appendix: Overview of the lower value surgical procedures (as reported in the NICE do not do recommendations) that were no longer performed in the Netherlands and therefore excluded from the analyses

Surgical procedure	Diagnosis/indication	Recommendation in do not do
Dilatation and curettage	Heavy menstrual bleeding	Dilatation and curettage should not be used as a therapeutic treatment
Surgical management for vesicoureteric reflux	vesicoureteric reflux in children	Surgical management of vesicoureteric reflux (VUR) is not routinely recommended
HIFU and cryotherapy	Prostate cancer	High-intensity focused ultrasound and cryotherapy are not recommended for men with locally advanced prostate cancer other than in the context of controlled clinical trials comparing their use with established interventions
Surgical prevention of metastatic spinal cord compression	Metastatic spinal cord compression	Patients with spinal metastases without pain or instability should not be offered surgery with the intention of preventing metastatic spinal cord compression expect as part of a randomized controlled trial
Posterior decompression	Metastatic spinal cord compression	Posterior decompression alone should not be performed in patients with metastatic spinal cord compression except in rare circumstances of isolated epidural tumor or neural arch metastases without bony instability
En bloc excisional surgery	Metastatic spinal cord compression	En bloc excisional surgery with the objective of curing the cancer should not be attempted, except in very rare circumstances (for example, confirmed solitary renal or thyroid metastasis following complete staging)
Intradiscal electrothermal therapy *	Low back pain	Do not refer people for intradiscal electrothermal therapy
Percutaneous intradiscal radiofrequency thermocoagulation *	Low back pain	Do not refer people for percutaneous intradiscal radiofrequency thermo coagulation
Transurethral microwave thermotherapy	Benign prostate enlargement	If offering surgery for managing voiding lower lower urinary tract symptoms presumed secondary to benign prostate enlargement, do not offer minimally invasive treatments
High-intensity focused ultrasound	Benign prostate enlargement	If offering surgery for managing voiding lower lower urinary tract symptoms presumed secondary to benign prostate enlargement, do not offer minimally invasive treatments
Transurethral ethanol ablation *	Benign prostate enlargement	If offering surgery for managing voiding lower lower urinary tract symptoms presumed secondary to benign prostate enlargement, do not offer minimally invasive treatments
Laser coagulation*	Benign prostate enlargement	If offering surgery for managing voiding lower lower urinary tract symptoms presumed secondary to benign prostate enlargement, do not offer minimally invasive treatments
Scleral expansion surgery	Presbyopia	It is recommended that sclera expansion surgery presbyopia should not be used
Corneal implants	Refractive error	Corneal implants should not be used for the treatment of refractive error in the absence of other ocular pathology such as keratoconus
Laparoscopic uterine nerve ablation	Chronic pelvic pain	Laparoscopic uterine nerve ablation for chronic pelvic pain should not be used
Transmyocardial laser revascularization	Refractory angina pectoris	Transmyocardial laser revascularization for refractory angina pectoris should not be used
Percutaneous laser revascularization	Refractory angina pectoris	Percutaneous laser revascularization for refractory angina pectoris should not be used

Coronary revascularization	Systolic left ventricular impairment	Coronary revascularization should not be routinely considered in patients with heart failure due to systolic left ventricular impairment, unless they have refractory angina
Anterior spinal fusion	Tuberculosis	In patients with spinal tuberculosis, anterior spinal fusion should not be performed routinely
Caesarean section	Hepatitis B	Mother-to-child transmission of hepatitis B can be reduced if the baby receives immunoglobulin and vaccination. In these situations pregnant women with hepatitis B should not be offered a planned Caesarian section because there is insufficient evidence that this reduces mother-to-child transmission of hepatitis B virus.
Caesarean section	Hepatitis C	Women who are infected with hepatitis C should not be offered a planned Caesarian section because this does not reduce mother-to-child transmission of the virus.
Caesarean section	Planned	The risk of respiratory morbidity is increased in babies born by Caesarian section before labor, but this risk decreases significantly after 39 weeks. Therefore planned Caesarian section should not routinely be carried out before 39 weeks
Metallic stent placement	Colorectal cancer	Do not place self-expanding metallic stents in low rectal lesions or to relieve right-sided colonic obstructions or if there is clinical or radiological evidence of colonic perforation or peritonitis.
Risk reducing surgery	Familial breast cancer	Do not offer risk-reducing surgery (for people with personal history of breast cancer) to people with comorbidities that would considerably increase the risk of surgery or who have limited life expectancy from their cancer or other conditions
Percutaneous posterior tibial nerve stimulation	Urinary incontinence in women	Do not offer percutaneous posterior tibial nerve stimulation for overactive bladder syndrome unless: there has been a multidisciplinary team review; and conservative management including drug treatment had not worked adequately; and the woman does not want botulinum toxin A7 or percutaneous sacral nerve stimulation
Laparoscopic colposuspension	Urinary incontinence in women	Do not offer laparoscopic colposuspension as a routine procedure for the treatment of stress urinary incontinence in women. Only an experienced laparoscopic surgeon working in an multidisciplinary team with expertise in the assessment of urinary incontinence should perform the procedure

**No specific code in Dutch hospital data*

Chapter 3

An international comparison of the management of the neck in early oral squamous cell carcinoma in the Netherlands, UK, and USA

Tim M Govers

Tijs B H de Kort

Matthias A W Merkx

Stefan Steens

Maroeska M Rovers

Remco de Bree

Robert P Takes

J Craniomaxillofac Surg. 2016 Jan;44(1):62-9

Abstract

Background: Early oral cavity squamous cell carcinoma (OCSCC) management appears to vary both within and between countries. Variation in practice can be an indicator of absence of evidence-based management and may negatively influence survival and morbidity of OCSCC patients not receiving optimal treatment. The exact variation in practice and the relationship of the variation to differences in guidelines are unknown. This study aimed to report on these international variations in the Netherlands, UK, and USA, and to evaluate them.

Methods: Information regarding the international variation in OCSCC management strategies was obtained from a questionnaire sent to representatives of head and neck cancer centers in the Netherlands, UK, and USA. Within-country and between-country variations were also assessed in light of the different guidelines and health systems.

Results: In total, representatives of 45 HNC centers completed the questionnaire; 10 from the Netherlands, 26 from the UK and 9 from the USA. Our results demonstrate a distinct variation in the management strategy of OCSCC, both within and between countries. This variation was present in diagnostics, treatment and follow up. Only a small amount of variation between countries could be linked to differences in guidelines.

Conclusions: In conclusion there is high variation in the management of the neck in OCSCC. There seem to be a need for direct evidence about optimal management decisions to establish more evidence-based management and more uniform practice patterns.

Introduction

Cancers of the oral cavity are an important public health issue in many countries.[61] The vast majority of these cancers are oral cavity squamous cell carcinomas (OCSCC).[11] Regional lymph node metastases occur frequently and can decrease survival drastically compared with no regional lymph node involvement.[14,62] Therefore, adequate diagnostic assessment and treatment of the neck is considered to be crucial. However, the head and neck region is anatomically complex, and treatments can have a high impact on quality of life.[18,19,23] Therewith, decisions regarding both the assessment and treatment of the neck have been the subject of debate for many years. Multiple diagnostic modalities are available for detecting lymph node metastases, ranging from palpation to imaging techniques, such as computerized tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), and ultrasound-guided fine-needle aspiration cytology (USgFNAC). Even when no metastases are demonstrated by these diagnostic modalities (i.e. clinically negative node [cN_0]), a high rate of occult metastases remains. Traditionally, treatment of the cN_0 neck consists of an elective dissection of the neck or a watchful waiting (WW) policy. More recently, the sentinel node biopsy (SNB) procedure has been added to the diagnostic armamentarium. With this procedure, the metastatic status of the neck can be more accurately assessed.[63-65]

Until recently, (modified) radical neck dissection was considered to be the appropriate surgical treatment if lymph node metastases were detected. After selective neck dissection became the standard type of dissection for the cN_0 neck, the efficacy of selective neck dissections in the clinically positive neck (cN_+) has been explored as well.[26] The use and duration of regular follow-up of OCSCC patients after neck surgery and WW have also been questioned over the past few decades.[66]

Despite the amount of research published in the literature, the management of early OCSCC appears to vary both within and between countries; even between countries with a similar level of development. Variation in practice can be an indicator of absence of evidence-based management and may negatively influence survival and morbidity in those OCSCC patient populations that do not receive optimal treatment. Moreover, the variation in practice may also hamper the execution of international multicenter trials.

However, the exact variation in practice is unknown. Furthermore, it is not known whether some variation could be explained by differences in organization of healthcare between countries or by differences in guidelines.

The aim of the current study was to report on the international variations in the management of early (T_{1-2}) OCSCC in the Netherlands, UK, and USA. Furthermore, we will report on differences in guidelines and healthcare systems.

3

Materials and Methods

Design

This was a cross-sectional study using a convenience sample of centers treating OCSCC and an electronic online questionnaire (survey monkey).[67]

Questionnaire development

The questionnaire was based on both the literature and interviews with several specialists in the field of head and neck cancer (HNC), including two otolaryngologist/head and neck surgeons, three oral and maxillofacial head and neck surgeons, a radiologist specialized in HNC, and an oral cancer researcher. We piloted the questionnaire using five surgeons, and made some final adjustments based on their comments and suggestions before general distribution.

Study group

To obtain information regarding the international variation in the management strategies for early OCSCC, questionnaires were sent to representatives of HNC centers in the Netherlands, UK, and USA. These countries were chosen because of their assumed high quality of care, but very different healthcare systems. In the Netherlands, the questionnaires were distributed in cooperation with the Dutch Head and Neck Society, representing all eight HNC centers and their partner centers. Questionnaires were sent by e-mail to the members of the eight centers. These members were asked to distribute the questionnaire to possible partner centers. In the UK, the questionnaires were distributed in cooperation with the British Association of Head & Neck Oncologists (BAHNO) a multi-disciplinary society for healthcare professionals involved in HNC. The link to the questionnaires was distributed by e-mail and on the website of the BAHNO. In the USA, HNC specialists involved in treatment and research in the field of oral cancer were directly contacted by e-mail in which a link to the questionnaire was attached. The centers were assured that the results were anonymous.

Questionnaire content

The questionnaire was divided into the following two parts: questions on diagnostics, management of the cN_0 neck, management of the cN_+ neck, and standard follow-up strategy; and choice of a preferred strategy for two hypothetical cases of OCSCC with cN_0 necks in two different scenarios (see Box 1 for a description of the cases).[1,2] Participants were ask to choose a treatment strategy for each patient according to two different scenarios, consisting of different primary tumor characteristics. In scenario one, the largest surface tumor diameter is 1 cm and the depth of the invasion is 4 mm. In scenario two, the largest surface tumor diameter is 2 cm and the depth of the invasion is 7 mm.

Box 1: Description of hypothetical patient examples

Patient I

Man, 75 years of age with squamous cell carcinoma of the tongue and a clinically negative neck. This man has been smoking 20 cigarettes and drinking 2 units of alcohol on average every day for the past 50 years. He has multiple co-morbidities and has been living in a retirement home for 5 years. His quality of life is relatively low at present. He has some problems performing his daily activities, has pain (in his back and shoulder), and is quite anxious after receiving the diagnosis of oral cavity cancer. He also has some shoulder complaints. He is unable to lift objects above his head because of these complaints. However, the shoulder complaints are not always present.

Patient II

Woman, 55 years of age with a clinically negative squamous cell carcinoma of the tongue. This woman does not smoke, but drinks a few units of alcohol on average every day. She has no co-morbidity and works full time as a schoolteacher at a primary school. Her quality of life is currently relatively high. She has no problems with her daily activities and no pain, but is slightly anxious after receiving the diagnosis of oral cavity cancer. She has no shoulder complaints at present.

Healthcare systems and guidelines

As sources of information on the distribution of healthcare systems, we used both the internet site and the report 'OECD in Figures' of the Organisation for Economic Co-operation and Development (OECD). For guideline comparisons, we used guidelines provided by the professional HNC organizations in the included countries.[68-70]

Data analysis

Within-country evaluations were performed for the Netherlands, UK, and USA to explore the current situation in the centers treating HNC. Between-country comparisons were performed to assess differences in trends among neck treatment in OCSCC of the included countries. These differences were also assessed with regard to the different guidelines and healthcare systems

Results

Participants

In total, representatives of 45 different HNC centers completed the questionnaire: 10 from the Netherlands; 26 from the UK; and nine from the USA. Table 1 shows the number of patients treated per center in 2013.

Table 1: Number of patients with oral cavity squamous cell carcinoma at treated at the responding centers

Number of patients treated in 2013	Netherlands Centers (%) *n* = 10	UK Centers (%) *n* = 26	USA Centers (%) *n* = 9
Clinically negative neck			
0–10	0 (0%)	2 (7.7%)	1 (11.1%)
10–25	2 (20.0%)	13 (50.0%)	2 (22.2%)
25–50	3 (30.0%)	5 (19.2%)	2 (22.2%)
50–100	2 (20.0%)	4 (15.4%)	3 (33.3%)
100+	1 (10.0%)	0 (0%)	1 (11.1%)
Not specified	2 (20.0%)	2 (7.7%)	0 (0%)
Clinically positive neck			
0–10	0 (0%)	4 (15.4%)	0 (0%)
10–25	4 (40.0%)	10 (38.5%)	1 (11.1%)
25–50	3 (40.0%)	5 (19.2%)	6 (66.6%)
50–100	0 (0%)	2 (7.7%)	1 (11.1%)
100+	1 (10.0%)	3 (11.5%)	1 (11.1%)
Not specified	2 (20.0%)	2 (7.7%)	0 (0%)

Diagnosis

Table 2 provides an overview of the diagnostic modalities used to assess the neck in the Netherlands, UK, and USA, as well as the recommendations in the national guidelines. More specific information about diagnosis, i.e. which modality is deemed most reliable and the criteria for USgFNAC, is presented below.

In the Netherlands, all centers use USgFNAC, which is deemed to be the most reliable modality by all centers. The minimum axial size of the largest lymph node that should be present to indicate USgFNAC differed among the centers, as well as the method to measure this axial size. A minimum axial size of 5 mm is used as the criterion in seven centers (70%), whereas the other three centers (30%) use 3, 4, or 10 mm cut-off points, respectively. Seven centers (70%) measure the minimum axis diameter of the suspected lymph nodes, while two (20%) measure the long axis diameter, and one (10%) does not use a standard axis.

In the UK, 24 centers (92.3%) use USgFNAC. However, only 12 centers deem USgFNAC (46.2%) to be the most reliable method, whereas eight centers deem positron emission tomography (PET), PET combined with computerized tomography (PET-CT), or CT the most reliable diagnostic modality. Minimum axial size of the suspected lymph node, as an indication for USgFNAC, is 10 mm in nine centers and 5 mm in two centers; the other centers do not use a strict cut-off point.

In the USA, all centers use CT and all centers deem CT or PET/PET-CT to be the most reliable method. The minimal axial size of the lymph node required to perform an USgFNAC is 10 mm in all four centers using USgFNAC. Two centers perform fine needle aspiration without US guidance, and one center performs CT-guided fine needle aspiration biopsies.

Table 2: Diagnostic modalities performed to detect neck lymph node metastases.*

Diagnostic modalities	**Netherlands Centers (%) n = 10**	**UK Centers (%) n = 26**	**USA Centers (%) n = 9**
Palpation	9 (90.0%)	12 (46.2%)	7 (77.8%)
Ultrasound	2 (20.0%)	22 (84.6%)	1 (11.1%)
USgFNAC	10 (100.0%)	24 (92.3%)	4 (44.4%)
CT	8 (80.0%)	19 (73.1%)	9 (100.0%)
PET/PET-CT	5 (50.0%)	15 (57.7%)	8 (88.9%)
MRI	9 (90.0%)	22 (84.6%)	3 (33.3%)
Other modality	3 (30.0%)	3 (11.5%)	3 (33.3%)
Guideline recommendation	**USgFNAC is preferred**	**CT or MRI mandatory, High sensitivity of USgFNAC mentioned but not recommended**	**CT and/or MRI as indicated**

CT = computed tomography, MRI = magnetic resonance imaging, PET = positron emision tomography, USgFNAC = ultrasound-guided fine needle aspiration cytology.
*Multiple diagnostic modalities could have been selected by one center.

Management of the clinically negative neck

Table 3 provides an overview of the treatment strategies in patients with a cN_0 neck in the Netherlands, UK, and USA, and the recommendations of the respective guidelines. Table 4 provides an overview of the important primary tumor characteristics used in management decisions concerning the neck. More detailed information about these tumor characteristics is given in Tables A1–A3 in the appendix. Tables A4–A6 show details of the techniques used to measure the depth of invasion of the primary tumor, the techniques to detect sentinel nodes and the chosen strategies depending on the results of the SNB, respectively.

In the Netherlands, both WW and elective neck dissection (END) are frequently used strategies. Furthermore 40% of the centers perform an SNB procedure in at least some of their patients. Primary tumor size is the main characteristic that influences the decision for a certain strategy.

In the UK, most centers use END and WW as strategies for the cN_0 patients. Five centers (19%) sometimes perform an SNB procedure. Tumor size and depth of invasion are the main characteristics used to make treatment decisions.

In the USA, all centers perform ENDs, whereas about half of the centers use WW. In 22% of the centers, an SNB is performed. Depth of invasion, tumor size, and tumor site are all deemed to be important for treatment decisions.

Table 3: Treatment strategies performed in early stage clinically negative oral cavity squamous cell carcinoma.*

Treatment strategy	Netherlands Centers (%) n = 10	UK Centers (%) n = 26	USA Centers (%) n = 9
Elective neck dissection	9 (90.0%)	24 (92.3%)	9 (100.0%)
Watchful waiting	10 (100.0%)	23 (88.5%)	5 (55.6%)
Sentinel node biopsy	4 (40.0%)	5 (19.2%)	2 (22.2%)
Guideline recommendation	**Watchful waiting can be considered, END when probability of occult metastases >20%. SNB should not yet be considered as standard modality**	**END should be offered for patients with more than 15–20% of occult metastases. Only high sensitivity of SNB mentioned, but SNB is not recommended outside research**	**END strongly considered for invasion depth >4 mm, for invasion depth <2 mm END only in highly selective situations, between 2 and 4 mm, clinical judgment must be utilized. SNB is an alternative to END in centers with expertise**

END = elective neck dissection, SNB = sentinel node biopsy.
*Multiple strategies could have been selected by one center.

<u>Management of the clinically positive neck</u>
Table 5 shows the primary treatment strategies in the case of a cN_+ neck, and the recommendations of each of the national guidelines.
In the Netherlands, all centers use modified radical neck dissection as a strategy for cases with cN_+ necks. Seven centers (70%) also perform selective neck dissections in selected cases with cN_+ necks, of which two centers (29%) perform neck dissections of levels I–III and the other five (71%) level I-IV. Radiotherapy is only used in one (10%) of the centers in some of the cN_+ cases.
In the UK, 15 (58%) centers perform modified radical neck dissection for cases with cN_+ necks; 23 (50%) centers use selective neck dissections, which involve levels I–IV in most cases (65%). Radiotherapy is used in half of the included centers in cN_+ necks.
In the USA, eight centers (89%) use modified radical neck dissections, while six centers (67%) use selective neck dissections for cases with cN_+ necks.

Table 4: Primary tumor characteristics influencing the choice for the strategy for a early stage clinically negative neck.*,†

Primary tumor characteristics	Netherlands Centers (%)	UK Centers (%)	USA Centers (%)	Appendix table
Elective neck dissection	*n* = 9	*n* = 24	*n* = 9	
Size	8 (88.9%)	16 (66.7%)	8 (88.9%)	A1
Site	4 (44.4%)	9 (37.5%)	7 (77.8%)	A2
Depth of invasion	7 (77.8%)	21 (87.5%)	9 (100.0%)	A3, A4
Watchful waiting	*n* = 10	*n* = 23	*n* = 5	
Size	8 (80.0%)	15 (65.2%)	5 (100.0%)	A1
Site	2 (20.0%)	7 (30.4%)	3 (60.0%)	A2
Depth of invasion	7 (70.0%)	19 (82.6)	5 (100.0%)	A3
Sentinel node biopsy	*n* = 4	*n* = 5	*n* = 2	
Size	2 (50.0%)	2 (40.0%)	0 (0%)	A1
Site	1 (25.0%)	3 (60.0%)	1 (50.0%)	A2
Depth of invasion	1 (25.0%)	1 (20.0%)	1 (50.0%)	A3
Guideline recommendation	**No clear recommendation**	**No clear recommendation**	**Depth of invasion best predictor (see also recommendation in Table 3)**	

*Multiple characteristics could have been selected by one center.

†More details about characteristics involved can be found in the mentioned appendix tables.

3

Table 5: Chosen primary treatment strategies in the case of a clinically positive neck.*

Treatment strategy	Netherlands Centers (%) *n* = 10	UK Centers (%) *n* = 26	USA Centers (%) *n* = 9
Modified radical neck dissection	10 (100.0%)	15 (57.7%)	8 (88.8%)
Radiotherapy	1 (10.0%)	13 (50.0%)	0 (0%)
Selective neck dissection	7 (70.0%)	23 (88.5%)	6 (66.7%)
In case of selective neck dissection, dissection of: – Levels I–III – Levels I–IV – Not clear	 2 (28.6%) 5 (71.4%) 0 (0%)	 1 (4.3%) 15 (65.2%) 7 (30.4%)	 2 (33.3%) 2 (33.3%) 2 (33.3%)
Guideline recommendation	**IV–V can be performed for N+ neck, for N1 neck I–III can be sufficient in specific cases**	**Treatment of N+ neck, should be guided by treatment to the primary site. Selective neck dissection is adequate treatment for pN1 neck disease without adverse histological features**	**Selective neck dissection may be appropriate in certain patients with N1 to N2 disease.**

*Multiple treatments could have been selected by one center.

Follow-up

Table 6 shows the time intervals used by the centers to follow their patients, and the recommendations in each of the national guidelines. The diagnostic modalities used to detect regional failure during follow-up are listed in Table A7.

In the Netherlands, palpation and USgFNAC are used by 90% of the centers in the follow-up. In the UK, most centers (81%) use palpation in the follow-up. Ultrasound, USgFNAC, CT, PET/PET-CT and MRI are all modalities that are used in about 50% of the centers. In the USA, most centers (89%) use palpation, but also CT (78%) or PET/PET-CT (44%) are frequently used during follow-up.

Table 6: Time intervals between follow-up consultations

Time interval	Netherlands Centers (%) *n* = 10	UK Centers (%) *n* = 26	USA Centers (%) *n* = 9
First year			
2 weeks	0 (0%)	1 (3.8%)	0 (0%)
4 weeks	0 (0%)	12 (46.2%)	1 (11.1%)
6 weeks	5 (50.0%)	10 (38.5%)	3 (33.3%)
8 weeks	4 (40.0%)	1 (3.8%)	1 (11.1%)
10 weeks	0 (0%)	0 (0%)	0 (0%)
12 weeks	1 (10.0%)	2 (7.7%)	4 (44.4%)
Second year			
1 month	0 (0%)	1 (3.8%)	0 (0%)
2 months	1 (10.0%)	17 (65.4%)	2 (22.2%)
3 months	9 (90.0%)	7 (26.9%)	3 (33.3%)
4 months	0 (0%)	1 (3.8%)	4 (44.4%)
Third year			
3 months	0 (0%)	18 (69.2%)	1 (11.1%)
4 months	9 (90.0%)	8 (30.8%)	4 (44.4%)
5 months	0 (0%)	0 (0%)	0 (0%)
6	1 (10.0%)	0 (0%)	4 (44.4%)
Fourth year			
4 months	0 (0%)	11 (42.3%)	0 (0%)
5 months	0 (0%)	1 (3.8%)	0 (0%)
6 months	10 (10.0%)	14 (53.8%)	9 (100%)
Guideline recommendation	**2–3 monthly in first year, 3 monthly in second year, 4–6 monthly in third year and 6 monthly in subsequent years**	**At least 2 monthly in first 2 years and 3–6 monthly in subsequent years**	**1–3 monthly in first year, 2–6 monthly in second year, 4–8 monthly in third to fifth year and 12 monthly in subsequent years**

Case examples with a cN_0 neck

Table 7 shows the chosen strategies of the centers in two hypothetical cases of OSCC with a cN_0 neck, in two different scenarios. The cases and scenarios are presented in Box 1.

Table 7: Chosen strategies for two patients with a clinically negative neck in two different scenarios

Chosen strategy	Netherlands Centers (%) *n* = 9	UK Centers (%) *n* = 21	US Centers (%) *n* = 9
Patient I, scenario I			
END	0 (0%)	7 (33.3%)	6 (66.7%)
WW	6 (66.7%)	10 (47.6%)	2 (22.2%)
SNB	3 (33.3%)	4 (19.0%)	1 (11.1%)
Patient I, scenario II			
END	6 (66.7%)	17 (81.0%)	9 (100.0%)
WW	0 (0%)	1 (4.8%)	0 (0%)
SNB	3 (33.3%)	3 (14.3%)	0 (0%)
Patient II, scenario I			
END	2 (22.2%)	10 (47.6%)	5 (55.6%)
WW	3 (33.3%)	6 (28.6%)	2 (22.2%)
SNB	4 (44.4%)	5 (23.8%)	2 (22.2%)
Patient II, scenario II			
END	7 (77.8%)	18 (85.7%)	9 (100.0%)
WW	0 (0%)	0 (0%)	0 (0%)
SNB	2 (22.2%)	3 (14.3%)	0 (0%)

END = elective neck dissection, SNB = sentinel node biopsy, WW = watchful waiting.

Health statistics and guidelines

In 2011, 16.9% of the gross domestic product was used for healthcare in the USA, 12.1% in the Netherlands, and 9.3% in the UK. In the same year, the number of doctors per 1000 inhabitants was 3.1 in the Netherlands, 2.7 in the UK, and 2.4 in the USA, and the number of CT scanners available per 1 million inhabitants was 40.9 in the USA, 12.5 in the Netherlands, and 8.0 in the UK. In 2010, the number of MRI scanners per 1 million inhabitants was 31.5, 12.2, and 6.2 for the USA, Netherlands, and UK respectively.

Some differences exist in guidelines between the countries. The currently used Dutch guideline is from 2004, but it is being updated, while the UK (2011) and USA (2014) have more recent (updated) guidelines.

The biggest difference between the guidelines exists in diagnostics. The UK and US guidelines mention that CT or MRI should be performed to detect lymph node metastases. In the Dutch guideline, USgFNAC is recommended. In the UK guideline, USgFNAC is mentioned as an alternative to detect lymph node metastasis, while USgFNAC is not mentioned in the US guideline.

On the management of cN_0 and cN_+ necks, the basic recommendations of the three guidelines generally match, although there are some marked differences. The US guideline is the only one that has a clear recommendation about the use of tumor depth for the choice of strategy. There is no hard recommendation about the use of SNB in any of the guidelines. The US National Comprehensive Cancer Network (NCCN) incorporated SNB in their guideline as follows: 'Sentinel lymph node biopsy is an alternative to elective neck dissection for the identification of occult cervical metastasis in patients with early (T_1 or T_2) oral cavity carcinoma in centers where experience for this procedure is available'. In the current Dutch guideline, it is stated that SNB should not yet be considered as standard modality. In the concept of the updated Dutch guideline, SNB is also considered to be an alternative to END [not yet published]. In the UK guideline, SNB is not recommended outside of clinical trials.

In the Dutch and US guidelines, it is stated that selective neck dissection could be considered in certain cases of limited nodal disease. In the guideline from the UK, selective neck dissection is explicitly recommended for pathologically confirmed disease with a single ipsilateral positive lymph node smaller than 3 cm (pN_1 disease).

Discussion

Our results demonstrate a distinct variation in the management of the neck in OCSCC both within and between countries.

One of the obvious differences concerns the use of USgFNAC, which is used in the Netherlands and the UK, versus the use of CT and PET/PET-CT in the USA. Also notable is that WW is not often considered to be an alternative for management of the cN_0 neck in the UK, and especially not in the USA. Furthermore, SNB is used in all three countries, but only in a limited number of centers. A lot of variability between centers was seen in the use of primary tumor characteristics in the management of the cN_0 neck. In the USA, most centers take tumor size, site, and depth of invasion into account when making a decision for END, WW, or SNB, while this was less in the UK and the Netherlands. In the case of cN_+ necks, there is variation in the type of neck dissections performed in all three countries, and the frequent use of primary radiotherapy for N_+ cases in the UK is remarkable. Finally, there is quite some variation in the standard follow-up interval that is used.

In the presented hypothetical cases, variation in the chosen strategies in the cN_0 neck was indeed apparent. The presence of variability in used management strategies was confirmed by the two cases. More variability was present in scenario I of both cases, reflecting that decisions are more complex for early stage tumors.

To the best of our knowledge, this study is the first study to provide insight into the differences in management of oral cavity cancer both within and between countries. The study covers a large part of the treatment pathway for these patients, and shows detailed information along all parts of the pathway. Others have performed similar surveys over the years within one country. Werning et al. used a survey with a clinical scenario which was comparable to scenario II of our cases to determine the approach to the cN_0 neck in the USA. Of the 313 otolaryngologists, 66% would perform END, while 16% preferred radiotherapy, and 13% preferred to observe the patient. In our hypothetical cases, all respondents from the USA would perform END.[71] In a study by van de Bree et al.a similar patient was presented to the eight main Dutch HNC centers.[18] Of these centers, 75% would treat this patient with END, while 25% would opt for WW.[72] In the current study, 67% or 78% would choose END, depending on the overall status of the patient. The remaining centers would use SNB. It seems like the SNB has replaced WW in this presented case in the Netherlands. Looking at these two comparisons with earlier performed surveys, it seems that surgeons let tumor control prevail above the morbidity of surgery.

The current study includes all main centers in the Netherlands and a high number of centers in the UK. The low number of responding centers in the USA was, however, the major limitation of this study. However, most of the responding centers from the USA were high-volume centers, treating a high number of patients per year (Table 1). We therefore, believe that the results provide a reasonable impression regarding the management of the neck in oral cancer in the USA.

Some of the observed variation could be related to differences in healthcare systems and/ or guidelines. The frequent use of CT in the USA and the high use of USgFNAC in the Netherlands could be related to the differences in guidelines. Why these differences exist in the guidelines, despite the same available evidence, is unclear but may be explained by 'tradition'. Furthermore, the depth of invasion was used in all centers in the USA to guide the management of the cN_0 neck. This may be caused by the fact that the US guideline is the only one that has a clear recommendation on the use of depth of invasion. Another finding that could be explained by the difference in guidelines is the higher use of selective neck dissection in cN_+ necks in the UK compared with the Netherlands and the USA.

Besides differences in guidelines, differences in healthcare systems are also present between the included countries. This could be an explanation for some of the variations like the high use of CT in de USA. Although, it is difficult to directly link information about the healthcare systems to the existing variations, it is important to keep such differences in mind when evaluating the variation in used strategies, especially between countries.

Although some variation could be explained by differences in healthcare systems between countries, it is likely that an important part of the differences in the results is due to the absence of clear evidence about optimal management of the neck in early stage OCSCC. Much research on management of the neck in OSCC patients has already been published, but the evidence apparently is insufficient to guide decisions in an uniform manner.

Conclusion

In conclusion, there is high variation in the management of the neck in early OCSCC. Direct evidence about optimal management decisions, deriving from sources such as direct comparison in decision models or comparative effectiveness studies, seems to be needed to establish more evidence-based management and more uniformity in clinical practice. It is likely to result in better quality of care and facilitates international multicenter studies.

Acknowledgments

We would like to thank all head and neck cancer centers who participated in this research and answered our questionnaire. Furthermore, we would like to thank the Dutch Head & Neck Society (NWHHT) and BAHNO for the access to the centers affiliated with them.

Appendix: More specific results of the questionnaire.

Table A1: Tumor size influencing the choice of primary treatment in early stage clinically negative oral cavity squamous cell carcinoma.

Tumor size characteristics	Netherlands Centers (%)	UK Centers (%)	US Centers (%)
Elective neck dissection	*n* = 9	*n* = 24	*n* = 9
Size does not influence choice for END	1 (11.1%)	8 (33.3%)	1 (11.1%)
More often performed in T_2 tumors	8 (88.9%)	16 (66.7%)	8 (88.9%)
Only performed in T_2 tumors	0 (0%)	0 (0%)	0 (0%)
Watchful waiting	*n* = 10	*n* = 23	*n* = 5
Size does not influence choice for WW	2 (20.0%)	8 (34.8%)	0 (0%)
More often performed in T_1 tumors	7 (70.0%)	10 (43.5%)	4 (80.0%)
Only performed in T_1 tumors	1 (10.0%)	5 (21.7%)	1 (20.0%)
Sentinel node biopsy	*n* = 4	*n* = 5	*n* = 2
Size does not influence choice for SNB	2 (50.0%)	3 (60.0%)	1 (50.0%)
More often performed in T_1 tumors	2 (50.0%)	2 (40.0%)	1 (50.0%)
Only performed in T_1 tumors	0 (0%)	0 (0%)	0 (0%)
More often performed in T_2 tumors	0 (0%)	0 (0%)	0 (0%)
Only performed in T_2 tumors	0 (0%)	0 (0%)	0 (0%)

END =elective neck dissection, SNB sentinel node biopsy, WW, watchful waiting.

Table A2: Tumor sites where the strategy concerned is more often performed in clinically negative oral cavity squamous cell carcinoma.*

Tumor site characteristics	Netherlands Centers (%)	UK Centers (%)	USA Centers (%)
Elective neck dissection	*n* = 9	*n* = 24	*n* = 9
Not site-dependent	5 (56.6%)	15 (62.5%)	2 (22,2%)
Floor of the mouth	4 (44.4%)	9 (37.5%)	7 (77.8%)
Tongue	4 (44.4%)	9 (37.5%)	7 (77.8%)
Lip	1 (11.1%)	0 (0%)	1 (11.1%)
Hard palate	1 (11.1%)	1 (4.2%)	3 (33.3%)
Retromolar trigone	4 (44.4%)	7 (29.2%)	7 (77.8%)
Buccal mucosa	4 (44.4%)	6 (25.0%)	7 (77.8%)
Alveolar ridge	3 (33.3%)	7 (29.2%)	4 (44.4%)
Watchful waiting	n = 10	n = 23	n = 5
Not site dependent	7 (70.0%)	16 (69.6%)	2 (40.0%)
Floor of the mouth	0 (0%)	3 (13.0%)	2 (40.0%)
Tongue	0 (0%)	4 (17.4%)	2 (40.0%)
Lip	3 (30.0%)	5 (21.7%)	3 (60.0%)
Hard palate	3 (30.0%)	5 (21.7%)	2 (40.0%)
Retromolar trigone	0 (0%)	1 (4.3%)	2 (40.0%)
Buccal mucosa	0 (0%)	1 (4.3%)	1 (20.0%)
Alveolar ridge	0 (0%)	0 (0%)	3 (60.0%)
Sentinel node biopsy	n = 4	n = 5	n = 2
Not site-dependent	3 (75.0%)	2 (40.0%)	1 (50.0%)
Floor of the mouth	1 (25.0%)	1 (20.0%)	1 (50.0%)
Tongue	1 (25.0%)	3 (60.0%)	1 (50.0%)
Lip	0 (0%)	1 (20.0%)	1 (50.0%)
Hard palate	0 (0%)	1 (20.0%)	0 (0%)
Retromolar trigone	0 (0%)	1 (20.0%)	0 (0%)
Buccal mucosa	0 (0%)	2 (40.0%)	0 (0%)
Alveolar ridge	0 (0%)	1 (20.0%)	0 (0%)

*Multiple sites could have been selected by one center.

Table A3: Depth of invasion cut-off points influencing the choice for primary treatment in early stage clinically negative oral cavity squamous cell carcinoma.

Invasion depth	Netherlands Centers (%)	UK Centers (%)	USA Centers (%)
Elective neck dissection	*n* = 9	*n* = 24	*n* = 9
Not depending on depth	2 (22.2%)	3 (12.5%)	0 (0%)
>2 mm	0 (0%)	2 (8.3%)	0 (0%)
>3 mm	0 (0%)	4 (16.7%)	3 (33.3%)
>4 mm	5 (55.6%)	9 (37.5%)	5 (55.6%)
>5 mm	2 (22.2%)	4 (16.7%)	0 (0%)
No clear cut-off point	0 (0%)	2 (8.3%)	1 (11.1%)
Watchful waiting	*n* = 10	*n* = 23	*n* = 5
Not depending on depth	3 (30.0%)	4 (17.4%)	0 (0%)
<2 mm	0 (0%)	3 (13.0%)	0 (0%)
<3 mm	1 (10.0%)	6 (26.1%)	1 (20.0%)
<4 mm	4 (40.0%)	6 (26.1%)	4 (80.0%)
<5 mm	2 (20.0%)	2 (8.7%)	0 (0%)
<7 mm	0 (0%)	1 (4.3%)	0 (0%)
No clear cut-off point	0 (0%)	1 (4.3%)	0 (0%)
Sentinel node biopsy	*n* = 4	*n* = 5	*n* = 2
Not depending on depth	3 (75.0%)	4 (80.0%)	1 (50.0%)
<4 mm	1 (25.0%)	1 (20.0%)	1 (50.0%)

CT = computed tomography, END = elective neck dissection, MRI = magnetic resonance imaging.

Table A4: Diagnostic techniques performed to measure invasion depth of the primary tumor (in centers that perform END and use invasion depth as a parameter).

Techniques performed	Netherlands Centers (%) (*n* = 9)	UK Centers (%) (*n* = 24)	USA Centers (%) (*n* = 9)
Palpation	2 (22.2%)	5 (20.8%)	4 (44.4%)
CT	2 (22.2%)	5 (20.8%)	0 (0%)
MRI	6 (66.7%)	7 (29.2%)	1 (11.1%)
Ultrasound	1 (11.1%)	2 (8.3%)	0 (0%)
Biopsy	3 (33.3%)	16 (66.7%)	3 (33.3%)
Histology of the dissected specimen	5 (55.6%)	14 (58.3%)	7 (77.8%)

Table A5: Techniques used to detect the sentinel node(s) in centers performing SNB

Techniques performed	Netherlands Centers (%) (*n* = 4)	UK Centers (%) (*n* = 5)	USA Centers (%) (*n* = 2)
Lymphoscintigraphy	4 (100.0%)	5 (100.0%)	2 (100.0%)
Gamma probe	4 (100.0%)	5 (100.0%)	2 (100.0%)
Freehand SPECT	1 (25.0%)	4 (80.0%)	1 (50.0%)
Blue dye detection	3 (75.0%)	4 (80.0%)	0 (0%)

SNB = sentinel node biopsy, SPECT = single-photon emission computed tomography.

Table A6: SNB results and consequences in centers performing SNB.

SNB strategies	Netherlands Centers (%) *n* = 4	UK Centers(%) n = 5	USA Centers(%) n = 2
No sentinel node detected			
Watchful waiting	3 (75.0%)	2 (40.0%)	2 (100%)
Dissection level I–III	1 (25.0%)	0 (0%)	0 (0%)
Dissection level I–IV	0 (0%)	2 (40.0%)	0 (0%)
Not specified	0 (0%)	1 (20.0%)	0 (0%)
Metastases present in level I sentinel node			
Dissection level I–III	0 (0%)	1 (20.0%)	1 (50.0%)
Dissection level I–IV	2 (50.0%)	4 (80.0%)	0 (0%)
Dissection level I–V	2 (50.0%)	0 (0%)	1 (50.0%)
Metastases present in level II sentinel node			
Dissection level I–III	0 (0%)	0 (0%)	1 (50.0%)
Dissection level I–IV	1 (25.0%)	5 (100.0%)	0 (0%)
Dissection level I–V	3 (75.0%)	0 (0%)	1 (50.0%)
Metastases absent in sentinel node			
Watchful waiting	4 (100.0%)	4 (100.0%)	2 (100.0%)

SNB = sentinel node biopsy.

Table A7: Diagnostic modalities used to routinely detect regional failure during follow-up.*

Diagnostic modalities	Netherlands Centers (%) *n* = 10	UK Centers (%) *n* = 26	USA Centers (%) *n* = 9
Palpation	9 (90.0%)	21 (80.8%)	8 (88.9%)
Ultrasound	2 (20.0%)	15 (57.7%)	0 (0%)
USgFNAC	9 (90.0%)	15 (57.7%)	2 (22.2%)
CT	1 (10.0%)	13 (50.0%)	7 (77.8%)
PET (PET-CT)	1 (10.0%)	14 (53.8%)	4 (44.4%)
MRI	3 (30.0%)	15 (57.7%)	0 (0%)

CT = computed tomography, MRI = magnetic resonance imaging, PET = positron emission tomography, USgFNAC = ultrasound-guided fine needle aspiration cytology.
*Multiple modalities could have been selected by one center.

Chapter 4

Innovating head and neck cancer care: an interactive evaluation

Ester A Rake

Tim M Govers

Maroeska M Rovers

Rob P Reuzel

Submitted

Abstract

Background: Contemporary operating theaters (OTs) are built that are targeted at research and development and evidence-based surgical innovation. We explored the benefits of an interactive evaluation (IE) in innovating HNC-care related to these OTs.

Methods: An IE involves anonymously exchanging concerns, claims, and issues between stakeholders during interviews. Interviews were transcribed and analyzed. After the first interview cycle, we circulated a summary of important issues among stakeholders, to which they responded in the second cycle. Finally, we summarized topics and recommendations on which consensus was achieved.

Results: Most stakeholders agreed that the time interval between diagnosis and treatment is too large. Secondly, communication and information across disciplines should be improved. Furthermore, all stakeholders agreed that research on resection margins is highly valuable.

Conclusions: This pilot study showed that IE could be a valuable method for innovating HNC-care and other multidisciplinary fields.

Introduction

Every year half a million people worldwide are diagnosed with head and neck cancer (HNC). Involved risk factors are substantial alcohol use and smoking.[8,73] Anatomic and pathological variability of HNC makes treatment complex and this contributes to the fact that quality of life of patients is affected by treatment.[74] An important aspect of the complexity of treatment is the relative high incidence of occult metastasis; new techniques, are aimed at optimizing diagnosis and treatment of metastases.[75] Furthermore, a lot of research and innovations are designed to improve survival and quality of life.[8] Because of these complexities, the management of HNC patients urges a multidisciplinary approach.[74]

This study focuses on innovations in HNC care. In the Radboud University Medical Center, Nijmegen, the Netherlands (Radboudumc), state of the art operating theaters (OT's) are being built, targeted at research and development of evidence based surgical innovations including the whole process before and after surgery. The ultimate goal is to deliver more efficient care and improve short- and long-term patient outcomes and satisfaction.[76] To reach this goal we have to know what stakeholders expect and how the health care pathway should be adapted to improve patient outcome and satisfaction.

Successfully innovating the HNC health care pathway requires that complexities are acknowledged and that innovations conform to the perspectives of experts involved. Our hospital believes that this includes the perspective of the patients, as patient involvement is crucial to improving health service delivery and quality.[77] Moreover, this implies that the whole health care process should be considered, not only the actual surgery in the OT.[78]

Interactive evaluation is a kind of fourth generation evaluation that is aimed at systematically eliciting the views of persons involved, particularly their norms and values, and on this basis identify shared values.[79] Next, it seeks to develop a shared view on where the technology should develop, what are desirable scenarios, and what are appropriate outcome measures for assessment.

Interactive evaluation matches the contemporary ideal of participatory medicine, a model of cooperative health care that seeks to achieve active involvement by patients, professionals, caregivers, and others across the continuum of care on all issues related to an individual's health. Choices that are made in policy research should be such, that the various stakeholders consider the results relevant and valid. To help achieve this, it has been suggested that a scoping exercise should be conducted. That is, efforts should be made to elicit stakeholders' perspectives and invoke these in the assessment. There are several reasons for performing such participatory research. First, it is acknowledged that a technology (defined as prevention and rehabilitation, vaccines, pharmaceuticals and devices, medical and surgical procedures, and the systems within which health is protected and maintained) is not just an isolated artifact or device, but something that is interwoven with a societal context to be addressed. Participatory research leads to a better understanding of what a technology actually does in a societal context. Second, participation opens large sources of relevant expertise and thus generates potentially important information, i.e. it helps to identify a fairly comprehensive list of perspectives and issues related to a technology. Interactive evaluation as suggested in this proposal distinguishes itself from participatory evaluation in that participants are not only consulted, but interact with each other. The resulting learning process helps to ensure that views are well considered, that the expertise, knowledge and creativity available within the field is used, and that stakeholders have an opportunity to develop a shared view on the technology under consideration. Furthermore, the synthesis of perspectives may result in fairly creative solutions. In other words, interactive evaluation will help to 1) identify criteria of merit and acceptability of the new technology; 2) achieve agreement about regulations (policy recommendations) to be implemented to meet these criteria; 3) formulate relevant research questions.

An interactive evaluation seems a suitable tool to investigate all these views of stakeholders on the innovation of the HNC health care pathway. Therefore, the primary aim of this pilot study is to explore the benefits of an interactive evaluation in the context of innovating and improving HNC care. Hence, we answered the following questions: what is the perspective on improving the current head and neck cancer health care process shared by all stakeholders involved? How would these stakeholders define and prioritize targets for optimizing this process and goals for improvement?

Materials and Methods

We identified stakeholders on the basis of the HNC care pathway and snowball sampling(79). Subsequently, semi-structured, in-depth interviews were conducted with each of the stakeholders, who were visited in their natural environment, mostly at their workplace. After a first cycle of interviews, a second cycle ensued. Two interviewers were involved. Each stakeholder was interviewed twice by the same interviewer.

The interviews were recorded, transcribed verbatim and analyzed using 'Atlas.ti 7 software' for qualitative data analysis.[80] We analyzed the interviews by open coding to mark relevant issues and topics.[81] Moreover, we used the method of reconstructing interpretative frames to distinguish between problem definitions, solutions, empirical background knowledge, and normative preferences of each stakeholder.[82] After each interview, a summary was returned to the participant for validation (member checking). A network analysis was performed to generate a network view and identify overarching issues.

During interviews, stakeholders were invited to respond to rival claims as forwarded by other stakeholders. In order to prevent power and the quality of existing relations having an unwanted impact on the evaluation, this exchange of arguments was performed

anonymously. The idea is that the ensuing interaction leads to a process of vicarious learning and the development of a shared view on the quality of the Health Care Pathway (HCP).
During the first interview cycle, we focused on quality of care and on the current HCP. After this cycle, we circulated an overall summary of the most important issues among stakeholders to which the stakeholders could respond in the second cycle. Eventually, the results from the interviews provided policy recommendations, based on issues on which the various stakeholders had reached consensus.

Results

Respondents

Thirteen stakeholders were invited to participate in this study, of whom 12 were willing to take part (see Table 1). One stakeholder refused because of concerns about the interpretation of outcomes of the IE. Moreover, one stakeholder did not participate in the second interview cycle due to time constraints. Interviews lasted between 18 and 90 minutes. We performed two interview cycles and anonymity of stakeholders was warranted. The results of the network analysis are shown in Appendix A. Topics that surfaced most often will be described in the following paragraphs.

Current HCP

After visiting the General Practitioner or Dentist (GP/D), patients are sent (often via a general hospital) to a specialized hospital. Eight medical centers (with or without a preferred partner clinic) in the Netherlands provide HNC health care, which are all connected to the Dutch Head and& Neck Society (DHNS/NWHHT). In the Radboudumc, two departments preside in providing care for patients with (suspected) HNC: the department of Oral & Maxillofacial Surgery (OMF) and the department of Otorhinolaryngology and Head and Neck Surgery (ENT). All stakeholders were positive about this collaboration and they mentioned that many health care professionals are very dedicated to providing care for the HNC patients. The diagnostic process contains two days in which several examinations are done in order to diagnose patients. Subsequently, a multidisciplinary team meets weekly to discuss treatment options of specific patients. These treatment options are discussed with the patient, usually consisting of surgery, radiotherapy, chemotherapy or a combination of these. Consequently, the patient will be treated and afterwards follow-up care is given at the outpatient clinic.

Table 1: Overview of participating stakeholders

Stakeholder	Participated in interview 1?	Participated in interview 2?	Interviewer A or B
Oral & Maxillofacial Surgeon	Yes	Yes	A
Ear-, Nose-, Throat physician	Yes	Yes	A
Nurse of H&N surgery ward	Yes	Yes	A
Patients association 'Stichting Klankbord'	Yes	Yes	A
Physiotherapist representative	Yes	No	A
Anaesthesiologist	Yes	Yes	A
Representative of the H&N outpatient clinic and of the paramedical working field	Yes	Yes	A
Patient	Yes	Yes	A
Patient	Yes	Yes	B
Radiologist	Yes	Yes	B
Medical oncologist	Yes	Yes	B
Radiotherapist	Yes	Yes	B

<u>Continuity of care</u>

Constraints and improvements were mentioned concerning the continuity of the HCP. Stakeholders described that the pathway from GP/D to one of the eight HNC centers, spans an unduly long time interval of approximately six weeks. Patients said they experienced much uncertainty during this period, caused by a lack of information. According to some stakeholders, creating awareness among health care professionals in primary health care, and consequently accelerating the referral to a specialized hospital can save time.

All stakeholders, except for one patient, mentioned that the waiting time between diagnosis and surgery should be reduced. The guideline of the NWWHT states 'waiting time' is defined, as that 80% of new HNC patients has to start their curative therapy within 30 calendar days.[83] They stated that being diagnosed in two days is out of proportion to a waiting time of about a month to the start of treatment. The capacity of the operating theatres (OT's) was mentioned as the main cause of the long time interval. Another factor, which was also identified in relation to other issues, is the large fluctuation in the number of newly diagnosed patients.

Ideas were suggested to shorten this 'waiting-time' interval, such as a flexible planning of available time at OT's for ENT and OMF. Due to fluctuations in newly diagnosed patients and the sometimes complex and time-consuming procedures, it is not possible to exactly predict the hours needed at the OT's. As a result, the only possibility to deliver care without knowing how many patients will need surgery a week, is to adjust the planning to allow for more flexibility. A second solution would be to extend the current working hours and days. For example, health care professionals should start working in the evening or at a Saturday morning. Stakeholders were positive about this solution. However, more personnel should be hired to fill these hours, which raised the question whether this would be affordable.

Information and communication

Continuity of care was also associated with communication between health care professionals. Many of the communication failures between health care professionals seemed to be caused by incorrect communication and provision of information. Four stakeholders mentioned that a better understanding of the work of colleagues in adjacent fields of health care would be beneficial for the health care process, partly because of such understanding could improve mutual acceptance of acknowledgement of choices made. In order to improve teamwork, communication, and streamlining of the process, two stakeholders proposed to implement a yearly audit with all health care professionals involved (GP/D, paramedical professionals, surgeons) and patient associations. One stakeholder proposed to create a head and neck oncology department at which all professionals within HNC care should be located. Most of the other stakeholders did, however, not agree. The most important reason was that this would be very costly and not efficacious.

Many stakeholders said that treatment-information is hard to find and not clearly described. Patients experienced unmet information needs during the time interval between referral from a GP/D to a medical center. All stakeholders supported the idea of creating a national website with accurate information of all medical centers treating HNC. Moreover, stakeholders described that the post-treatment period is difficult for patients. For example, patients experience eating- and speaking problems, and have great difficulties finding relevant information. Patients no longer have (weekly) appointments at the medical center and they have to face their problems by themselves. Stakeholders said that patients cope differently with their HNC-related problems.

Surgical innovation

An issue raised by both surgeons and patients, was the completeness of resected tumors. The surgical margins are an important factor for prognosis and for determining whether further treatment is necessary. Patients said they lived in uncertainty for about a week after surgery, before receiving the results of the pathological examination of these margins. One of the patients mentioned that receiving unfavorable results account for the most impactful event during the disease. The surgeons mentioned that during surgery, they do not know whether they removed enough tissue to eliminate all tumor cells. Both patients and physicians indicated that solving this problem is of high importance and would have a significant impact on head and neck cancer. They suggested that studies at the research

Ots should be focused on imaging of the primary tumor and resection margins during surgery. From the perspective of radiotherapy, imaging by a CT- or MRI-scans could be feasible to improve the accuracy of radiotherapy.

Currently, the status of the neck is a controversial topic in treatment of HNC, because of the relative low sensitivity of current diagnostics for cervical lymph node metastases. A sentinel node biopsy is a relatively new technique that has a high sensitivity for these metastases. The sentinel node biopsy technique (SNB) was discussed during the interviews with stakeholders. Stakeholders did, however, not agree as to whether this technique should be used regularly in the near future, because according to them, currently available literature is not consistent about this topic. Therefore, they stated that research into SNB at the new Ots would be valuable.

Discussion

In this study, we explored the potential value of interactive evaluation for innovation in the field of HNC care. As a result, we found that stakeholders agree on the following issues. First of all, the continuity of the HCP should be improved. The most important issue in this respect was the 'waiting time' between diagnosis and treatment. Most stakeholders stated that this interval was too large, and called for more flexible planning of the OT capacity and possibly extended working hours. Secondly, the study showed that the quality of information and communication across disciplines should be improved. Thirdly, all stakeholders agreed that research on resection margins is highly valuable and ought to be an important focus of research.

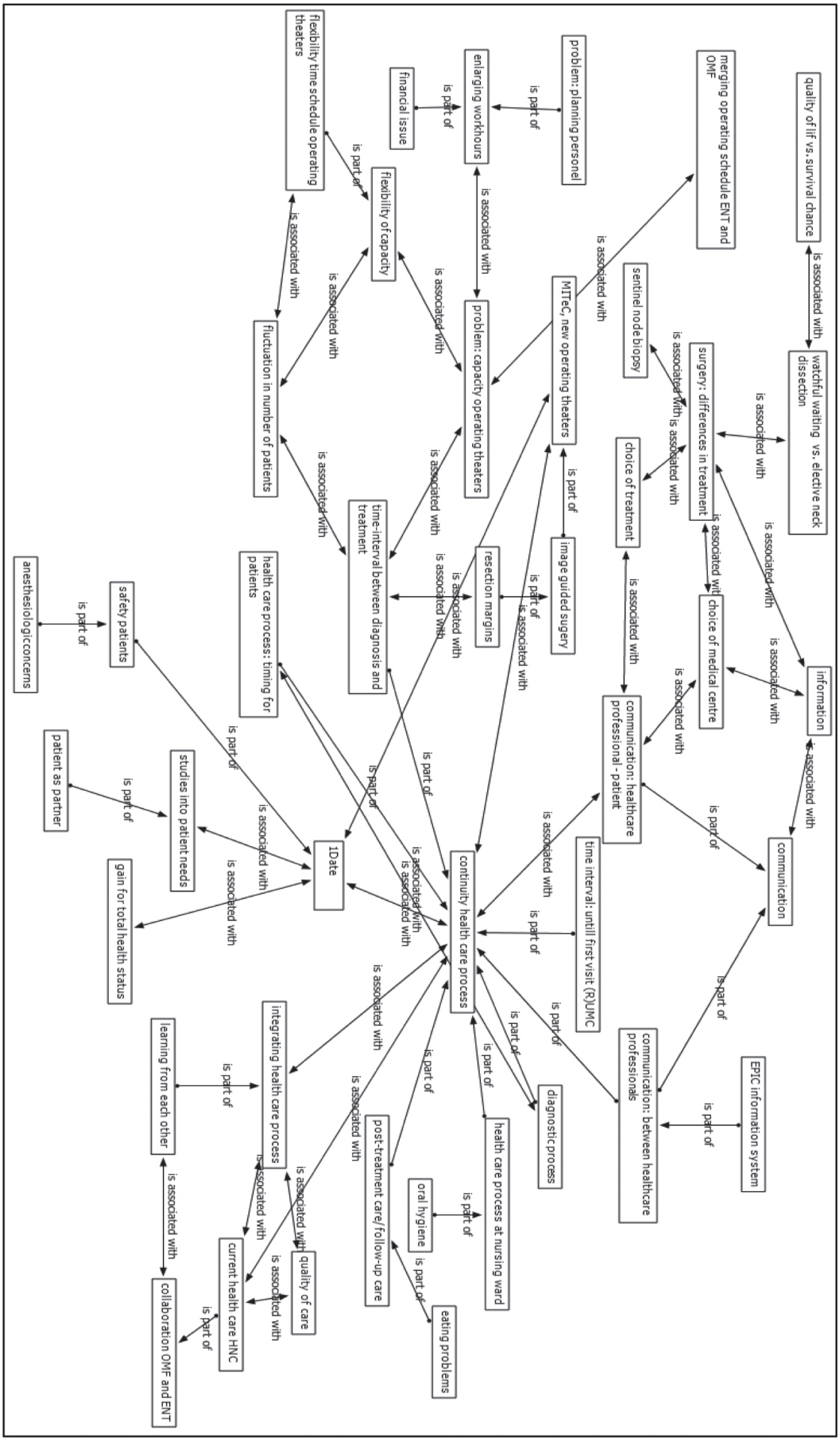
As far we know this is the first interactive evaluation to study the quality of HNC care and identify priority targets for innovation. Although this has been an exploratory study, we already perceived some important benefits of our approach. First of all, we were able to include perspectives of different stakeholders in the field, and succeeded in identifying a few widely endorsed recommendations, whilst a few other suggestions could efficiently be shown to be less effective, feasible, or appropriate. Secondly, by including various stakeholders, we were able to study HNC care from an *integrated* perspective. As a result, continuity of care could properly be discussed across disciplines and departments, whereas otherwise such discussion would have been confined by the boundaries of a specific discipline or department (e.g. OT's) and opportunities for improvement would have been limited accordingly. Furthermore, an important feature of this method is the incorporation of the patients' perspective. The added value of this shows for instance in the issue of the interval between diagnosis and treatment, on which not every patient agreed with the health care professionals. Finally, this method prioritized the issues we should focus on in the future. The results showed that patients as well as health care professionals agreed on investigating innovative techniques regarding resection margins.

Importantly, we believe that the procedure of involving stakeholders ought to be valued in itself, apart from the results. Six stakeholders mentioned that they were glad to read that other stakeholders shared their vision regarding specific topics. One stakeholder said that he learned from views of other stakeholders by participating in our study. By providing persons involved with a sense of ownership, future improvements in HCN care probably meet with more acceptance and therefore are more feasible.
Some limitations of this study should also be discussed. First, we included a limited number of stakeholders in this study, although we do believe that the included stakeholders represented the most important perspectives. Second, we were confronted with a stakeholder refusing to participate. Since the evaluation was targeted at the local situation and hence the person in question is properly regarded an expert rather than a stakeholder, it was not too problematic. However, it does alert us to the fact that by its very nature interactive evaluation depends on participants willing to discuss the status quo and abide by the results of the procedure. In this sense, interactive evaluation affects the power balance that usually shapes practice.[84]
In conclusion, we showed that an interactive evaluation can be used to develop a widely endorsed perspective on how to identify and prioritize viable innovations in health care. The results obtained in this study suggest that innovation of HNC care should focus on flexible OT planning to reduce the time between diagnosis and treatment, improved interdisciplinary communication, and research into resection margins.

Acknowledgements

We thank all stakeholders for their participation in this study. Especially we would like to thank the patients that were willing to participate and share their story and ideas about head and neck cancer care.

Appendix A: network view



Part II

Chapter 5

Sentinel node biopsy for squamous cell carcinoma of the oral cavity and oropharynx: A diagnostic meta-analysis

Tim M Govers

Gerjon Hannink

Matthias AW Merkx

Robert P Takes

Maroeska M Rovers

Published: Oral Oncology 2013, Aug; 49 (8): 726-32

Abstract

Background: The aim of the study was to systematically assess the accuracy of a sentinel lymph node biopsy (SLNB) in cT_1/T_2N_0 oral cavity and oropharyngeal squamous cell carcinoma patients.

Methods: We searched electronic databases, including EMBASE and MEDLINE (Pubmed) up to November 7 2012, by combining oral cancer keywords with sentinel node biopsy keywords. We included diagnostic accuracy studies which used neck dissection or follow-up as a reference test for the sentinel node biopsy. Study characteristics and measures of accuracy were extracted. Diagnostic accuracy was calculated from 2 x 2 tables.

Results: 21 studies (847 patients) could be included. Most of these patients had oral cavity squamous cell carcinoma (OCSCC). The pooled data showed an overall sensitivity of 0.93 [95% CI 0.90-0.95]. Subgroup analysis showed no significant differences between subgroups.

Conclusions: The high sensitivity of SLNB supports a role in the diagnostic work-up of OCSCC.

Introduction

Oral cavity and oropharyngeal squamous cell carcinoma (OCSCC and OPSCC) are considered an important part of the global burden of cancer, mainly due to the widespread use of tobacco and alcohol.[61] The most important prognostic factor is the presence of cervical lymph node metastases, which can decrease the 5-year survival rates to lower than 50%.[14] Exact staging of the neck is therefore crucial in managing this type of cancer. Staging by palpation and imaging techniques (eg. MRI, CT, ultrasound-guided fine needle aspiration cytology (USgFNAC) are not sensitive enough in detecting micrometastases, resulting in a high incidence of occult metastases in the neck.[85] In the literature rates between 23% and 43% have been reported.[86] Because of this, elective neck dissection (END) is the standard of care in clinically negative necks (cN_0) of early stage (T_1/T_2) oral cavity and oropharyngeal squamous cell carcinomas in most institutions. However, this implies overtreatment and treatment-associated morbidity in the majority of patients.[18]

Sentinel lymph node biopsy (SLNB) has emerged as an alternative or additional staging procedure. The SLNB procedure is based on the concept that tumor cells will spread from the primary site to a single node or group of nodes (the sentinel nodes), before progressing to the remainder of the lymph nodes. A radiotracer, possibly in conjunction with colored dye injected into the primary tumor allows for identifying the sentinel nodes. Radiolocalization of the sentinel node consists of a preoperative lymphoscintigraphy either or not with SPECT/CT and the intraoperative use of a hand-held gamma probe and/or portable gamma-camera.[87] Histopathological evaluation of sentinel nodes may allow accurate prediction of the disease status. False negative results can have several causes including uneven radionuclide injection, obscuring of sentinel lymph nodes by the radioactive signal of the primary tumor, and lymphatic obstruction by gross tumor, resulting in redirection of lymphatic flow.[88]

Although SLNB is still an invasive procedure, it gives less morbidity than selective neck dissection by sparing relevant structures.[63,64] Therefore, SLNB can have great consequences for the treatment of OCSCC and OPSCC patients. More accurate staging of the neck by SLNB could be a serious alternative for direct elective neck dissection.

The aim of this diagnostic meta-analysis therefore was to study the diagnostic accuracy of SLNB in cT_1/T_2N_0 oral cavity and oropharyngeal squamous cell cancer patients.

5

Materials and Methods

Data sources and searches

We searched electronic databases, including EMBASE and MEDLINE (Pubmed) from inception up to November 7 2012, by combining oral cavity cancer keywords with sentinel node biopsy keywords (see Appendix). No restrictions on language were used in the search. All citations identified by the search were imported into a EndNote bibliographic database.[89]

Study selection

Initially the titles and abstracts of the search results were screened. Subsequently, the reports were reviewed according to pre-defined inclusion and exclusion criteria.

We included studies if they fitted the following criteria: Human cT_1/T_2N_0 oral cavity and/or oropharyngeal cancer patients; Full text available in English; Possible to derive absolute numbers of observations (True Positives [TP], False Positives [FP], False Negatives [FN] and True Negatives [TN]; Reference standard was a neck dissection or a follow-up of at least 18 months; The target condition was metastasis in one on the lymph nodes of the neck (pN_+). Case studies were excluded.

Data extraction and quality assessment

In order to obtain 2 x 2 contingency tables from the included studies, we extracted or calculated TP, FN and TN. In sentinel node biopsies FP results are not possible. The number of observations (TP, FP, FN and TN) were extracted separately for OCSCCs and OPSCCs when possible.

We designed a data extraction form specifically to collect details from selected studies. The following data were extracted for each individual study: Sentinel node detection rate (percentage of patients in whom at least 1 sentinel node was detected for biopsy), previous diagnostics (N0 determination), age and gender distribution, SLN identification (lymfoscintigraphy, gamma probe and/or blue dye substance), histology (H&E staining and/or immunohistochemistry), levels of neck dissection or mean follow-up time as reference standard, study design, aim of study, country, year, and whether or not it was a multicenter study. The quality of the included studies was assessed using the QUADAS-2 tool for diagnostic studies.[90]

Analysis

We constructed 2 x 2 contingency tables of TP, FP, FN and TN cases in RevMan 5.[91] We calculated sensitivity with a 95% confidence interval (CI) for each study separately. Because SLNB cannot be false positive, the specificity and the positive predictive value of the test are always 1. We used the the statistical software package SAS to carry out the meta-analyses.[92,93] Two subgroups of clinically relevant covariates; primary location (oral cavity or oropharynx) and sentinel node protocol were analyzed on the sensitivity of SLNB. Furthermore, we performed subgroup analyses to assess possible differences in sensitivity between studies using a neck dissection or follow-up as the reference test.

Results

In total 1884 studies were identified. After removal of duplicates and screening of titles and abstracts, 116 studies were selected for a full text evaluation. 21 studies fulfilled the inclusion criteria and a data extraction form was completed for these studies (Figure 1).[88,94-113] Study characteristics and treatment protocols of the 21 included studies are listed in Table 1A. Population characteristics are listed in Table 1B.
The studies contained 847 cases in total. Most studies also mentioned the detection rate of sentinel nodes. In these studies at least one sentinel node was detected in almost all patients and a sentinel node biopsy could thus be performed in 835 patients (Table 2). In 8 studies it was possible to split the results of SLNB in OCSCCs and OPSCCs, in 3 studies it was not possible to split the results of OCSCCs and OPSCCs, 9 studies only included oral cavity patients and 1 study included only OPSCC patients. As a result, 508 cases had OCSCCs, 131 cases had OPSCCs cancers and in 195 cases it was unclear if they had oral cavity or oropharyngeal cancer.

5

Figure 1: Flowchart of systematic search

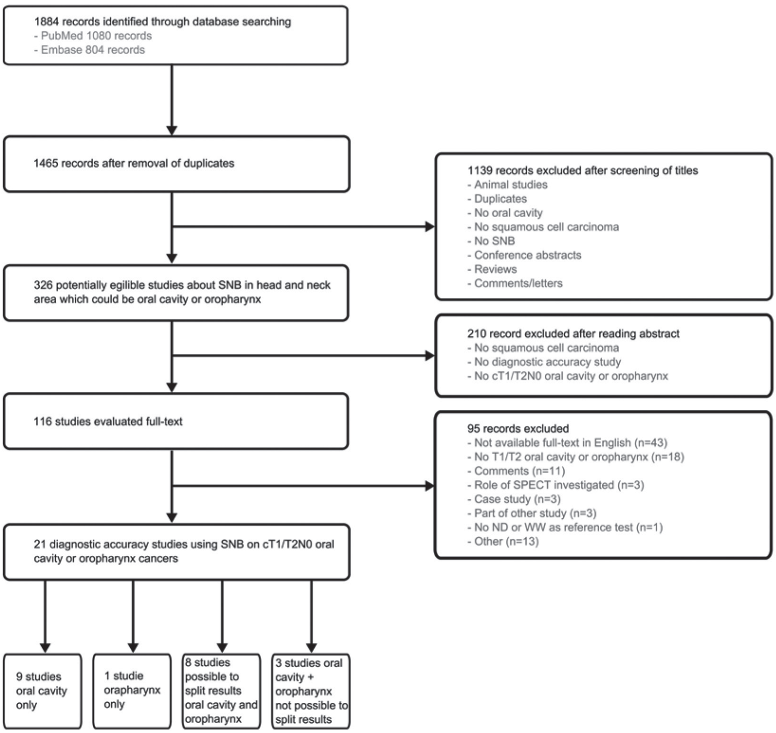


Table 1A: Study characteristics and protocols

Author	Year	Patient enrollment	Multi-center study	Study design	Primary location*	N0 determination	Index test†	Histology (of Sentinel nodes)‡	Reference test ND (type) or WW (mean follow-up time)#
Amezaga et al.	2007	Consecutive	Unclear	Prospective	Oral	CT + clinical exploration	L+G+B	H+I	ND (I-III or I-V)
Alkureishi et al.	2010	Unclear	Yes	Prospective	Oral+Orop	Varying between institutions	L+G+B	H+I	ND (Unclear)
Barzan et al.	2002	Unclear	Unclear	Prospective	Oral/Orop	CT+preoperative evaluation	L+G	Unclear	ND (Selective)
Broglie et al.	2012	Consecutive	Yes	Prospective	Orop	Palpation+CT+MRI+ USgFNAC	L+G (SPECT)	H+I	FU (25 months)
Burcia et al.	2009	Consecutive	No	Prospective	Oral+Orop	CT+Clinical	L+G	H+I	ND (Unclear)
Burns et al.	2009	Unclear	No	Prospective	Oral/Orop	Palpation+CT+MRI	L+G+B	H+I	ND (Unclear)
Civantos et al.	2006	Unclear	Unclear	Prospective	Oral	CT	L+G	H+I	ND (I-IV)
Civantos et al.	2010	Unclear	Yes	Prospective	Oral	Clinical	L+G	H+I	ND (I-IV)
Hart et al.	2005	Unclear	No	Prospective	Oral/Orop	CT+clinical	L+G	H+I	ND (Unclear)
Hoft et al.	2004	Unclear	Unclear	Prospective	Oral/Orop	Ultrasound	L+G	H+I	ND (Unclear)
Jeong et al.	2006	Consecutive	Unclear	Prospective	Oral	Palpation+CT	L+G	H+I	ND (I-III or I-IV)
Kontio et al.	2003	Consecutive	No	Prospective	Oral	Palpation+CT+MRI	L+G+B	H+I	ND (I-III or I-IV)
Melkane et al.	2012	Consecutive	No	Prospective	Oral	Clinical+ Radiological	L+G	H+I	ND (I-III)
Minimikawa et al.	2005	Unclear	No	Retrospective	Oral/Orop	Unclear	B	Unclear	ND (Unclear)
Pezier et al.	2012	Consecutive	Unclear	Prospective	Oral	Clinical+CT+MRI	L+G+B	H+I	FU (22.5 months)
Pitman et al.	2002	Unclear	Unclear	Prospective	Oral/Orop	Clinical	L+G	Unclear	ND (Selective)
Rigual et al.	2005	Consecutive	Unclear	Prospective	Oral	Palpation+CT	L+G+B	H	ND (I-III or I-IV or I-V)
Stoeckli et al.	2007	Consecutive	Unclear	Prospective	Oral/Orop	Palpation+CT	L+G	H+I	ND (I-III)
Taylor et al.	2001	Unclear	Unclear	Prospective	Oral/Orop	Clinical	L+G	Unclear	ND (I-IV)
Terada et al.	2010	Unclear	Unclear	Prospective	Oral	Clinical	L+G (SPECT)	H	FU (46 months)
Vorburger et al.	2012	Consecutive	No	Prospective	Oral+Orop	Palpation+CT+MRI+ USgFNAC	L+G (SPECT)	H+I	FU (40 months)

* Oral/Orop: separation in results between primary locations (Oral cavity/Oropharynx) possible , Oral+Orop: separation in results not possible

† L= Lymfoscintigraphy, G = gamma probe, B = blue dye substance, SPECT=Additional single photon emission computed tomography

‡ H = H&E staining , I = Immunohistochemisty

ND=neck dissection, FU=watchful waiting

Table 1B: Population characteristics

Author	Year	No. of patients	Age, mean (year)	Age, Range (year)	Gender distribution m/v (%)
Amezaga et al.	2007	15	65.9*	Unclear	68/32*
Alkureishi et al.	2010	55	Unclear	Unclear	Unclear
Barzan et al.	2002	12	64*	36-84*	77/23*
Broglie et al.	2012	111	58	28-91	64/36
Burcia et al.	2009	53	55.3 (median)	29-87	70/30
Burns et al.	2009	12	59.2*	38-80*	69/31*
Civantos et al.	2006	37	63 (median)*	29-89*	70/30*
Civantos et al.	2010	140	58 (median)	24-90	60/40
Hart et al.	2005	14	62.8*	35-80*	65/35*
Hoft et al.	2004	30	Unclear	Unclear	80/20
Jeong et al.	2006	20	53	35-68	75/25
Kontio et al.	2003	15	63.8	35-81	53/47
Melkane et al.	2012	53	56*	28-86*	71/29*
Minimikawa et al.	2005	15	60*	28-83*	65/35
Pezier et al.	2012	59	62.5 (median)	38-90	61/39
Pitman et al.	2002	12	55.9	24-81	Unclear
Rigual et al.	2005	20	Unclear	Unclear	60/40
Stoeckli et al.	2007	28	57	36-81	75/25
Taylor et al.	2001	9	61.9	22-80	44/56
Terada et al.	2010	45	62 (median)	30-85	67/33
Vorburger et al.	2012	92	58 (median)	28-84	63/29

*Characteristics are only known from the total population of the study which can include $>T_2$, N_+ or other primary sites than oral cavity or oropharynx. Characteristics are oral cavity + oropharynx if applicable

<u>Quality of included studies</u>

Risk of bias varied among the studies (Figure 2). Risk of bias regarding patient selection was high in seven (33%) studies mostly due to inappropriate exclusion of deeply invasive tumors. Risk of bias regarding index test and reference standard was unclear in 15 (71%) and 17 (81%) studies, respectively. In almost all these studies it was unclear if the index test and reference test were interpreted independently from each other. For risk of bias in flow and timing there were eleven (55%) studies with low risk and six (29%) studies with high risk. There was less concern about the applicability of the studies (Figure 3). Only in two (10%) studies there was concern about the applicability because of the patient selection, and in 1 (5%) study about the reference test .

5

Table 2: Detection rate

Author	Year	No. of patients	At least 1 SN	Detection rate (%)
Amezaga et al.	2007	15	14	93.3
Alkureishi et al.	2010	55	53	96.4
Barzan et al.	2002	12	11	91.7
Broglie et al.	2012	111	110	99.1
Burcia et al.	2009	53	50	94.3
Burns et al.	2009	12	12	100
Civantos et al.	2006	Unclear	37	Unclear
Civantos et al.	2010	140	140	100
Hart et al.	2005	14	14	100
Hoft et al.	2004	30	29	96.6
Jeong et al.	2006	20	20	100
Kontio et al.	2003	15	15	100
Melkane et al.	2012	53	53	100
Minimikawa et al.	2005	Unclear	15	Unclear
Pezier et al.	2012	59	57	96.6
Pitman et al.	2002	12	11	91.7
Rigual et al.	2005	20	20	100
Stoeckli et al.	2007	28	28	100
Taylor et al.	2001	9	9	100
Terada et al.	2012	Unclear	45	Unclear

Results are for oral cavity and oropharyngeal cancers if applicable

Diagnostic accuracy

The forest plot in Figure 4 shows the number of TPs, FPs, FNs, TNs and the sensitivity of all included studies. The pooled estimate of the sensitivity of SLNB was 0.93 [95% CI 0.90-0.95]. Negative predictive values were ranging from 0.88 to 1 (Table 3).

Subgroup analysis

Figure 5 shows the number of TPs, FPs, FNs, TNs and the sensitivity of the OCSCCs only. The pooled estimate of the sensitivity of SLNB in OCSCCs was 0.92 [0.86-0.95].
One study included oropharyngeal cancers only.[113] There were low number of OPSCCs in the studies in which it was possible to split the results for oral cavity and oropharyngeal cancers. In six of the eight studies for which it was possible to split OCSCC and OPSCC, only one or two patients had oropharyngeal cancer. These numbers were too low to calculate a pooled estimate for OPSCCs only.

Figure 2: Results of QUADAS-2

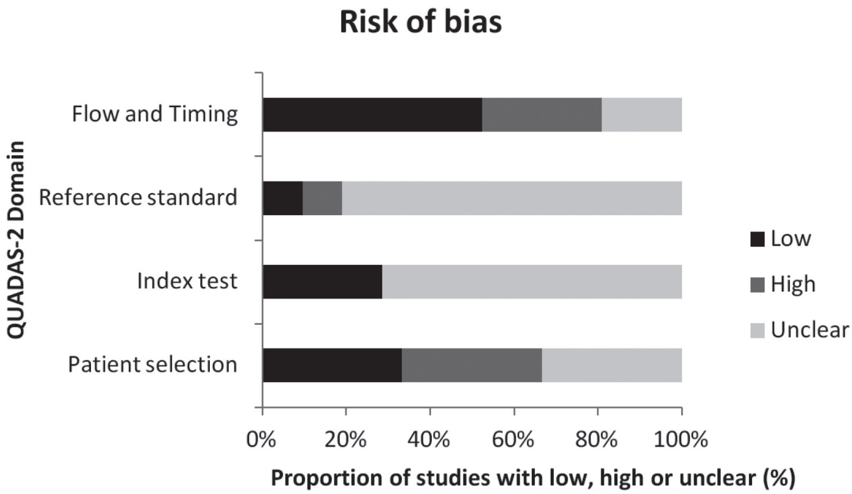
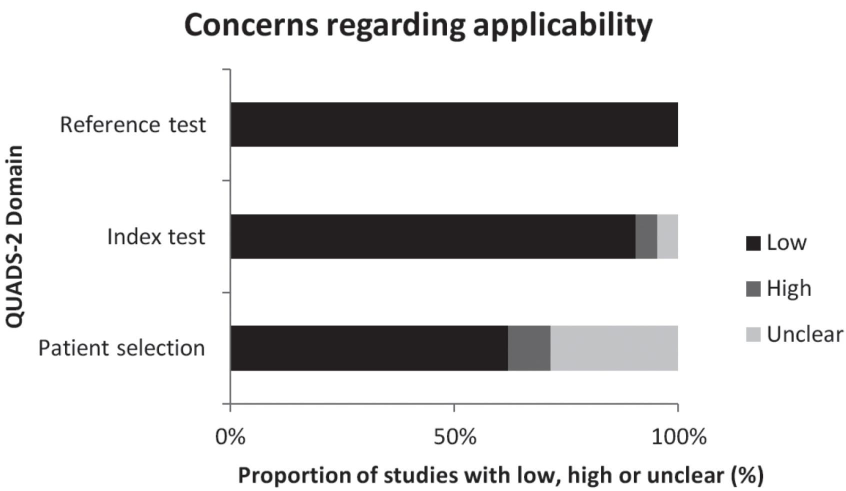


Figure 3: Results of QUADAS-2



In six studies the triple procedure was used with lymphoscintigraphy, blue dye substance and the use of a (handheld) gamma probe. One study used blue dye only. The other 14 studies did not use blue dye and did a procedure with lymphoscintigraphy and a gamma probe only. The procedure with lymphoscintigraphy and a gamma probe had a pooled sensitivity of 0.93 [0.91-0.98]. Triple procedure had a pooled sensitivity of 0.92 [0.77-0.96]. In 16 studies the results of a neck dissection were used as reference test, whereas the other four studies followed the patients to study whether the SLNB had predicted the neck status correctly. The pooled sensitivity of the studies with ND as reference test was

0.94 [0.90-0.97]. The pooled estimate of studies with follow-up as reference was 0.91 [0.84-0.95].

The pooled estimates of all studies and the different subgroups are shown in Figure 6. Subgroup analyses with the N_0 determination procedure or the reference test used was not possible since these were not described clearly enough.

Figure 4: Forest plot total, Oral cavity and Oropharyngeal cancers

Study	TP	FP	FN	TN	Sensitivity
Terada2010	7	0	3	35	0.70 [0.35, 0.93]
Barzan2002	3	0	1	7	0.75 [0.19, 0.99]
Kontio2003	3	0	1	11	0.75 [0.19, 0.99]
Rigual2005	10	0	2	8	0.83 [0.52, 0.98]
Minamikawa2005	5	0	1	9	0.83 [0.36, 1.00]
Civantos2010	37	0	4	99	0.90 [0.77, 0.97]
Vorburger2012	34	0	3	55	0.92 [0.78, 0.98]
Broglie2012	42	0	3	65	0.93 [0.82, 0.99]
Pezier2012	17	0	1	39	0.94 [0.73, 1.00]
Alkureishi2010	22	0	1	30	0.96 [0.78, 1.00]
Civantos2006	15	0	0	22	1.00 [0.78, 1.00]
Hoft2004	10	0	0	19	1.00 [0.69, 1.00]
Hart2005	2	0	0	12	1.00 [0.16, 1.00]
Burns2009	4	0	0	8	1.00 [0.40, 1.00]
Burcia2009	18	0	0	32	1.00 [0.81, 1.00]
Amezaga2007	3	0	0	11	1.00 [0.29, 1.00]
Taylor2001	4	0	0	5	1.00 [0.40, 1.00]
Pitman2002	2	0	0	9	1.00 [0.16, 1.00]
Stoeckli2007	9	0	0	19	1.00 [0.66, 1.00]
Jeong2006	6	0	0	14	1.00 [0.54, 1.00]
Melkane2012	12	0	0	41	1.00 [0.74, 1.00]

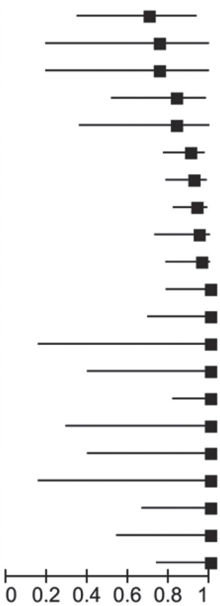


Figure 5: Forest plot oral cavity cancers

Study	TP	FP	FN	TN	Sensitivity
Terada2010	7	0	3	35	0.70 [0.35, 0.93]
Kontio2003	3	0	1	11	0.75 [0.19, 0.99]
Rigual2005	10	0	2	8	0.83 [0.52, 0.98]
minamikawa2005	5	0	1	8	0.83 [0.36, 1.00]
Civantos2010	37	0	4	99	0.90 [0.77, 0.97]
Pezier2012	17	0	1	39	0.94 [0.73, 1.00]
Pitman2002	2	0	0	8	1.00 [0.16, 1.00]
Taylor2001	3	0	0	5	1.00 [0.29, 1.00]
Stoeckli2007	6	0	0	19	1.00 [0.54, 1.00]
Hoft2004	5	0	0	13	1.00 [0.48, 1.00]
Jeong2006	6	0	0	14	1.00 [0.54, 1.00]
Hart2005	2	0	0	11	1.00 [0.16, 1.00]
Melkane2012	12	0	0	41	1.00 [0.74, 1.00]
Burns2009	3	0	0	7	1.00 [0.29, 1.00]
Civantos2006	15	0	0	22	1.00 [0.78, 1.00]
Amezaga2007	3	0	0	11	1.00 [0.29, 1.00]
Barzan2002	2	0	0	7	1.00 [0.16, 1.00]

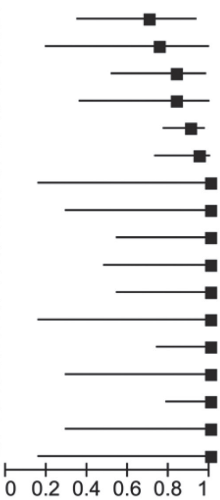
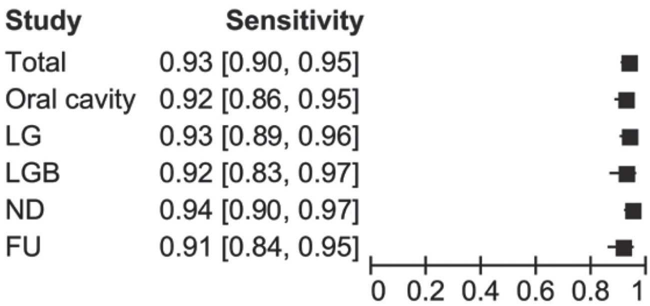


Table 3: Prevalence and negative predictive value

Author	Year	Prevalence of metastases(%)	NPV
Amezaga et al.	2007	27	1
Alkureishi et al.	2010	43	0.97
Barzan et al.	2002	36	0.88
Broglie et al.	2012	41	0.96
Burcia et al.	2009	36	1
Burns et al.	2009	33	1
Civantos et al.	2006	41	1
Civantos et al.	2010	29	0.96
Hart et al.	2005	14	1
Hoft et al.	2004	34	1
Jeong et al.	2006	30	1
Kontio et al.	2003	24	0.92
Melkane et al.	2012	23	1
Minimikawa et al.	2005	40	0.90
Pezier et al.	2012	32	0.98
Pitman et al.	2002	18	1
Rigual et al.	2005	60	0.80
Stoeckli et al.	2007	32	1
Taylor et al.	2001	44	1
Terada et al.	2010	22	0.92
Vorburger et al.	2012	40	0.95

Results are for oral cavity and oropharyngeal cancers if applicable

Figure 6: Forest plot pooled data



Total = oral cavity and oropharynx, Oral cavity = Oral cavity only,

LG= Lymfoscintigraphy + Gamma probe

LGB = Lymfoscintigraphy + Gamma probe + Blue dye substance

ND= Neck dissection as reference test

FU= Follow-up as reference test

Discussion

The results of this diagnostic meta-analysis demonstrate that sentinel node biopsy appears to be a sensitive method in the detection of neck metastases in cT_1/T_2N_0 OCSCC. In almost all patients at least one sentinel node was detected and therefore a biopsy could be taken which means that the procedure appears to be applicable as well.
These results are in line with a previous conducted meta-analysis which also showed a high sensitivity of SLNB in CT_1/T_2N_0 patients.[95] The major strengths of our study apart from being more up to date, is that we made a distinction in OCSCCs and OPSCCs and performed additional sub(group) analyses. OCSCCs and OPSCCs are often considered as one group under the heading of oral cancers. However they have different characteristics, which can cause different results for SLNB. Moreover, we performed a quality assessment on included studies.
However, some potential limitations should be discussed. First, the quality assessment showed that there is high risk of bias in patient selection in most studies, e.g. due to inappropriate exclusions of minimal invasive or small tumors. There was also some risk of bias in flow and timing because not all patients received the same reference standard and there was possible selection of patients in some studies. Second, the high sensitivity could only be ensured for oral cavity cancers. Only one study performed SNLB on a substantial amount of patients with oropharyngeal cancers. The low number of patients with oropharyngeal cancers in the other studies did not allow pooling the results of this type of tumor. Therefore, it is not certain that the high sensitivity of SLNB also applies for oropharyngeal cancers. Third, one study used blue dye only for detecting the sentinel nodes, which might be considered as an inappropriate method. A sensitivity analysis , in which we excluded this study, however, showed similar pooled outcomes. We therefore decided to include this study to provide a complete overview of the literature. Fourth, the methods used to classify the nodal status as N_0 clinically, were unclear in too many studies to allow subgroup analysis. However, it is expected that the method of N_0 determination influences only the percentages of occult metastases and therefore not the sensitivity. In most included studies the reference test was histopathological examination of a neck dissection specimen. Levels included in the neck dissection varied among the studies, but was at least level I-III. The probability of metastases skipping the first three levels is expected to be very low and therefore no significant differences are expected due to these variations in the reference test.
In conclusion, the high detection rate of the SLNB and the high sensitivity of the test justify an important role of SLNB in the diagnostic pathway of cT_1/T_2N_0 oral cavity squamous cell carcinoma patients.

Acknowledgements

The authors thank Maarten de Rooij of the department of Radiology, Radboud University Medical Centre, the Netherlands, and Hans Reitsma of the Julius Centre, University Medical Centre Utrecht, The Netherlands for their advice and help with the analyses.

Appendix: Search strategy

MEDLINE (Pubmed)

	Search item	**Fields**
	Head and neck neoplasm's	Mesh
OR	Mouth neoplasm's	Mesh
OR	Mouth	Mesh
OR	Oral	Title/Abstract
OR	Mouth	Title/Abstract

AND

	Search item	**Fields**
	Sentinel lymph node biopsy	Mesh
OR	Lymphoscintigraphy	Mesh
OR	Sentinel lymph node	Title/Abstract
OR	Sentinel node	Title/Abstract
OR	Lymph node excision	Title/Abstract
OR	SNB	Title/Abstract
OR	SNLB	Title/Abstract
OR	SLN	Title/Abstract
OR	SLNB	Title/Abstract
OR	Lymphoscintigraphy	Title/Abstract

EMBASE

	Search item	**Fields**
	Head and Neck	Title or abstract
OR	Mouth	Title or abstract
OR	Oral	Title or abstract

AND

	Search item	**Fields**
	Sentinel lymph node	Title or abstract
OR	Sentinel node	Title or abstract
OR	Lymph node excision	Title or abstract
OR	SNB	Title or abstract
OR	SNLB	Title or abstract
OR	SLN	Title or abstract
OR	SLNB	Title or abstract
OR	Lymphoscintigraphy	Title or abstract

Chapter 6

Management of the N_0 neck in early stage oral squamous cell cancer: A modeling study of the cost-effectiveness

Tim M Govers

Robert P Takes

M Baris Karakullukcu

Gerjon Hannink

Matthias AW Merkx

Janneke PC Grutters

Maroeska M Rovers

Published: Oral Oncology 2013, Aug; 49 (8): 771-7

Abstract

Objectives: To assess the cost-effectiveness of five strategies for diagnosing and treating $cT_{1\text{-}2}N_0$ oral cavity squamous cell cancer.

Materials and methods: A Markov decision analytic model was used to evaluate the cost-effectiveness of 1) elective neck dissection (END), 2) watchful waiting (WW), 3) gene expression profiling (GEP) followed by neck dissection (ND) or WW, 4) sentinel lymph node (SLN) procedure followed by ND or WW, and 5) GEP and SLN (for positive GEP) followed by ND or WW. Uncertainty was addressed using one-way and probabilistic sensitivity analyses

Results: Base-case analysis showed that SLN procedure followed by ND or WW was the most effective and most cost-effective strategy. Compared with direct END the incremental cost-effectiveness ratio was €3,356 per QALY gained. Uncertainty analysis showed that the model was sensitive to changes in assumed occult metastases incidence and utility values. SLN was found to have the highest probability (66%) of being cost-effective of the five strategies, at a willingness to pay of €80,000 per QALY.

Conclusion: Given the current evidence and costs the SLN procedure followed by ND or WW appears to be the most cost-effective strategy for diagnosing and treating oral cavity squamous cell cancer patients. Our model provides the foundation for future diagnostic and therapeutic research in this field and shows that further information on quality of life in this population is highly valuable.

Introduction

Head and neck squamous cell cancer (HNSCC) is considered an important part of the global burden of cancer, mainly due to the widespread use of tobacco and alcohol.[61] The oral cavity is the most predominant location in the head and neck region for primary malignant tumours, and more than 90% consists of squamous cell carcinoma (SCC). The primary treatment of these tumours is surgery in most cases. For patients with oral cavity SCC (OCSCC) prognosis drops to nearly half when regional lymph nodes are involved.[14] Since regional metastases occur relatively frequently, adequate diagnosis and treatment of the neck are important besides optimal local tumour control. If nodal metastases are detected, therapeutic neck dissection as an integral part of regional treatment is beyond debate, to be followed by postoperative radiotherapy, with or without chemotherapy on indication. There is, however, still vigorous debate on how to manage the neck in patients without clinical evidence of metastases. The question is whether elective neck dissection (END), which usually consists of removal of lymph nodes of level I-III, is indicated in these patients.

In most centres patients with clinically uninvolved (cN_0) necks undergo an elective neck dissection if the known incidence of clinically occult metastases is greater than 20%. This implies a possible overtreatment, including its treatment associated morbidity and impact on quality of life, of up to 80% in this specific group. The alternative for END is a watchful waiting (WW) strategy, which may result in undertreatment.[30] The prognostic impact of such a WW policy versus elective surgery is influenced by the delay in detection and treatment of any neck metastases.[39]

Decision analyses by Weiss et al. and Okura et al. suggested thresholds of 20% and 44% of occult metastasis as cut off in the decision to operate or observe.[114,115] Neither of them took into account the costs in relation to health outcomes (cost-effectiveness) and both are dated. New diagnostic techniques have been introduced, such as sentinel lymph node (SLN) procedures and gene expression profiling (GEP), which might add to better prediction of occult metastasis.[116] The sentinel node concept is based on the assumption that tumour will spread from the primary site to a single node or group of nodes (the sentinel nodes), before progressing to the remainder of the lymph nodes. Histopathological evaluation of sentinel lymph nodes may allow accurate prediction of the disease status. Although SLN biopsy is still an invasive procedure gives less morbidity than a neck dissection (ND) by sparing relevant structures.[63] GEP is a test to distinguish metastasizing cancers from non metastasizing cancers on basis of the expression of a set of genes of the primary tumour. Cost and morbidity could be reduced if one or a combination of these novel diagnostic techniques could avoid unnecessary elective neck dissections in cN_0 oral cancer patients.[117] An updated decision model, which takes into account these factors, is therefore warranted.

We developed a new decision model addressing the important decisions for current management of the cN_0 neck, considering both cost and effects. With this model, we aim to make an important step to an evidence based decision for oral cavity cancer patients regarding possible regional metastases by determining the most cost-effective strategy of current relevant options.

Materials & Methods

Population

The target population consists of patients with clinical $T_{1\text{-}2}N_0$ OCSCC. cN_0 was classified by palpation, computed tomography (CT) and/or Magnetic resonance Imaging (MRI) and ultrasound-guided fine needle aspiration cytology (USgFNAC).

Decision model

We developed a decision-analytic Markov model to evaluate the cost-effectiveness of five different diagnostic and treatment strategies for possible occult regional metastases in $cT_{1\text{-}2}N_0$ patients with OCSCC. A decision model is a mathematical method to weigh risks, benefits, patient preferences, and costs of clinical strategies.[118]

A decision tree was used to model the diagnostic pathways, and a Markov model represented the subsequent follow-up. The strategies were: 1) elective neck dissection (END), 2) watchful waiting (WW), 3) gene expression profiling (GEP) followed by neck dissection (ND) or WW, 4) sentinel lymph node (SLN) procedure followed by ND or WW, and 5) GEP and SLN (for positive GEP) followed by ND or WW. The created clinical settings are based on published clinical guidelines, published diagnostic and intervention studies, and expert opinion.[119] Adjuvant radiotherapy on the primary tumour (which may involve part of the adjacent lymph nodes) is not explicitly included in the model as it is assumed to occur independent of the different strategies of the neck.

In the model a hypothetical cohort of OCSCC patients moves between health states according to a set of transition probabilities. After each cycle (i.e. one year) patients in the cohort have a probability to move to different health states. In each health state, patients are assigned the corresponding costs and outcomes. This is modelled for a total of five cycles of one year each, representing five years.

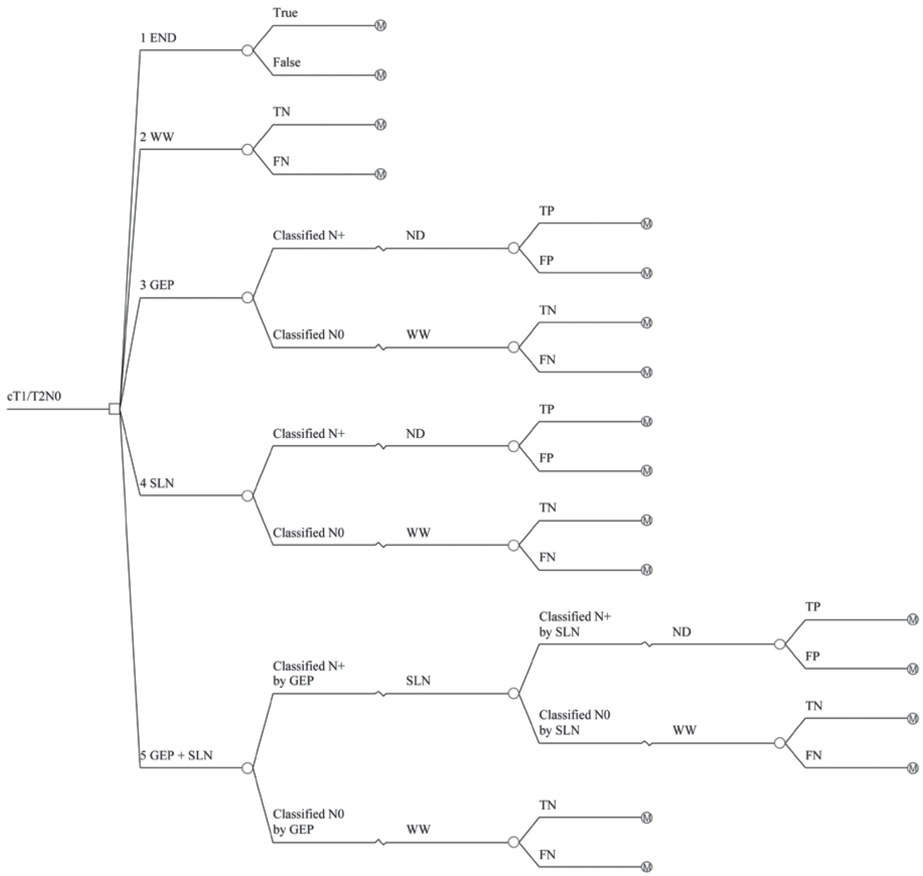
Model structure

The diagnostic pathways of the five strategies are shown in Figure 1. Direct END (strategy I) or direct WW (strategy II) for all patients is considered to be correct or incorrect, true or false. In the END strategy ‘true’ means detected nodal metastasis on histopathological examination. In the WW strategy ‘true’ means no detectable metastases (regional failure) in follow up. Those patients classified as “N_+” by the additional tests (strategy III-V), which can either be false positive (FP) or true positive (TP), are all treated with a ND.

Those classified as “N_0”, which can be either False Negative (FN) or True Negative (TN), all undergo WW. After these classifications and corresponding treatment, patients move to the different health states in the Markov part of the model which represents the follow-up of patients.

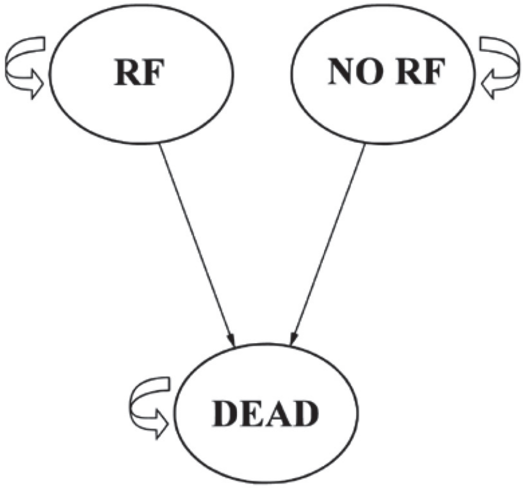
Possible health states for patients were: regional failure (RF), no regional failure, or dead. The structure of the Markov part of the model was the same for each diagnostic outcome (Figure 2). However, the probabilities of entering the different states (i.e. the chance of RF) were different among the diagnostic outcomes.

Figure 1: Model structure for the initial diagnostic and treatment pathway



The decision node is shown as the open square, chance nodes as open circles and Markov nodes as open circles with M.

Figure 2: Influence diagram of Markov part of the model



The Markov part of the model represents follow-up of patients after diagnostic outcome and corresponding treatment. Patients that experience regional failure (RF) enter the Markov part of the model in the RF state and the other patients in the no RF state. Patients can either survive and stay in the state they entered in or die and go to the dead state. The no RF health state is different after WW, SLN procedure and ND. The RF health state is independent of previous treatment.

Transitions

An overview of the probabilities used in the model is shown in Table 1. The diagnostic accuracy data used for transition probabilities in strategies with GEP were derived from a recent Dutch multicentre study.[116] The accuracy data of SLN biopsies were derived from a meta-analysis of 13 studies which we performed alongside this study.[88,95,96,98-107] Data on percentage of occult metastases, regional failure probability and survival data after regional failure and no regional failure of patients that underwent a neck dissection were derived from the eight Dutch head and neck oncological centres. These results were based on 96 patients. WW outcomes regarding regional failure probability and survival data were derived from one centre, where WW is the standard for all $cT_{1\text{-}2}N_0$ patients. This was based on 69 patients. All cause mortality data were analyzed with Kaplan-Meier methods.

Table 1: Transition probabilities

Parameter	**Point estimate**	**SE**	**Source**
Sensitivity gene expression	0.86	0.07	Van Hooff et al.[10]
Specificity gene expression	0.45	0.06	Van Hooff et al.[10]
Sensitivity sentinel node	0.93	0.04	Govers et al. [no ref]
Specificity sentinel node	1	Fixed	Govers et al. [no ref]
Percentage occult metastasis	0.39	0.05	Empirical study with 96 patients
Regional failure after END with occult metastases (TP)	0.42	0.08	Empirical study with 96 patients
Regional failure after WW with occult metastases (FN)	1	Fixed	Empirical study with 69 patients
Survival with regional failure		Fixed	Empirical study with 30 patients
Year 1	0.80		
Year 2	0.78		
Year 3	0.93		
Year 4	0.83		
Year 5	1		
Survival without regional failure		Fixed	Empirical study with 135 patients
Year 1	0.95		
Year 2	0.90		
Year 3	0.93		
Year 4	0.91		
Year 5	1		

Cost information

Cost analysis was performed from a health care perspective. Unit costs of neck dissection and SLN procedure were calculated according to the Dutch pharmaco-economic guidelines using available sources from the department of Otorhinolaryngology and Head and Neck Surgery of the Radboud University Nijmegen Medical Centre (RUNMC). Costs of GEP were obtained from Agendia BV (Amsterdam, The Netherlands).[120] The volume of hospital days and medical specialist hours were collected from existing registries of the RUNMC and were multiplied by reference prices from the Dutch pharmaco-economic guideline.[120]
In each strategy a percentage of patients will experience regional failure. It was assumed that regional failure was followed by salvage therapy with the costs of a (modified) radical neck dissection. No differences in costs were expected for follow-up of patients between

the strategies and these costs were not included in the analysis. Also no differences were expected for the number of hospital days in the base case analysis since this is mainly determined by surgery of the primary tumour. However, the costs of the hospital days were included to be able to vary them in the deterministic sensitivity analysis and to give an indication of the total diagnostic and treatment costs of each strategy. The costs of 11.8 hospital days were, therefore, included in each strategy. Costs were based on the year 2011, and future costs were discounted to their present value by a rate of 4%.[121] An overview of the cost data used in our model is provided in Table 2.

Table 2: Costs

Parameter	Resource use	Unit price	Costs (€)	Sources (Unit Price)
Neck dissection				
-OR time + materials	2.5 hours	650	-1,625	-Unit cost calculation, Department of Otorhinolaryngology (RUNMC)
-Medical specialist	2.5 hours	116	-290	-Pharmaco-economic guidelines[25]
Total			**1,915**	
Salvage therapy				
-OR time + materials	4 hours		-2,665	-Unit cost calculation, Department of Otorhinolaryngology (RUNMC)
-4 hours medical specialist	4 hours		-412	-Pharmaco-economic guidelines[25]
Total			**3,077**	
Gene expression profile	1	2,675*	**2,675**	-Agendia BV
Sentinel lymph node procedure				
-hour OR time + materials	1	650	-650	-Unit cost calculation, Department of Otorhinolaryngology (RUNMC)
-hour medical specialst	1	103	-103	-Pharmaco-economic guidelines
-Nuclear materials + services	1	210	-210	-Department of nuclear medicine[25] (RUNMC)
-day treatment	1	238	-238	-Pharmaco-economic guidelines[25]
Total			**1,201**	
Hospital days				
- Hospital days (included in each strategy)	11.8 days	575	**6,785**	-Pharmaco-economic guidelines[25]

Resource use was determined using data of the Radboud University Nijmegen Medical Centre (RUNMC)

<u>Outcome measures</u>

Health-related quality of life was used as outcome measure in the model. It was considered as a single index utility, on a scale from 0 (representing death) to 1 (representing perfect health).[122] The use of utility scores allows the calculation of Quality Adjusted Life Years (QALYs) and costs per QALY. We used the utility data from the decision model of Weiss et al. in our model. These utilities are presented as disutility (reduction of quality of life) in relation to WW patients without regional failure (to whom a utility of 1 was assigned).[114] These values were derived from expert consultation. We assumed a similar disutility for patients who underwent ND after SLN and those who only underwent ND. The disutility of patients with a WW strategy after the SLN procedure was assumed to be half of the disutility of ND. GEP was assumed to have no influence on quality of life. When regional failure occurred the disutility was assumed to be independent of previous strategy because of complete neck dissection after regional failure. An overview of (dis)utilities for each health state is provided in Table 3. Effects were discounted with a constant rate of 1.5% according to the pharmacoeconomic guidelines.[121]

Table 3: Utility values

Parameter	Point estimate	SE	Source
Utility no regional failure; after WW	1	fixed	Weiss et al.[5]
Disutility no regional failure; after END	0.03	0.015	Weiss et al.[5]
Disutility no regional failure; after sentinel node followed by WW	0.015	0.015	Assumption based on Weiss et al.[5]
Disutility regional failure	0.06	0.030	Weiss et al.[5]

<u>Data analysis</u>
Strategies were compared in terms of mean costs, mean effects (in QALYs) and incremental cost-effectiveness ratios (ICERs). ICERs are calculated by dividing the additional costs of one treatment over the other by the additional QALYs. They thereby represent the extra costs that are needed to gain one QALY. Whether a strategy is deemed cost-effective depends on the willingness-to-pay (WTP) for a QALY. A WTP threshold of €80,000 per QALY was used, as recommended by the Dutch Council for Public Health and Care.[123] This means that a strategy is deemed cost-effective compared to another strategy when it costs €80,000 or less to gain an extra QALY, i.e. the ICER is lower than €80,000 per QALY. ICERs were calculated compared with the elective neck dissection strategy, and in a full incremental analysis. For the incremental analysis we ordered the strategies from lowest (A) to highest costs (E). If strategy B was more cost-effective than strategy A, we used strategy B as the comparator for strategy C. If strategy A was more cost-effective, strategy A remained the comparator for strategy C. We completed the incremental analysis by repeating this action for strategy D and E.

We performed one-way deterministic sensitivity analyses on the assumptions made regarding the diagnostic accuracy of SLN and GEP, percentage of occult metastasis, costs, probability of regional failure after ND, and the difference in quality of life (after SLN, ND and WW). We changed these parameters over a range of values to study the potential consequences on the outcome of the model.

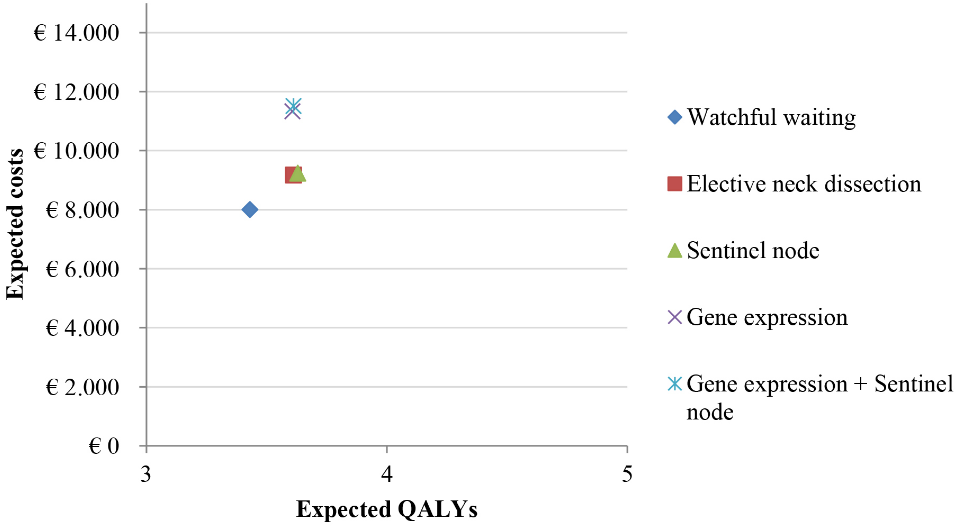
In addition, probabilistic sensitivity analysis was performed to assess the impact of all uncertainty on the outcomes. This means that we assigned distributions to model parameters, to reflect the uncertainty in the estimation of that parameter when possible. All distributions are listed in Table 1 and 3. Cost parameters were not included in the probabilistic sensitivity analysis. Parameter values were drawn at random from the assigned distributions, using Monte Carlo simulation with 10,000 iterations. To illustrate the results of the simulation, cost-effectiveness acceptability curves (CEACs) were calculated. CEACs show the probability that a strategy is cost-effective for a range of WTP thresholds.
As uncertainty exists, there is always a chance that the 'wrong' decision will be made. The expected value of perfect information (EVPI) reflects the expected costs of uncertainty relating to all input parameters in the decision model, based on the probability of making an incorrect decision and its consequences. We calculated the EVPI to measure the value of acquiring additional empirical information per patient.[124] The population EVPI was calculated by multiplying this value by the effective population, which is based on the incidence of the disease. Also the expected value of partial perfect information (EVPPI) was calculated. The EVPPI represents the expected costs of uncertainty for one or a group of parameters. This shows which (groups of) parameters are most valuable for further research.

Results

<u>Base-case analysis</u>

WW was the least effective strategy with 3.43 expected QALYs while SLN procedure was the most effective strategy with 3.63 expected QALYs. WW was also the least costly strategy with on average €8,003 in costs per patient. The average costs for WW patients consist of hospital days related to the treatment of the primary tumor, and salvage therapy for regional failure for a percentage of patients. GEP + SLN procedure was the most expensive strategy with €11,515. The baseline deterministic results of the five strategies are presented in Table 4a and in Figure 3. SLN procedure was the most cost-effective strategy in the incremental analysis (Figure 4a). The costs of the SLN strategy were slightly higher than for the END strategy (€9,241 versus €9,180, respectively), whereas the corresponding QALYs were 3.63 and 3.61. The ICER was €3,356 per QALY gained. All other strategies were also compared to END, the results are presented in Table 4b.

Figure 3: Baseline results of the five strategies



Expected mean costs and QALYs per patient over five years

Table 4a: Incremental analysis

Strategy	Costs (€)	Effects (QALY)	Comparator	Incremental costs (€)	Incremental effects (QALY)	ICER (€/QALY)	Outcome
Watchful waiting	8,003	3.4296	x	x	x	x	x
Elective neck dissection	9,180	3.6108	WW	1,177	0.1812	6,493	Cost-effective
Sentinel node	9,241	3.6291	END	61	0.0183	3,356	Cost-effective
Gene expression	11,335	3.6068	Sentinel node	2,094	-0.0223	x	Dominated
Gene expression + Sentinel node	11,515	3.6114	Sentinel node	2,274	-0.0177	x	Dominated

6

Table 4b: ICERs compared to END

Strategy	Costs (€)	Effects (QALY)	Incremental costs (€)	Incremental effects (QALY)	ICER (€/QALY)	Outcome
Elective neck dissection	9,180	3.6108	x	x	x	x
Watchful waiting	8,003	3.4296	-1,177	-0.1812	-6,493	Not cost-effective
Sentinel node	9,241	3.6291	61	0.0183	3,356	Cost-effective
Gene expression	11,335	3.6068	2,155	-0.004	412,933	Dominated
Gene expression + Sentinel node	11,515	3.6114	2,335	0,0006	3,892,287	Not cost-effective

Deterministic sensitivity analysis

Deterministic sensitivity analysis showed that variations in the diagnostic accuracy, costs and regional failure rate after ND had no significant effect on the results, assuming realistic variations for this parameters. The results of this analysis showed that the results are sensitive to variations in the percentage of occult metastasis and quality of life.

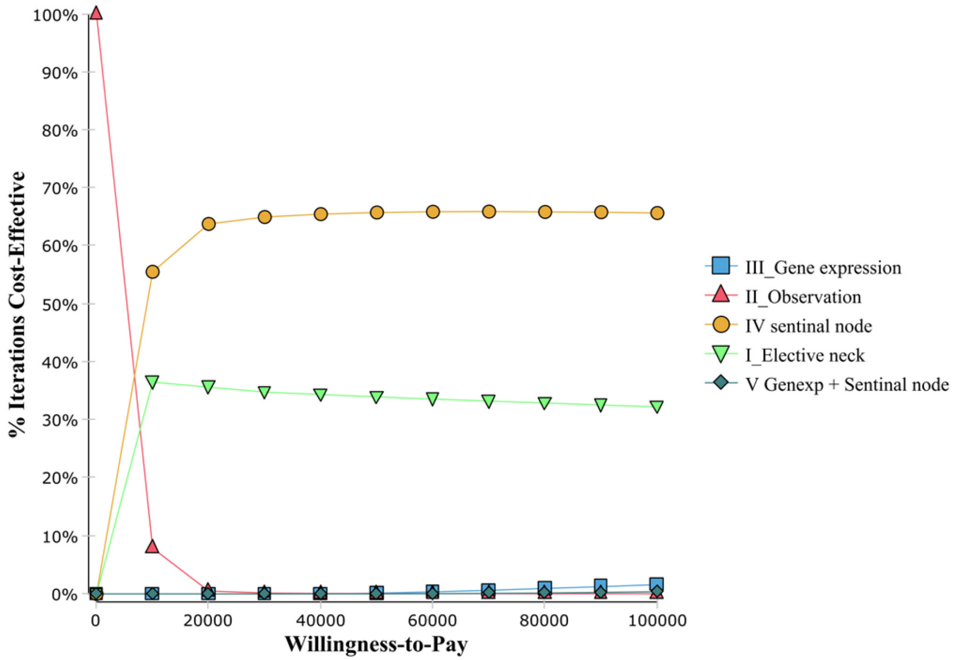
SLN was the most cost-effective strategy when we modelled percentages of occult metastasis between about 11 and 53%. When we modelled percentages above 54%, END was the most cost-effective strategy, and with percentages of 11% or lower, WW was most cost-effective.

The outcome of the model also changed when we changed the disutilities of patients who underwent ND and those assuming to undergo a SLN procedure, relative to the utility of patients with a WW strategy. Depending on the utility for the health state without regional failure following ND, WW (utility No RF after ND lower than 0.80), GEP followed by SLN (utility No RF after ND between 0.80 and 0.87), SLN strategy (utility No RF after ND between 0.88 and 0.98) and END alone (utility No RF after ND higher than 0.98) were found to be the most cost-effective strategy respectively. In these situations, the health state without regional failure after SLN always had a utility of half that of no regional failure after ND.

Probabilistic sensitivity analysis

By performing probabilistic sensitivity analysis, it was possible to determine the optimal strategy for a range of values regarding willingness to pay for a QALY gained given the existing uncertainty. Figure 4 presents the cost-effectiveness acceptability curve (CEAC) for all strategies. With a WTP threshold of €80,000 per QALY, SLN procedure and END were cost-effective in 66%, and 33% of the simulations, respectively. Above a WTP threshold of €7,500/QALY SLN procedure appears to be the most cost-effective strategy. With a WTP threshold of €7,500 or less, WW had the highest probability of being cost-effective. The value-of-information analysis demonstrated an EVPI of €997 per patient. About 350 patients are diagnosed with a cT_1/T_2N_0 OCSCC per year in the Netherlands. Over five years the discounted population EVPI will be 486 000 Euros, i.e. the expected costs for uncertainty relating to all input parameters in the decision model for all cT_1/T_2N_0 OCSCC patients in the Netherlands over 5-year. The EVPPI of utility values only was the highest with €780 per patient.

Figure 4: Acceptability curve for the five strategies



This figure presents in which percentage of simulations the strategies were cost-effective for a certain WTP

Discussion

The results of our study show the potential cost-effectiveness of five diagnostic and treatment strategies for patients with $cT_{1\text{-}2}N_0$ OCSCC. SLN strategy was found to be on average more effective than WW and slightly more effective than the other strategies. The extra effectiveness of the SLN strategy was about 0.2 QALYs compared with WW and approximately 0.02 QALYs compared with the other strategies over the first five years. This is equivalent to an increase of 2.5 months in full health compared with WW and one week extra in full health compared with the other strategies. Costs of the SLN strategy were somewhat higher than for the END strategy, but were lower than for strategies with GEP. The slightly higher effectiveness of SLN compared with END, and the slightly higher costs of SLN, resulted in a favorable incremental cost-effectiveness ratio for the SLN strategy. Consequently, on the basis of currently available evidence, SLN appears to be the most cost effective strategy.

Our results follow on previous studies of Weiss et al. and Okura et al. whom reported that END was more effective than WW above thresholds of 20% and 44% of occult metastasis, respectively. However, both studies did not take into account the costs. Furthermore, we studied two new diagnostic techniques, i.e. SLN procedure and GEP, which have not been studied before in a decision analysis.

Some potential limitations should also be discussed. First, our sensitivity analysis showed that the percentage of occult metastasis had some impact on the outcome. However, the SLN procedure was most cost-effective within the range of 23% to 43% reported in literature.[86] Second, because utility values were not available for all health states, we had to make several assumptions regarding (dis)utility values in the different health states. We assumed that patients who underwent a SLN procedure, followed by WW had a utility between that of ND patients without regional failure and WW patients without regional failure. We also assumed that utility values of patients with regional failure were independent of previous treatment and were the lowest. Also, since the utilities were not collected from OCSCC patients themselves but by expert opinion, larger differences in utilities after ND, SLN procedure and WW may exist than the small differences included in the model. Our deterministic sensitivity analyses showed the impact of quality of life on the outcome in patients and the optimal strategy. It showed that a relatively larger difference in this strategy associated quality of life will favour GEP in combination with a SLN procedure (strategy 5), or when the difference is very large, WW alone will become the most cost-effective strategy. Third, we built a model with a time horizon of five years. Although differences in survival and quality of life narrow over time, a life-time model could have resulted in different outcomes due to possible life-time shoulder morbidity. Therefore, this choice of time horizon can be deemed a conservative approach. Fourth, the survival after regional failure was assumed independent of the initial treatment to prevent possible confounding by indication. It is possible that this assumption overestimates the survival of patients after ND. Fifth, the input for this model is based on the Dutch situation using costs in one centre. As costs in particular can vary between settings, the question may arise whether these costs are applicable in other centres and countries. Sensitivity analyses showed that costs could vary considerably without changing the outcome, i.e. SLN strategy was the most cost-effective strategy even when we doubled the costs of an SLN procedure.

Despite these limitations our model provides clinically relevant information. Our model demonstrates that a SLN procedure is a valuable step in the diagnostic pathway of OCSCC patients since it will improve the outcome for patients without significant extra costs for healthcare by preventing unnecessary burdensome END. It is important to realize that the impact of strategies on the patients quality of life (utility) is important in the decision making process. We expect that the disutilities after ND and SLN procedure, used in the model now, turn out to be different when obtained from the OCSCC population themselves. This could mean that SLN procedure could even be more effective but possibly also that GEP should be performed before making a choice for a SLN procedure, to avoid unnecessary sentinel node biopsies as well. For example, if GEP can be performed at lower costs in the near future, the combined procedure (i.e. GEP and SLN) might become the most cost-effective strategy.

The EVPI showed that future empirical research is warranted to reduce uncertainty of the outcome. Further research was found to be most valuable for the utility values, the EVPPI (expected costs of uncertainty for a single group of parameters) of the utility values showed that about 80% of the total expected costs were caused by uncertainty in utility values. Therefore, it is worthwhile to measure quality of life in patients who underwent ND, WW or a SLN procedure first. Second, survival data should be gathered in a group of patients with a (more) random selection of treatment (WW or END). In this way we could use different survival after regional failure after WW or ND when required.

In conclusion, given the current evidence and cost the SLN procedure followed by a neck dissection or watchful waiting appears to be the most cost effective strategy for diagnosing and treating patients with $cT_{1-2}N_0$ oral cavity squamous cell cancer. Our model provides the foundation for future diagnostic and therapeutic research in this field and shows that further information on quality of life in this population is highly valuable.

Chapter 7

Quality of life after different procedures for regional control in oral cancer patients: Cross-sectional survey

Tim M Govers

Willem H Schreuder

W Martin Klop

Janneke PC Grutters

Maroeska M Rovers

Matthias AW Merkx

Robert P Takes

Published: Clinical Otolaryngology 2015, July

Abstract

Objectives: To examine health utilities in patients with cT_{1-2} oral cavity squamous cell carcinoma following different diagnostic and treatment modalities for the neck, and to investigate the relation between shoulder morbidity and health utility.

Design: Cross-sectional survey

Setting: Two Dutch hospitals

Participants: Four subgroups of patients with oral cavity cancer who underwent watchful waiting, sentinel lymph node biopsy, elective supraomohyoid neck dissection, or therapeutic modified radical neck dissection

Main outcome measures: Patients received the EuroQol-5D-3L questionnaire and the shoulder disability questionnaire. Mean health utility, visual analogue scale, and shoulder disability scores were calculated.

Results: 181 patients (62%) returned the questionnaires. Mean health utilities, adjusted for age, gender and time since treatment, were 0.804, 0.863, 0.834 and 0.794 for the watchful waiting, sentinel lymph node biopsy, supraomohyoid neck dissection and modified radical neck dissection subgroups, respectively. Mean shoulder disability scores (higher score means more shoulder complaints) for these subgroups were 8.64, 10.57, 18.92 and 33.66. Patients with shoulder complaints had a mean utility of 0.78 while patients without shoulder complaints had a mean utility of 0.90.

Conclusions: This study shows that more invasive procedures appear to result in lower health utility. The high health utility for patients after sentinel lymph node biopsy supports a role for this procedure in oral cancer patients.

Introduction

Effects of cancer treatment on Health-related quality of life (HRQoL) are gaining importance next to oncological effectiveness. Besides this, armamentarium for the diagnosis and treatment of cancer is expanding. These two factors make decision making around the optimal treatment strategy more complex. Decision analysis is very useful in this perspective. In these analyses, quality-adjusted life years (QALYs) are the preferred health outcome because they combine HRQL with oncological outcome.[122] To calculate QALYs it is necessary to express HRQoL as a value that is anchored on a numeric scale ranging from death (0) to perfect health (1), the so-called utility score.[125]
In the diagnosis and treatment of lymph nodes in the neck in oral cavity squamous cell cancer (OCSCC) patients, the trade-off between HRQoL and oncological outcome is important for decision making. More invasive procedures are more likely to be oncologically effective. However, they are also more likely to have a negative impact on HRQoL. Decrease in HRQoL in these patients is mostly caused by shoulder complaints that are associated with neck treatment.[18,19] Other studies have confirmed that more invasive treatment indeed results in more shoulder complaints.[23,25,126] However, it is unclear what the impact of these shoulder complaints is on health utilities. Therefore, the results of these studies cannot be used to calculate QALYs and are not useful for decision analyses. The importance of health utilities after different treatment modalities for choosing the optimal treatment modality for early stage oral cavity cancer patient with a clinical negative neck is also confirmed in our recent decision analytical model.[127]
Therefore, the aims of this study are: 1) to examine health utility in patients with cT_{1-2} oral cavity cancer following different diagnostic and treatment modalities and 2) to investigate the relation between shoulder morbidity and health utility.

Patients and Methods

Ethical considerations

The study was approved by the institutional review boards of the participating centers.

Study design and patient enrollment

We performed a cross-sectional survey of patients with early stage (T_{1-2}) OCSCC in the Radboud University Medical centre (Radboud UMC) and the Antoni van Leeuwenhoek hospital Amsterdam (AvL) between 2001 and 2013. Four subgroups of patients who underwent different diagnostic and treatment modalities for the neck were selected: watchful waiting (WW), sentinel lymph node biopsy (SLNB), supraomohyoid neck dissection (SOHND) and modified radical neck dissection (MRND). MRND was performed in patients with node positive necks (detected by ultra sound guided fine needle aspiration cytology, SLNB or in a few cases after regional failure). The other three modalities were performed, as standard, in patients with cN_0 necks for a certain period in one of the centers. The SLNB group consisted of patients who had a negative sentinel node and did not receive any additional treatment for the neck. For this study, the WW group could be seen as a reference to show the influence of the additional treatment of the neck in these patients. Electronic patient databases were used to gather information about the patients. The following data were retrieved from the medical records: date and type of treatment, tumor stage, tumor subsite, date of birth, gender, previous mucosal head and neck tumors, second primary tumors, local recurrences and the administration of (chemo-)radiotherapy in the head and neck region.

The EuroQol-5D questionnaire (EQ-5D-3L) and the shoulder disability questionnaire (SDQ) were sent to the selected patients guided by a letter containing informed consent and instructions. We did not send reminders to the patients because informed consent stated that when they did not want to participate, patients did not have to respond to the questionnaire.

Questionnaires

The EQ-5D-3L is the most frequently used multi-attribute health status classification system and is recommended by the National Institute for Health and Clinical Excellence.[128] The EQ-5D-3L consists of five dimensions of generic HRQL (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) with three answer possibilities per dimension (severe problems, some problems, no problems).[129] The answers to these questions can be combined to calculate health utilities using a scoring function, i.e. for each answer that represents a problem in one of the dimensions, the health utility of a patient is lowered with a certain amount.[122] This scoring function is determined by the general population by valuing the possible health states (each possible combination of answers represent a health state). Scoring functions vary among countries depending on the preferences of

that particular population. These scoring functions for countries are called tariffs. Health utilities range from a negative score (health state worse than death, severe problems in all five dimensions), through 0 (death) to 1 (full health, no problems in all dimensions). Next to the five questions, a visual analogue scale (VAS) is included in the EQ-5D that ranges from 0 (worst imaginable health state) to 100 (best imaginable health state). The VAS measures the patients' self-rated HRQoL (patients' perspective).
The SDQ questionnaire consists of 16 items and is designed to evaluate functional status limitation in patients with shoulder disorders.[130] The SDQ score was calculated as the ratio of the number of items with an affirmative answer over the number of applicable items and was multiplied by 100.

<u>Analyses</u>
Health utilities were calculated from the EQ-5D-3L outcomes, using the Dutch tariff.[131]
We performed analysis of covariance (ANCOVA) to calculate mean health utility, SDQ and VAS scores for the different procedures. HRQL scores are likely to decrease with increasing age, tend to be lower for females and can change after treatment due to recovery. [132-134] Therefore, we adjusted for potential confounding by age (at time of questionnaire), gender and time since treatment. Time since treatment was calculated as time since surgery of the primary tumor for the WW group. For the SOHND, MRND and SLNB groups this was time since the neck dissection or sentinel node procedure, which generally corresponds to the time since surgery of the primary tumor, since these are mostly performed during the same procedure. The analysis was repeated for the group of patients without any previous mucosal malignancy in the head and neck region, local recurrence or second primary tumor because this can influence the choice for a certain strategy. The analysis was repeated, a third time, in which also patients who received any (chemo)radiotherapy in the head and neck region were excluded.
We used correlation to assess the relation between the health utility and SDQ score.
Finally, we examined which percentage of patients had shoulder complaints at the moment of completing the questionnaire (SDQ>0) in each of the subgroups and checked whether these patients had a lower mean health utility than patients without shoulder complaints (SDQ=0).
All analyses were performed with SPSS version 20.

Results

Population characteristics

Of the 291 eligible patients (39 WW, 36 SLNB, 168 SOHND and 48 MRND), 181 patients (62%) returned the questionnaires. Most of these patients underwent a SOHND strategy (60%), followed by MRND (15%), WW (14%) and SNB (11%). The response rate was between 56 and 67 percent among the subgroups. In all treatment groups the majority of respondents were males. Overall mean age at completion of the questionnaire was 64.4 years. The mean age was higher in the WW group (71.4 years) as compared with the other subgroups. In the WW and SLNB group most tumors were staged T_1 while in the SOHND and MRND group most tumors were T_2. In all groups most tumors were located on the tongue. Population characteristics of the respondents are presented in Table 1. Some patients only filled out one of both questionnaires or did only fill out half of the EQ-5D-3L questionnaire (only the questions or only the VAS). 174 Patients answered the five questions of the EQ-5D-3L (and thereforea health utility could be calculated) , while 173 completed the VAS. 168 patients answered the questions of the SDQ. However, 17 patients (13 SOHND, 1 MRND, 2 SLNB and 1 in the WW group) answered all questions of the SDQ with non applicable and therefore no SDQ score could be calculated for these patients in accordance with the instruction of the questionnaire.(130) For 146 patients it was possible to both calculate a health utility and a SDQ score and with these results assessment of the relation between shoulder complaints and health utility was possible.

Health utility, SDQ and VAS scores

The adjusted mean utility scores for WW, SLNB, SOHND and MRND were 0.804, 0.863, 0.834 and 0.794 respectively (p value 0.700). The adjusted mean SDQ scores were 8.64, 10.57, 18.92 and 33.66 respectively (p value 0.019) and the VAS scores 69.7, 79.6, 76.2 and 71.5 respectively (p value 0.234). The mean scores were adjusted for age, gender and time since treatment (Table 2a).

In Table 2b the adjusted mean utility scores for patients who did not have any previous mucosal malignancy, local recurrence or second primary tumor are presented. Table 2c shows adjusted mean utility scores for patients who also did not receive adjuvant (chemo) radiotherapy.

Relation shoulder complaints and health utility

The correlation coefficient between SDQ score and health utility was -0.499 (p-value 0.000) and the beta coefficient was -0.003 (se 0.001), i.e. patients with more shoulder complaints appear to have a lower HRQo.

Table 1: Patient characteristics

	WW	SLNB	SOHND	MRND
Responders (response rate %*)	26 (67%)	19 (53%)	109 (65%)	27 (56%)
Sex				
Male (%)	17 (65.4)	13 (68.4)	70 (64.2)	15 (55.6)
Female (%)	9 (34.6)	6 (31.6)	39 (35.8)	12 (44.4)
Age (years at time of questionnaire)				
Mean (sd)	71.4 (11.4)	63.6 (9.4)	62.7 (12.2)	64.8 (10.6)
Range	54.8 – 91.6	44.9-80.2	29.5 – 84.6	40.5- 96.5
Time after treatment (year)				
Mean (sd)	4.8 (1.8)	1.9 (1.4)	5.2 (2.6)	5.2 (3.2)
Range	2.3 – 9.2	0.4- 4.1	1.6 – 12.2	0.4 – 11.0
T-stage				
T1 (%)	23 (88.5)	16 (84.2)	42 (38.5)	12 (44.4)
T2 (%)	3 (11.5)	3 (15.8)	67 (61.5)	15 (55.6)
Subsite				
Tongue (%)	14 (53.8)	10 (52.6)	59 (53.6)	12 (44.4)
Floor of mouth (%)	5 (19.2)	8 (42.1)	35 (31.8)	13 (48.1)
Other subsite (%)	7 (26.9)	1 (5.3)	16 (14.5)	2 (7.4)
Adjuvant (Chemo)radiotherapy				
Yes (%)	4 (15.4)	0 (0.0)	44 (40.4)	22 (85.5)
No (%)	22 (84.6)	19 (100.0)	65 (59.6)	5 (18.5)
Previous malignancy in head and neck region				
Yes (%)	2 (7.7)	1 (5.3)	11 (10.1)	2 (7.4)
No (%)	24 (92.3)	18 (94.7)	98 (89.9)	25 (92.6)
Local recurrence or second primary in head and neck region				
Yes (%)	4 (15.4)	0 (0.0)	8 (7.3)	6 (22.2)
No (%)	22 (84.6)	19 (100.0)	101 (92.7)	21 (78.8)

Patient Characteristics of the study population per subgroup: supraomohyoid neck dissection (SOHND), watchful waiting (WW), sentinel lymph node biopsy (SLNB), modified radical neck dissection (MRND). *The response rate is the percentage of returned questionnaires. Not every question was answered by every respondent.

Table 2a: Adjusted mean quality of life scores

	WW	SLNB	SOHND	MRND	p-value
Mean health utility (se)	0.804 (0.04) (n=26)	0.863 (0.05) (n=19)	0.834 (0.02) (n=104)	0.794 (0.04) (n=25)	0.700
Mean SDQ score (se)	8.64 (5.6) (n=25)	10.57 (7.7) (n=14)	18.92 (2.9) (n=91)	33.66 (6.0) (n=21)	0.019
Mean VAS score (se)	69.7 (3.7) (n=25)	79.6 (4.8) (n=16)	76.1 (1.8) (n=107)	71.5 (3.3) (n=25)	0.234

Mean scores (with standard errors) for: Health utility (number between 0 and 1 in which 1 stands for perfect health), SDQ score (higher score means more shoulder complaints) and VAS score (higher VAS means higher self-rated quality of life). P-value is between groups.

Table 2b: Adjusted mean quality of life scores in patients without previous mucosal malignancies, local recurrences or second primary tumors in the head and neck region

	WW	SLNB	SOHND	MRND	p-value
Mean health utility (se)	0.849 (0.05) (n=21)	0.859 (0.05) (n=18)	0.841 (0.02) (n=86)	0.800 (0.05) (n=20)	0.827
Mean SDQ score (se)	10.2 (4.8) (n=21)	8.3 (7.5) (n=14)	18.4 (3.1) (n=74)	34.4 (6.6) (n=16)	0.028
Mean VAS score (se)	71.2 (4.3) (n=20)	79.2 (5.1) (n=15)	76.9 (2.3) (n=89)	69.3 (4.1) (n=21)	0.239

Mean scores (with standard errors) for: Health utility (number between 0 and 1 in which 1 stands for perfect health), SDQ score (higher score means more shoulder complaints) and VAS score (higher VAS means higher self-rated quality of life). P-value is between groups.

Patients with shoulder complaints compared to patients without shoulder complaints

When we defined SDQ>0 as shoulder complaints, the percentage of respondents with complaints in the complete subgroups (at the moment of filling out the questionnaire) are: 32 percent in the WW group (8 out of 25), 36 percent in the SLNB group (5 out of 14), 43 percent in the SOHND group (39 out 91) and 77 percent in the MRND group (16 out of 21).

The mean utility of patients with shoulder complaints (SDQ>0, n=67 patients) was 0.78, while patients without shoulder complaints (SDQ=0, n=79 patients) had a mean utility of 0.90 (p-value 0.043).

Table 2c: Adjusted mean quality of life scores in patients without previous mucosal malignancies, local recurrence or second primary tumors and without adjuvant (chemo)radiotherapy in the head and neck region

	WW	SLNB	SOHND	p-value
Mean health utility (se)	0.826 (0.05) (n=20)	0.858 (0.05) (n=18)	0.849 (0.03) (n=53)	0.955
Mean SDQ score (se)	13.30 (5.3) (n=20)	8.16 (6.6) (n=14)	16.1 (3.3) (n=49)	0.771
Mean VAS score (se)	71.4 (4.4) (n=19)	79.2 (5.2) (n=15)	79.2 (2.5) (n=55)	0.314

Mean scores (with standard errors) for: Health utility (number between 0 and 1 in which 1 stands for perfect health), SDQ score (higher score means more shoulder complaints) and VAS score (higher VAS means higher self-rated quality of life). P-value is between groups. Number of patients with MRND were too low (n=4 for health utility and VAS and n=2 for SDQ score) to present meaningful adjusted mean scores.

Discussion

This study provides health utility scores for patients who underwent different procedures for the neck in early OCSCC and the impact of shoulder complaints on health utilities. As expected, our results show that more invasive procedures for diagnosing and treating neck metastases result in more shoulder complaints, and that shoulder complaints lead to a lower health utility.

<u>Strengths</u>

To our knowledge, this is the first study describing health utilities for patients who underwent different diagnostic and treatment modalities for the neck. Furthermore, we are also the first examining the relation between shoulder complaints and health utility in the OCSCC population. Other studies did examine shoulder complaints after neck dissections, and our results are in agreement with these studies regarding shoulder complaints after the different procedures.[25,63,126]

<u>Limitations</u>

Some potential limitations should also be discussed. First, the number of patients in the WW and SLNB groups was low. This could explain the relatively low health utility after WW, which was caused by a few outliers. Also the low VAS scores for the WW group indicate that the WW group is probably selective. We expect that if more patients were included, the mean utility after WW would have been, relatively, higher than reported now. Although not statistically significant, utilities showed a trend in being favorable in patients with less invasive procedures against more invasive procedures. A power calculation using the number of available patients in this study showed that the difference that could be shown with an alpha of 5% and a power of 80% was 0.18, whereas the reported minimally important difference (MID) is 0.06.[135] To detect this MID of 0.06 we should have included a minimum of 174 patients per group. However, our results show differences which are in a range around the MID and are, although not statistically significant due to the low numbers, thus clinically important. Second (partially coherent with the first limitation), we could not adjust for baseline quality of life as this was not available in our population. The WW group could be seen as a reference for the other groups to study the difference in quality of life and occurrence of shoulder complaints. However, confounding by indication could have influenced the results. We did try to correct for T-stage and subsite and the influence of these parameters seemed to be very low. Besides this we performed subgroup analyses in which we only included patients without previous or second malignancies in the head and neck region and patients without radiotherapy treatment. The differences in mean health utilities between WW, SLNB, SOHND and MRND were different in these subgroups without previous or second malignancies and patients without radiotherapy, as compared to the analyses in which all patients were included. Especially the health utility in the WW subgroup was relatively higher in these subgroup analyses. However, we still expected relatively higher values in the WW group and therefore expect that also other factors caused the lower scores in the WW group besides previous malignancies or radiotherapy. Besides this, WW, SLNB and SOHND are strategies used for the clinical negative neck, while MRND is used for the positive neck. Using the WW group as a reference group should therefore be done with caution. Third, we used the Dutch tariffs for the utilities and therefore results could be different in other countries. However, calculating utilities using the UK and US tariffs showed similar results. We, therefore, believe that our utilities could also be used to support treatment decisions in other countries. Fourth, the response rate of 62 percent may have introduced some selection bias. It is possible that for example only the patients with the best quality of life returned their questionnaire. However it is expected that this would have caused bias in the same direction in each of the groups and therefore did not influence the results greatly.

Clinical implications

This study provides important new information about quality of life after several diagnostic and treatment modalities in OCSCC patients. Choices which are made with respect to management of the cN_0 neck seem to influence quality of life. Therefore, quality of life should be included in the decision making process for $cT_{1\text{-}2}N_0$ OCSCC patients. Health utilities can be used in combination with survival to calculate QALYs which should be used in decision analysis. Prospective measurement of health utilities could be performed to improve the quality of these values. However, prospective research takes time and until this is performed the utility measures from this study are the best available and could help in treatment decisions.

Conclusion

The relatively low percentage of shoulder complaints and the high health utility value after SLNB, in combination with the high sensitivity reported in our recent meta-analysis support a strategy including SLNB.[65] When we introduced the utilities, calculated in this study, into our previously constructed decision analytical model (in which they are weighed against oncological outcomes), the strategy including SLNB was indeed the most (cost-)effective strategy and was even more (cost-)effective than previously calculated.[127]

Chapter 8

Cost-effectiveness of selective neck dissection versus modified radical neck dissection for treating metastases in oral cavity cancer patients; a modeling study

Tim M Govers

Sejal Patel

Robert P Takes

Matthias AW Merkx

Maroeska M Rovers

Janneke PC Grutters

Published: Head & Neck 2014, July

Abstract

Background: Choosing between a more or less extensive neck dissection implies a trade-off between survival, quality of life and costs. This study aims to determine if selective neck dissection (SND= level I-III or I-IV) is cost-effective compared with modified radical neck dissection (MRND = level I-V) in cT_{1-2} oral cavity squamous cell carcinoma (OCSCC) patients with singular nodal disease confined to level I or II.

Methods: A decision-analytic model was developed to model quality-adjusted life years (QALYs) and costs over a life time horizon, based on literature.

Results: The SND strategy resulted in an expected health loss of 0.06 QALY and savings of €1,351 per patient compared with MRND. The results were sensitive to differences in regional failure probabilities between the strategies.

Conclusion: With the evidence used in this model, SND was not cost-effective compared to MRND. Prospective research on regional failure is needed to provide optimal treatment for OCSCC patients.

Introduction

Oral cavity squamous cell carcinoma (OCSCC) is the sixth most common cancer worldwide, accounting for an estimated 4% of all cancers. Despite significant technical advances in the treatment of oral cancer, survival following a diagnosis of OCSCC remains poor with 5-year survival around 50% overall, with only limited improvement in the past decades.[136] OCSCC has a significant propensity of regional lymph node metastases. If nodal metastases are detected the usual treatment consists of neck dissection, to be followed by postoperative (chemo)radiotherapy on indication.[137]

Traditionally, modified radical neck dissection (MRND) is performed in these cases irrespective of the number of affected nodes and levels of the neck, a procedure in which all levels (I-V) of the involved neck side are removed. Common sequellae of this procedure are shoulder dysfunction and pain, loss of sensitivity of the overlying skin, cranial nerve impairment and anatomic deformities, all of which have an effect on the quality of life of the patient.[18] In more recent years, the alternative of performing a selective neck dissection (SND) is considered in cases with singular nodal disease confined to level I or II. Since SND is limited to level I-III or I-IV, shoulder morbidity is lower compared to MRND. [24,25] This difference is explained by the fact that removal of level V in particular is associated with this type of morbidity.[18] On the other hand, the preservation of certain lymph node levels may lead to higher rates of regional failure, requiring additional treatment with re-surgery and/or radiotherapy and possibly resulting in worse survival. Moreover, the burden and costs of regional failure and corresponding treatments may exceed those of one single surgical procedure removing all levels of the neck. The choice between more or less invasive treatment, therefore, implies a trade-off between survival, quality of life and costs. The resulting question is whether neck levels associated with higher morbidity (especially level V) could be spared in patients with limited nodal disease.

The aim of this study is to determine if selective neck dissection (SND level I-III or I-IV) is cost-effective compared with modified radical neck dissection (MRND level I-V) in patients with $T_{1\text{-}2}$ OCSCC and clinical singular nodal disease (stage cN_1 or cN_{2a}) confined to level I or II.

8

Methods

Decision model

A decision-analytic model was developed to evaluate survival, quality of life and costs associated with SND and MRND in $T_{1\text{-}2}$ OCSCC patients with clinical singular nodal disease (stage N_1 or N_{2a}) confined to level I or II (referred to as limited nodal disease for the rest of the article).

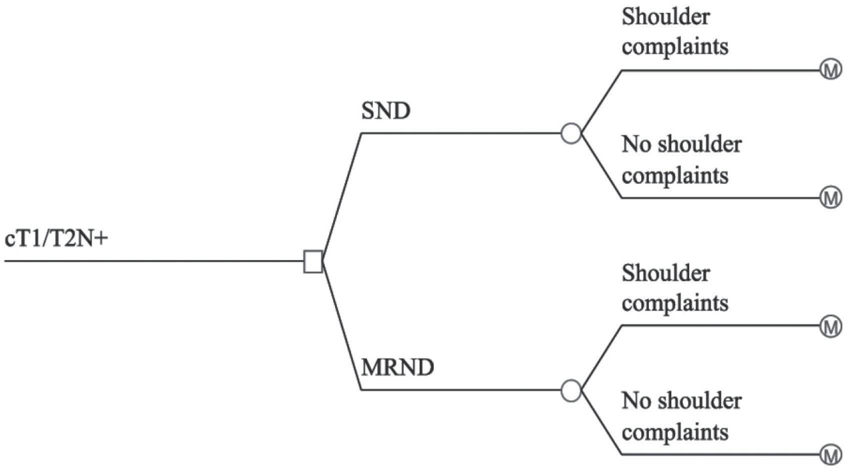
Decision modelling provides a framework for decision-making under conditions of uncertainty. It can provide a structure that reflects the possible prognoses of patients and how the treatments being evaluated may impact these prognoses. It offers an analytical framework within which evidence relevant to the study question can be synthesized and provides a means of translating this evidence into estimates of effects and/or costs for the options being compared. Besides this, it facilitates an assessment of uncertainty relating to the evaluation and therewith identifies priorities for future research.[122]

Model structure

The first part of the model consisted of a decision tree to model the direct effects of treatment for the two strategies. In both treatment options patients had a chance of suffering from shoulder morbidity or no shoulder morbidity after treatment (Figure 1). The number of cases requiring postoperative radiotherapy in addition to neck dissection was assumed to be equal in both treatment options and was therefore not explicitly included in the model. The decision tree was followed by a Markov model to simulate the long-term consequences. The Markov model involved the transition of patients across health states over time. The health states were 'regional failure with shoulder morbidity', 'regional failure without shoulder morbidity', 'no regional failure with shoulder morbidity', 'no regional failure without shoulder morbidity' and 'dead' (Figure 2). Regional failure was defined as having ipsilateral regional recurrence after neck dissection.

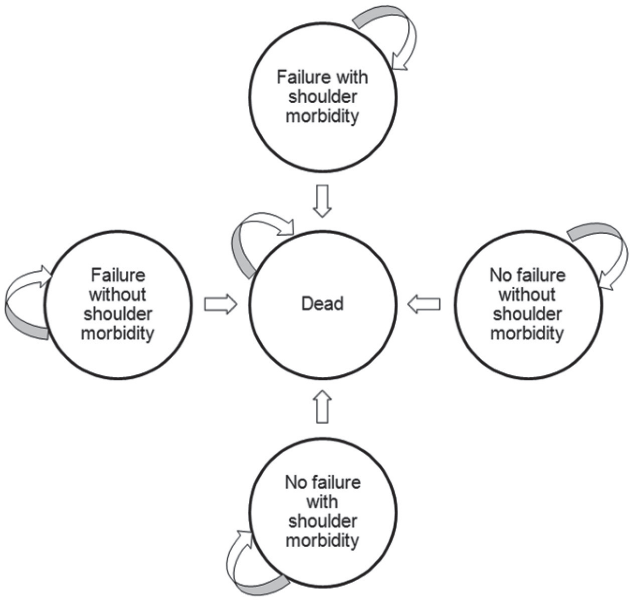
A hypothetical cohort of patients was sent trough the model to evaluate the consequences of both strategies. The model had a cycle length of one year and a lifelong time horizon was used. Patients who suffered from shoulder morbidity after neck dissection were assumed to experience these complications within one year after surgery. It was assumed that patients experienced these shoulder complaints for each following year. After each cycle the patients could stay in their health state or die. Costs and effects were assigned to each health state. The model was constructed and analysed using TreeAge Pro 2012 software (TreeAge Software Inc, Williamstown, MA, VS).

Figure 1: Model structure for the decision tree



The square represents a decision node, the open nodes are chance nodes and the open circles with M are Markov nodes (the Markov part of the model starts after the M nodes). Abbreviations: Selective neck dissection (SND), Modified radical neck dissection (MRND)

Figure 2: Influence diagram of Markov part of the model



Patients with shoulder complaints could experience regional failure and enter the Markov part of the model in the Failure with shoulder complaints state. Otherwise they enter the Markov state in the no failure with shoulder morbidity state. These states have different survival probabilities. The same applies to patients without shoulder complaints

Transition probabilities

Transition probabilities were included in the model to distribute patients over the pathways. The probabilities of shoulder morbidity and probability of regional failure differed between strategies. Probabilities of shoulder morbidity after SND and MRND were derived from a Dutch retrospective study based on 65 and 51 patients, respectively.[24] Probabilities for regional failure after neck dissection were derived from 64 patients of a published retrospective study.[138] Survival after regional failure was the same in both treatment options. Data on survival for the first 5 years (for patients with and without regional failure) were based on a previous modelling study.[127] In these survival rates it is included that not all patients are able (or want) to receive salvage therapy. After 5 years, survival was assumed to be equal to the overall mortality rates of the relevant age group, which were derived from Statistics Netherlands (Table 1).[139] An overview of the transition probabilities used in the model is presented in Table 2.

Table 1: Survival

Parameter	Point estimate (chance)	Sd	Source
With regional failure			
Year 1	0.80	Fixed	Govers, 2013(127)
Year 2	0.78	Fixed	Govers, 2013(127)
Year 3	0.93	Fixed	Govers, 2013(127)
Year 4	0.83	Fixed	Govers, 2013(127)
Year 5	0.99	Fixed	Govers, 2013(127)
Year 10	0.98	Fixed	CBS, 2012(139)
Year 15	0.97	Fixed	CBS, 2012(139)
Year 20	0.94	Fixed	CBS, 2012(139)
Year 25	0.89	Fixed	CBS, 2012(139)
Year 30	0.81	Fixed	CBS, 2012(139)
Year 35	0.68	Fixed	CBS, 2012(139)
Without regional failure			
Year 1	0.95	Fixed	Govers, 2013(127)
Year 2	0.90	Fixed	Govers, 2013(127)
Year 3	0.93	Fixed	Govers, 2013(127)
Year 4	0.91	Fixed	Govers, 2013(127)
Year 5	0.99	Fixed	Govers, 2013(127)
Year 10	0.98	Fixed	CBS, 2012(139)
Year 15	0.97	Fixed	CBS, 2012(139)
Year 20	0.94	Fixed	CBS, 2012(139)
Year 25	0.89	Fixed	CBS, 2012(139)
Year 30	0.81	Fixed	CBS, 2012(139)
Year 35	0.68	Fixed	CBS, 2012(139)

Abbreviations: standard deviation (sd)

Table 2: Transition probabilities

Parameter	Point estimate (chance)	Distribution	Distribution input	Source
Shoulder morbidity after selective neck dissection	0.20	Beta	α: 13 β: 52	van Wilgen, 2004(24)
Shoulder morbidity after modified radical neck dissection	0.33	Beta	α: 17 β: 34	van Wilgen, 2004(24)
Regional failure after selective neck dissection	0.11	Beta	α: 5 β: 40	Schiff, 2005(138)
Regional failure after modified radical neck dissection	0.05	Beta	α: 1 β: 18	Schiff, 2005(138)

Outcome measures

To be able to take into account both survival and quality of life, effectiveness was measured in quality-adjusted life years (QALYs). To calculate QALYs, health utilities were assigned to the health states and were combined with the survival rates (life years).

Data on utilities, for the group of patients with shoulder morbidity and without shoulder morbidity, were derived from a database including 136 patients who underwent SND or MRND (Table 3) (unpublished data). In the dataset, shoulder morbidity was measured using the shoulder disability questionnaire (SDQ). Patient with a score above 10 on the SDQ were defined as patients with shoulder morbidity. The utilities assigned to the health states were numbers between 0 (dead) and 1 (perfect health). The EuroQol-5D questionnaire (EQ-5D) was used to measure health utilities. This was done by combining the answers of the questionnaire with Dutch tariffs.[131] We did assume that utilities only changed due to shoulder complaints and not directly by a regional recurrence. A discount rate of 1,5% was applied to effects as recommended by the Dutch guidelines for pharmaco-economic research.[140]

Table 3: Quality of life

Parameter	Point estimate (utility)	Distribution	Distribution input	Source
Without shoulder morbidity	0.91	Beta	Sd: 0.12	Unpublished data
Shoulder morbidity	0.79	Beta	Sd: 0.23	Unpublished data

Abbreviations: standard deviation (sd)

Costs

The cost analysis was performed from a societal perspective. Volumes for hospital stay after the neck dissection were derived from an internal database from the Radboud University Medical Centre. Volumes for physical therapy treatment were obtained from expert consultation of a physical therapist. It was assumed that all patients underwent physical therapy if they suffered from shoulder morbidity. The number of outpatient consultations were based on the follow-up guidelines for patients with oral cavity and oropharyngeal carcinoma of the Dutch Head and Neck Society.[141] The number of days of absenteeism were derived using data from the Dutch patient association for head and neck cancer.[142] These volumes are shown in Table 4.

Costs of surgery for the two types of neck dissection were obtained from the Radboud University Medical Centre. Standard unit costs were used for hospital days, absenteeism, consultation and physiotherapy.[143] Academic unit costs were used for hospital days and consultation. Every patient who underwent a neck dissection was assumed to have received a pre- and postoperative consult by a physical therapist to examine shoulder function. The costs of these consultations are equal to the costs of one physical therapy session, each.

It was assumed that a regional failure was followed by salvage therapy with the costs of a complete cycle of radiotherapy, independent of type of neck dissection.[144] For all costs in the model, price index numbers from Statistics Netherlands were used to convert them to the 2012 price level.[145,146] The annual discount rate for costs was set at 4%.[140] Table 4 summarizes costs, in Euro, used in the model.

Table 4: Volumes

Parameter	Point estimate volume	Distribution	Distribution input	Source
Hospital days after selective neck dissection	11.8	Fixed	-	Radboud University Medical Centre
Hospital days after modified radical neck dissection	12.6	Fixed	-	Radboud University Medical Centre
Number of treatments physiotherapy for shoulder morbidity	24	Fixed	-	McNeely, 2004(147)
Number of out-patients consultation	14	Fixed	-	Nederlandse Werkgroep Hoofd-Halstumoren(141)
Absenteeism after neck dissection without shoulder morbidity (days)	17.5	Gamma	SE: €318.6 (Total costs)	Stichting Klankbord (142)
Absenteeism after neck dissection with shoulder morbidity (days)	35	Gamma	SE: €637.1 (Total costs)	Stichting Klankbord (142)

Abbreviations: standard error (SE)

Table 5: Unit costs

Parameter	Point estimate (€)	Source
Selective neck dissection (I-III)	2994	Radboud University Medical Centre
Modified radical neck dissection	3473	Radboud University Medical Centre
Regional failure after selective neck dissection	10,524	Ramaekers, 2013 (144) CBS StatLine (146)
Regional failure after modified radical neck dissection	10,524	Rameakers, 2013 (144) CBS StatLine (146)
Hospital day (one day)	611	Hakkaart-van Roijen (143) CBS StatLine (146)
Absenteeism (one day)	250	Hakkaart-van Roijen (143) CBS StatLine (145)
Preoperative consultation	38	Hakkaart-van Roijen (143) CBS StatLine (146)
Postoperative consultation	38	Hakkaart-van Roijen (143) CBS StatLine (146)
Physiotherapy (one consultation)	38	Hakkaart-van Roijen (143) CBS StatLine (146)
Out-patients consultation	137	Hakkaart-van Roijen (143) CBS StatLine (146)

<u>Analysis</u>

A base case analysis was conducted by calculating expected life years, QALYs and costs of the two strategies. In general there were three possible outcomes. First, one of the treatments is more effective (i.e. gains more QALYs) and is also less expensive. That treatment dominates the other treatment. Second, SND is more effective than MRND but also more costly. In this case the difference in effects has to be weighed against the difference in costs. This is done by calculating an incremental cost-effectiveness ratio (ICER), by dividing the estimated difference in costs by the difference in QALYs. This ICER represents the additional cost made to gain a QALY. Third, SOHND is less effective than MRND but also less costly. Again an ICER (weighing the difference in effects against the difference in costs) is calculated, which will represent the cost savings per QALY lost. We used a threshold value, i.e. the amount society is willing to pay to gain a QALY or willing to accept to lose a QALY, of €80,000.[148]

In addition, deterministic and probabilistic sensitivity analyses were performed to evaluate the sensitivity of the results to changes in parameter values. The deterministic sensitivity analyses included a one-way sensitivity analysis and an analysis from a healthcare perspective. First, one-way sensitivity analysis was performed on the transition probabilities of regional failure, by varying this parameter through a range of possible values. The probability of regional failure after MRND was fixed and the probability of regional failure after SND was varied between 0.05 and 0.11 using six intervals. In this way the impact of a number of possible values of this parameter on the outcome of the model was evaluated. Second, an analysis from a healthcare perspective was performed using only costs for healthcare. In this analysis costs for absenteeism were, therefore, excluded. To investigate sampling uncertainty, probabilistic sensitivity analysis was performed with 1000 (Monte Carlo) simulations. For this purpose, beta distributions were assigned to transition probabilities and utilities, and gamma distributions to costs of absenteeism (Table 2 and 4). The results were presented in a scatter plot, where each point represents a single iteration of the simulation.[149]

Results

Base case analysis

By performing MRND on average 16.87 life years were yielded compared to 16.48 with SND.

Taking into account quality of life (i.e. calculating QALYs), MRND was on average still the most effective strategy, yielding 12.38 QALYs, compared with 12.33 QALYs in the SND strategy (Table 6). SND was found to have lower expected costs (€17,291) than MRND (€18,642). The loss of 0.06 QALYs and savings of €1,351 in the SND strategy compared with the MRND strategy resulted in an ICER of €24,268 saved per QALY lost.

Table 6: Base case analysis based on societal perspective

Strategy	Mean cost (€)	Incremental cost (€)	Mean effect (QALY)	Incremental effect (QALY)	Average costs/effect (€/QALY)	ICER (€/QALY)
MRND	18,642	x	12.38	x	1,505	x
SND	17,291	-1351	12.33	-0.06	1,403	24,268

Abbreviations: Modified radical neck dissection (MRND), Selective neck dissection (SND), Quality adjusted life years (QALY)

Deterministic sensitivity analysis

For the base case analysis a transition probability of 0.11 for regional failure after SOHND was used, compared with a probability of 0.05 after MRND. When we changed the transition probability after SND to 0.10 the ICER was €801,724 per QALY lost. For each value below 0.10 the MRND strategy was dominated by the SND strategy. When the difference in probability of regional between the strategies was lower than 5 percent, SND was the most cost-effective strategy (Table 7).

The analysis from a healthcare perspective (using only healthcare costs) showed a saving of €768 with SND (€12,046) compared to MRND (€12,814). Using the earlier calculated effect of 0.06 QALYs lost, this resulted in an ICER of €13,798 saved per QALY lost (Table 8).

Table 7: Sensitivity analysis

Probability of failure after SOHND	Strategy	Costs (€)	Effects (QALYs)	ICER (€/QALY)*
0.11 (base case)	MRND	18,642	12.38	x
	SOHND	17,291	12.33	24,268
0.1	MRND	18,642	12.38	x
	SOHND	17,176	12.38	801,724
0.09	MRND	18,615	12,40	
	SOHND	17,072	12,43	Dominates MRND

*ICER was given in savings per QALY lost. Abbreviations: Modified radical neck dissection (MRND), Selective neck dissection (SND), Quality adjusted life years (QALY)

Table 8: Analysis based on health care perspective

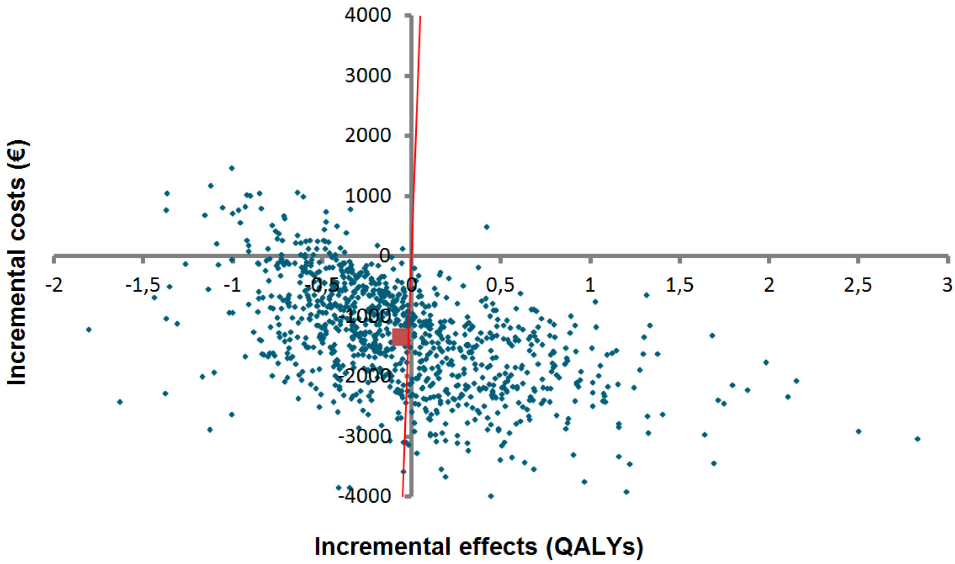
Strategy	Mean cost (€)	Incremental cost (€)	Mean effect (QALY)	Incremental effect (QALY)	ICER (€/QALY)
MRND	12,814	x	12.38	x	x
SND	12,046	-768	12.33	-0.06	13,798

Abbreviations: Modified radical neck dissection (MRND), Selective neck dissection (SND), Quality adjusted life years (QALY)

<u>Probabilistic sensitivity analysis</u>

The scatter plot of SND versus MRND presents the spread in incremental costs and effects of the 1000 Monte Carlo simulations (Figure 3). In 60% of the simulations SND was less effective than MRND. Only in 6% of the simulations SND was more costly. The red line in Figure 3 represents the WTA/WTP threshold of €80,000 per QALY. In the simulations which are to the right of the red line (41%), SND was cost-effective compared to MRND. In the simulations to the left of the red line (59%) MRND was more cost-effective.

Figure 3: Scatter plot of probabilistic sensitivity analysis



Each of the dots represents one of the 1000 Monte Carlo simulations. For these simulations, distributions were assigned to parameters in the model. Abbreviations: Quality adjusted life years (QALYs)

Discussion

Our model provides insight in treatment strategies for T_{1-2} OCSCC patients with limited nodal disease. On average, SND yielded less life years than MRND. Taking into account quality of life (i.e. calculating QALYs) MRND was still the most effective strategy, although differences between the strategies became smaller with lower quality of life after MRND. The better survival rate of MRND seems to outweigh the loss in quality of life after this more extensive procedure. SND was less costly than MRND. The incremental cost-effectiveness ratio was in favour of the MRND strategy, i.e. the savings in costs that can be reached with SND do not outweigh the loss in effectiveness. However, deterministic sensitivity analysis showed that the outcomes were highly sensitive to the assumed probabilities of regional failure after both strategies. When the difference in probability of regional failure between the strategies is lower than about 5 percent, SND is the most cost-effective strategy.
As far as we are aware, we are the first studying the cost-effectiveness of the two strategies in OCSCC patients with limited nodal disease. In this model we used QALYs to take into account both survival and quality of life. Furthermore we included indirect costs such as costs of physiotherapy and absenteeism. Therefore it seems to provide a comprehensive reflection of reality.

Some potential limitations should also be discussed. First, we used quality of life data on selective neck dissections of level I-III, whereas selective neck dissections for the cN+ neck will sometimes include level I-IV (depending on the involved levels and the preference of the surgeon). This could have overestimated the utility after selective neck dissection. However, dissection of level IV does not give much extra morbidity relatively to dissection of level V.[18] Second, we used a study of Schiff et al.for the regional failure probabilities of both strategies.[138] This study, however, also included some higher N-stages. Approximately 60% was stage N_1, 19% N_{2a}, 14% N_{2b} and 5% N_{2c} (Written personal communication with the authors). Therefore, a higher absolute risk of regional failure could have been found in these data for both strategies than one would find in a population with limited nodal disease only (cN_1 and CN_{2a}). This higher absolute risk in both strategies could have caused an overestimation of the relative risk of regional failure between the strategies (i.e. the difference in regional failure probabilities between SND and MRND). On the other hand, patients in the MRND group had higher pathological N-stages (pN-stages) compared to patients in the SND group, in this particular study. This in turn could have reduced a possible overestimation of the difference in regional failure probability between the strategies. We performed sensitivity analysis on the difference in probability of regional failure to assess the influence of these possible limitations.

Third, indications for radiotherapy in addition to neck dissection were assumed to be equal in both treatment options and were therefore not taken into account in this model. In the study of Schiff et al. more patients in the MRND group received postoperative radiotherapy. This could have reduced regional failure probability in this group. However, as mentioned above, these patients also had higher pathological N-stages.

Fourth, OCSCC is most prevalent among older, often retired, people. We, however, used average wages for absenteeism, which might have resulted in an overestimation of the costs of absenteeism. The analysis from a healthcare perspective, excluding the costs of absenteeism, did not alter our conclusion.

Fifth, in our model the same utility rate for shoulder morbidity was assigned every year. Since patients will receive physiotherapy it is expected that shoulder morbidity will decrease and the quality of life will improve over time. Furthermore, patients may learn to cope with their shoulder morbidity which could have a positive effect on their quality of life.[24] Possibly this will lead to an underestimation of the effects of both strategies, which seems largest for MRND where most shoulder complaints occur. On the other hand, shoulder complaints after MRND may be more permanent and less responsive to physiotherapy than after SND, which would imply a larger underestimation for SND.

Sixth, we assumed that salvage therapy after regional failure was equal in both strategies. In practice there will be a difference in salvage therapy after a regional failure between SND and MRND. It is likely that more effective treatment options may be available for regional failure after SND. To model the costs of salvage therapy we assigned the costs of a complete cycle of radiotherapy to every patient with regional failure. This may be an overestimation of the actual costs of salvage therapy. Since this overestimation was made in both strategies it did not influence the results.

Despite these possible limitations our model provides clinically important information and a concept of how to evaluate and give insight in the consequences of changes in treatment algorithms and strategies. To provide optimal treatment not only survival should be taken into account but also quality of life. In this study we combined these outcomes by using quality-adjusted life years. The model shows that, with used evidence, the overall survival gain with MRND outweighs the extra shoulder complaints and extra costs associated with this strategy. However, performing SND could save costs for society and healthcare. With raising healthcare costs, consideration of less expensive treatment options becomes more and more important. This raises the question if we are willing to drop effective conventional procedures to save costs. And, if so, how much savings would be required to accept a loss in effectiveness.[150,151] The €1,351 that are saved per patient when performing SND instead of MRND seem insufficient to accept the loss of 0.06 QALYs per patient.

The model shows that small changes in the assumed probabilities of regional failure have high impact on the outcomes. If the difference in probability of regional failure between the strategies is lower than 5%, SND was found to be more effective and cost-effective than MRND. Regional failure probabilities for both strategies could only be derived from a retrospective study with several limitations and were therefore uncertain. Therefore, prospective research on this parameter is highly valuable. With the results of such research (including subgroup analysis) our model can be used to show which strategy is optimal for each group of patients.

In conclusion, given the evidence used in our model, SND was not cost-effective compared to MRND for $T_{1\text{-}2}$ OCSCC with limited nodal disease ($N_{1\text{-}2a}$), since this strategy was on average more effective without significant extra costs compared with SND. Prospective research on regional failure probabilities is highly valuable to obtain better data to reduce uncertainty and provide optimal treatment for subgroups of HNSCC patients and OCSCC in particular.

Part III

Chapter 9

Decision modeling to inform evidence-based personalized decisions: Using neck management in oral cavity cancer patients as an example

Tim M Govers

Maroeska M Rovers

Marieke Brands

Emilie AC Dronkers

Robert J Baatenburg de Jong

Matthias AW Merkx

Robert P Takes

Janneke PC Grutters

Submitted

Abstract

Objective: To show how prediction models can be incorporated in decision models, to allow for personalized decisions; and to assess the value of this approach in the management of the neck in early stage oral cavity squamous cell carcinoma (OCSCC).

Study design and setting: In a decision model, three approaches were compared: a 'population-based' approach in which patients undergo the strategy that is optimal for the population; a 'perfectly predicted' approach, in which each patient receives the optimal strategy for that specific patient; and a 'prediction model' approach in which each patient receives the strategy that is optimal based on prediction models. The differences in costs and Quality Adjusted Life Years (QALYs) between these approaches were studied.

Results: The population-based approach resulted in 4.9158 QALYs with €8,675 in costs, per patient. The perfectly predicted approach yielded 0.21 more QALYs and saved €1,024 per patient. The prediction model approach yielded 0.0014 more QALYs and saved €152 per patient compared with the population-based approach.

Conclusion: The perfectly predicted approach shows that personalized care is worthwhile. However, current prediction models in the field of OCSCC have limited value. Incorporating prediction models in decision models appears to be a valuable method to assess the value of personalized decision making.

Introduction

In the last years personalized healthcare has gained increasing attention. Personalized healthcare may be thought of as the tailoring of medical treatment to individual characteristics, needs and preferences of a patient during all stages of care.[152] The evidence-based medicine (EBM) approach historically focuses on randomized clinical trials. In such comparative studies, which can include economic evaluations, often little attention is paid to patient heterogeneity. As a result, reimbursement decisions are generally based on average (cost-)effectiveness of the intervention in a selected population.[153] This is also reflected in clinical guidelines, in which recommendations are often based on studies that pertain to the average population. They often do not guide in dealing with individual patient characteristics.[154,155]

In recent years, EBM research increasingly takes into account patient heterogeneity.[156] For example, prediction models are developed to estimate a probability of a certain outcome in an individual, given his or her personal or disease characteristics.[157,158] These probability estimates can guide care providers as well as the individuals themselves in deciding upon further management.[159] Such models are becoming increasingly abundant in medical literature. Often multiple outcomes (e.g. survival and side effects) are relevant for one patient and therefore multiple prediction models are relevant for such an individual patient. This implies a trade-off between different outcomes, which makes the use of these prediction models for clinical decision making difficult.

This is for example the case in early stage oral cavity squamous cell carcinoma (OCSCC) patients. If they have no clinically detected lymph node metastases in the neck (cN_0), the decision whether to remove the lymph nodes or not implies a trade-off between survival, quality of life and costs. In a previous study we used decision analytical modeling to inform this trade-off for an average cohort of patients.[127] Sensitivity analysis showed that predictions for individual patients, based on their characteristics, such as a patient's probability of occult metastases or his quality of life after surgery, influence which strategy is deemed optimal. Therefore it seems worthwhile to include patient characteristics into this decision model to weigh risks, benefits and costs for individual patients, to allow for more personalized evidence-based treatment decisions.

The aim of this study was to show how prediction models can be incorporated in decision models, and to assess the value of personalized decisions based on these prediction models using the case of the management of the neck in early stage oral cavity squamous cell carcinoma (OCSCC) as an example.

Methods

The case

In about 30% of patients with early stage (T_{1-2}) OCSCC, who have no clinically detected regional lymph node metastases (cN_0), occult metastases are present in the lymph nodes of the neck.[86] Therefore the decision needs to be made whether to remove those lymph nodes at risk for involvement by metastatic cancer, or to follow a watchful waiting strategy. Removing the lymph nodes at risk by performing a neck dissection increases prognosis.[14] However, this invasive procedure is unnecessary in 70% of patients, and quality of life could be reduced due to shoulder complaints as a result of the neck dissection.[23] Watchful waiting on the other hand implies undertreatment in 30%, which could result in lower survival when occult metastases are present.[14]

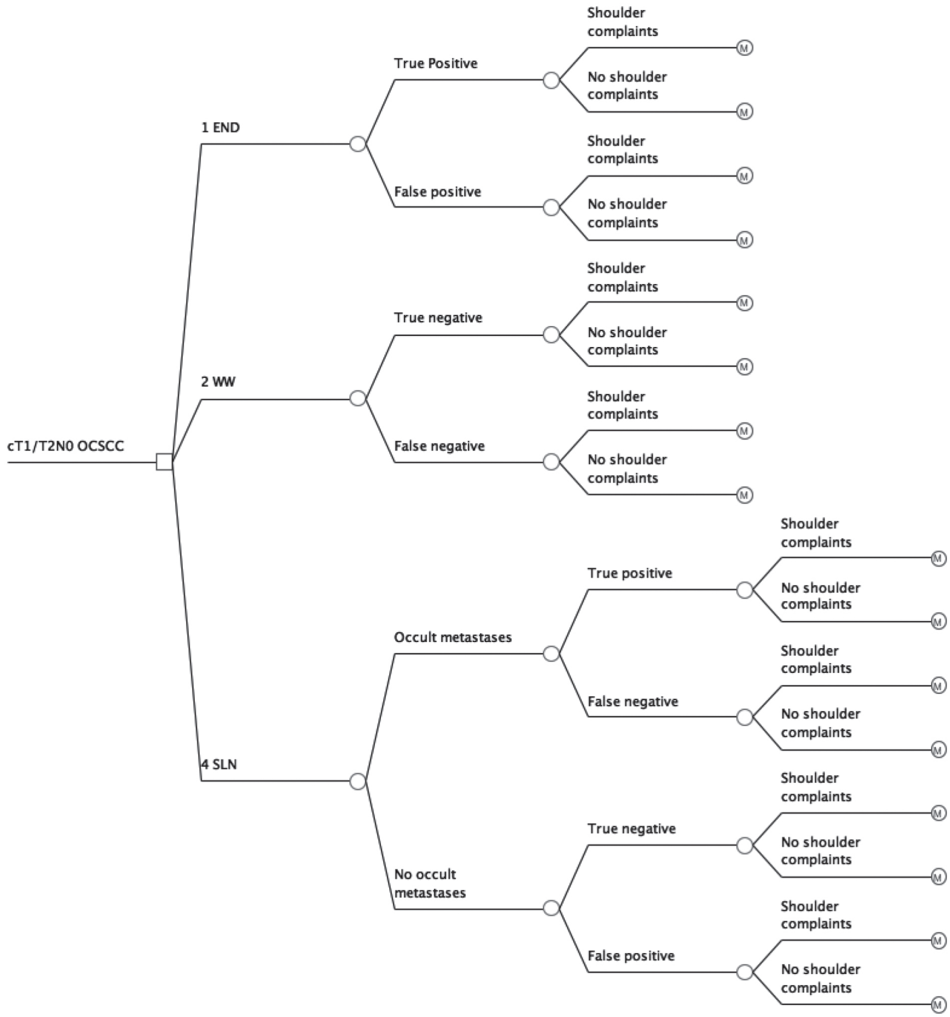
Decision Model

We used our previously published decision model as the basis for this study.[127] In this model five management strategies for the neck in $cT_{1-2}N_0$ OCSCC patients were evaluated. The strategies were: (1) elective neck dissection (END), (2) watchful waiting (WW), (3) gene expression profiling (GEP) followed by neck dissection (ND) or WW, (4) sentinel lymph node (SLN) procedure followed by ND or WW, and (5) GEP and SLN procedure (in case of positive GEP) followed by ND or WW.

A decision tree was used to model the diagnostic and treatment pathways, and a Markov model represented the subsequent follow-up. The diagnostic and treatment pathways are shown in Figure 1. Patients who actually do have occult metastases can either be correctly classified as positive (true positive) or falsely classified as negative (false negative). Those that do not have occult metastases can either be true negative or false positive. The false positive and true positive patients receive ND while the false negative and true negative patients undergo WW. All patients who are classified as positive are treated and therefore have a probability of having shoulder complaints. In the END (strategy 1) all patients are considered positive and are therefore true positive or false positive. In the WW strategy (strategy 2) all patients are considered negative and are therefore true negative or false negative. After this diagnostic and treatment pathway, patients move to the different health states in the Markov model, which represent the follow-up of patients. Possible health states for patients were: regional failure (RF), no RF, or dead. The structure of the Markov part of the model (Fig. 2) was the same for each diagnostic outcome. However, the probabilities of entering the different states (i.e. the chance of regional failure) were different among the diagnostic outcomes. Patient with occult metastases had a higher probability of regional failure if they are classified negative (false negative) than when they are classified positive (true positive). The utility values of the regional failure and no regional failure states depend on whether patients have shoulder complaints or not. Patients who are classified positive have a higher chance of having shoulder complaints and therefore on average, a lower health utility which was unnecessary in case of a false positive classification.

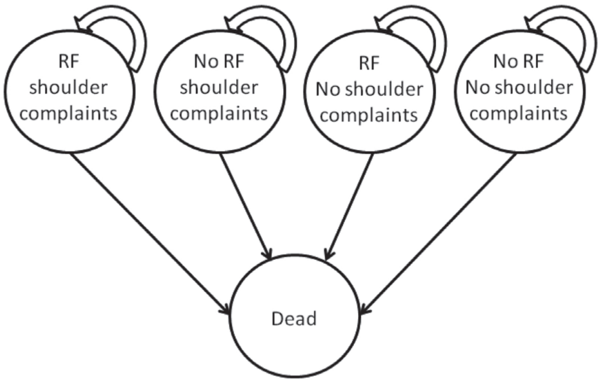
The time horizon of the model was 10 years, with yearly cycles. The model was built in Treeage Pro 2014.

Figure 1 Model structure decision tree



Decision Tree part of the model. Gene expression profiling (GEP) and GEP + sentinel lymph node (SLN) procedure were left out of the Figure. GEP had the same structure as the SLN strategy. GEP + SLN has an extra possibility after False positive to go to True negative.

Figure 2: Markov part of the model



Influence diagram of the Markov part of the model. The Markov part of the model represents follow-up after diagnostic outcome and corresponding treatment. Patients have a probability of entering the Markov model in the Regional failure (RF) or No RF state, depending on whether they have occult metastases and received treatment. These states could be with or without shoulder complaints. Patients can either survive and stay in the state they entered or die. The transition probabilities between the RF state and dead and the No RF state and dead were not depended on having shoulder complaints. However, the utility values of the RF or no RF state did differ between patients with and without shoulder complaints.

<u>Prediction models</u>

Our previous decision model showed that results were sensitive for changes in percentage of occult metastases and utility values. Although not tested in the previous study, we also expected survival to be an important parameter. Therefore we included prediction models for percentage (chance of) occult metastases, utility values (for patients with and without shoulder complaints) and survival (for patients with and without regional failure). Because of the inclusion of multiple prediction models it was important that we did not double count the effect of patient characteristics by incorporating the same effect of a characteristic in multiple prediction models. First, we searched for prediction models for these parameters in the literature which were suitable for the use in the decision model. In absence of suitable models we searched for data to develop such a model. Because the development and description of these models was not the aim of our study, we only present basic information about the models. However, we are aware that normally newly developed prediction models deserve a more comprehensive description. When these prediction models are further developed or used in practice, a more comprehensive description is warranted.[157]

Prediction model 1 - Probability of occult metastases

We searched for a multivariable logistic model to predict absolute risks of occult metastases in cN_0 patients. Multiple studies regarding predictors of occult metastases have been performed in OCSCC patients.[160-163] However, none of the identified models could be incorporated in our decision model. Some studies also included cN_+ patients [164,165], while other studies did not perform multivariable analyses, or absolute risks were neither provided nor calculable. [160-164,166] Therefore we developed a new prediction model based on registry data of the Radboudumc. This multivariable logistic regression model was based on 235 Dutch patients diagnosed with early stage cN_0 OCSCC between 2000 and 2013 in the Radboudumc. These patients were treated surgically for the primary tumor and underwent an END which was the standard treatment in this period. The predicted outcome was occult metastases (yes/no), which was defined as the pN stage (pN_+ or pN_0) after pathological examination of the neck specimen. The final model consisted of four predictors: T-stage (T_1/T_2), vasoinvasive growth (yes/no), spidery growth (yes/no), and depth of invasion (in mm). The C-index of the model was 0.725. Further details about this prediction model are listed in Appendix A. This model was incorporated in the decision model to define the probability of having occult metastases (Figure 1).

Prediction model 2 - Utility values

Because it is known that utility values depend on age and gender,we wanted to include a model with at least these two predictors.[132] No studies were available that predict utility values in OCSCC patients based on patient characteristics. Therefore we developed two multivariable linear regression models based on information we obtained in our previous quality of life study in 153 early stage OCSCC patients.[167] These patients were treated in the Radboudumc or the Netherlands Cancer Institute between 2001 and 2013 with different strategies for the neck (WW, SLN procedure or ND). One model predicted the utility value for patients with shoulder complaints (n=68), while the other model predicted the utility value for patients without shoulder complaints (n=83). Both models consisted of two predictors: age and gender. The R-squares of the models were 0.075 for the model for patients with shoulder complaints and 0.040 for the model for patients without shoulder complaints. Further details about these prediction models are listed in appendix B. The prediction models were used to assign utility values to patients with and without shoulder complaints, who could either be in the RF or no RF health states (Figure 2).

Prediction model 3 - Survival

We found relevant and validated prediction models for survival developed by a Dutch group which was suitable for our decision model.[168,169] We collaborated with this group to include two Cox proportional hazards models predicting survival for patients with and without regional failure (RF). These prognostic models were based on 1369 consecutive patients diagnosed with Head and Neck Squamous Cell Carcinoma (HNSCC) between 1981 and 1999. The yearly survival was predicted over 10 years in both models. Age at diagnosis and clinical tumor stage (cTNM) were taken into account in the original model.[168] Levels of co-morbidity (coded by the Adult Comorbidity Evaluation-27 (ACE27)) were added as a prognostic factor in an updated version of the model.[169,170] Internal validation of the updated model, including outcomes on patients with and without regional failure, showed a Harrell's Concordance Index of 0.76. A subset of this model was included in the Markov model to determine the probability of dying each year from the RF and no RF state for patients with early stage OCSCC (Figure 2). Details about this model are listed in appendix C.

<u>Other input decision model</u>

Input for parameters for which no prediction models were obtained was equal to our previous decision model.[127] Probabilities of having shoulder complaints were 32% after WW, 36% after SLN and 43% after (E)ND.[167] Patients with a positive SLN (strategy 3 and 5) received an ND and were assumed to have a probability of shoulder complaints of 43%. When the result of the GEP test was negative, the same probability of shoulder complaints as for the WW was assigned. A total overview of the input, including the average predictions of the prediction models, is listed in appendix D.

<u>Analysis</u>

For each of the five management strategies the expected costs and QALYs were simulated. The most cost-effective strategy was selected as the optimal strategy. To select the most cost-effective strategy a willingness to pay of €80,000 per QALY gained was used. For this purpose we simulated the costs and effects for 500,000 patients. Three approaches to select the most cost-effective strategy for patients were compared.

In the first approach, the population-based approach, all patients received the same strategy (of the five possible strategies): the one that is optimal for the average population. This was determined by using the basic input with average values (e.g. a probability of 22% for each patient to have occult metastases). The average costs and effects associated with this optimal strategy were calculated over all 500,000 patients.

In the second approach, the 'perfectly predicted' approach, each patient received the strategy that leads to optimal expected outcomes in that individual patient. In this approach we assumed that the optimal strategy for every individual patient could be perfectly predicted, i.e. a theoretical approach in which the ratio between observed and predicted was always 1 for each probability. For instance we assumed that we could perfectly predict which 22% of the patients had occult metastases and which patients belonged to the other 78% that did not have occult metastases. Again, we calculated the average costs and effects per patient when giving all 500,000 patients the optimal strategy.

In the third approach, the prediction model approach, each patient received the strategy that was deemed optimal based on the prediction models that were implemented in the decision model. For this prediction model approach patients with certain characteristics were simulated in the model. These characteristics were linked to the prediction models. Patients (including disease) characteristics were simulated based on a database of 248 patients from the Radboudumc, which was a small expansion on the database that was used to develop the prediction model for occult metastases. The ACE scores were not available in this database. The distribution of ACE scores was obtained from the database that was used for the survival model. The different distributions are presented in Table 1. Again, average costs and effects for the optimal strategies were calculated for the simulated patients. Similarly, we assessed the value of each of these prediction models separately, by using a fixed value (the average population value) for the other parameters.

Table 1 Distributions of patient characteristics of 248 patients from the Radboudumc database

Characteristic	Type of distribution	Values
Gender	Table distribution	Male = 0.617 Female = 0.383
Age	Normal distribution	Mean = 59.8 SD = 12.8
ACE-score*	Table distribution	ACE 1 = 0.45 ACE 2 = 0.25 ACE 3 = 0.21 ACE 4 = 0.09
Clinical T classification	Table distribution	T_1 = 0.302 T_2 = 0.698
Vasoinvasive growth	Table distribution	No = 0.685 Yes = 0.315
Spidery growth	Table distribution	No = 0.665 Yes = 0.335
Tumor depth of invasion	Gamma distribution	Mean = 7.9 SD = 5.1

*Obtained from the database used for the survival model All other distributions were based on 248 patients from the Radboudumc database

We compared the average costs and QALYs of the population-based approach with the perfectly predicted approach and the prediction model approach. The population-based approach was also compared with the prediction model approach for each of the three prediction separately. The difference between the perfectly predicted approach and the population-based approach indicates the maximum value of personalized care with respect to management of the neck in OCSCC patients. The difference between the prediction model approach and the population-based approach represents the value of personalized care based on the three prediction models together, or based on one of the prediction models, independently.

Results

The results of the three approaches are presented in Table 2. The SLN strategy was on average the most cost-effective strategy. On average, the SLN strategy resulted in 4.9158 QALYs and costs €8,675 per patient, which represent the results of the population-based approach. In the perfectly predicted approach each patient will receive his or her optimal treatment under the assumption that the optimal strategy for each patient could be perfectly predicted. This approach resulted in 5.1252 QALY per patient and €7,651 in costs, on average. Therewith, 0.21 QALYs were gained and €1,024 was saved per patient if management decisions could be perfectly predicted, compared to the population-based approach. These increments represent the maximum added value of personalizing care.

With the prediction model approach the SLN strategy was predicted as the optimal strategy in 78% of the patients while END and WW were both predicted as the optimal strategy in 11% of the patients. This resulted in an average effect of 4.9172 QALYs per patient and average costs of €8,523 per patient. This implies that the effects increased with 0.0014 QALYs per patient, while costs were reduced with €152 per patient, which represents the value of personalizing care base on the included prediction models.

Table 2 also shows the value of the separate prediction models, which represents the value of decision-making using each prediction model separately. The prediction model of the chance of occult metastases reduced the costs per patient with €136 and increased the expected QALY with 0.0004 per patient. The survival model slightly reduced the expected QALY per patient with 0.00002 and reduced expected costs with €17. The utility model slightly increased the expected QALY with 0.0002 and increased the expected cost with €5 per patient.

Table 2. Results of the three approaches (per patient)

Approach	Costs (€)	Effects (QALY)	Difference in QALYs compared to population approach	Difference in costs (€) compared to population based approach
Population based approach (sentinel lymph node strategy)	8,675	4.9158		
Perfectly predicted approach	7,651	5.1252	0.2094	-1024
Prediction model approach: all models	8,523	4.9172	0.0014	-152
Prediction model approach: occult metastases	8,539	4.9162	0.0004	-136
Prediction model approach: utility values	8,670	4.9160	0.0002	5
Prediction model approach: survival	8,658	4.9158	-0.00002	-17

Discussion

This study showed how prediction models can be incorporated in decision models, and assessed the value of personalized decisions regarding the management of the neck in early stage cN_0 OCSCC. In the perfectly predicted approach, 0.21 QALYs were won and €1,024 was saved per patient as compared to a population approach in which each patient receives the same treatment. This shows that there is considerable value of personalized care with respect to management of the neck in OCSCC patients. Selecting a strategy based on the prediction models yielded only 0.0014 extra QALYs and saved €152 per patient as compared to the population-based approach. The model for the chance of occult metastases was the most influential of the three prediction models.

The main strength of this study is that it combines multiple prediction models to assess the potential value of personalized care. We therewith combined existing methods to acknowledge heterogeneity in patients, i.e. predictions models, with methods to calculate the value of personalized care.[171-173] This provides insight in the actual efficiency of the efforts in personalizing healthcare.[57]

Some limitations of this study should also be discussed. First, the prediction models have not been validated in other populations, since this was not the aim of our study. If the models have sufficient value for improving clinical decisions, it is worthwhile to perform such an external validation.[157] Second, we did not incorporate the imprecision of the prediction models in our analysis. Hence, the consequences of making a wrong prediction were not taken into account. This could, however, be taken into account by implementing the uncertainty of the prediction models using probabilistic sensitivity analyses.[149] This makes the model more complex because variability and uncertainty are combined and require separate simulations.[174] As the main purpose of our study was to illustrate the combination of decision and prediction models, we decided to keep the model as simple as possible. Third, characteristics of the simulated patients such as age, tumor stage and depth of invasion were simulated independently, whereas it is known that for example stage and depth of invasion are related. This may have caused an overestimation of the added value of the prediction model approach. Because the model only includes early stage tumors, this overestimation is expected to be small. Fourth, we used the same utility values for each year that patients had shoulder complaints after an operation. In reality it is more likely that shoulder complaints cause more disutility in the first period after the operation.[25] This could have an influence on the optimal strategy because this implies a more important trade-off between quality of life (utilities) and survival in patients with low expected survival. For example, a high disutility in the first year after surgery might not be worth a slightly higher expected survival. Taking into account these different disutility values over time, by making a model with year to year predictions of utility values, could have caused a higher added value of the prediction model approach compared to the population approach, especially for the utility and survival prediction models. Further research is needed to obtain these data and make the model more precise.

Despite these limitations this model is an important step towards evidence-based and personalized decision making. With regard to the case of neck management in OSCC patients, the perfectly predicted approach showed that there is a considerable potential value of personalizing care. Although it is unrealistic that this full potential can be reached, it seems worthwhile to invest in personalizing OCSCC care. Since the prediction model approach, with the currently available prediction models, showed to have low added value for personalized decision making, more information should be obtained about the influence of patient heterogeneity in the field of OCSCC to expand the prediction models and improve the personalized decision model. For instance, the models for utility values had limited value in their current state. Since predicted utilities should be weighed against predicted survival, both the utility models and the survival model may be more valuable when utilities can be predicted over time. The model for prediction of chance of occult metastases showed the highest added value. It seems worthwhile to further develop this prediction model. There are indications that some characteristics in the occult metastases model, i.e. vasoinvasive growth and spidery growth, could only be reliably assessed after excision of the primary tumor. This would imply a different use of this model in practice, and should therefore be taken into account in further development of this prediction model.

More general, including multiple prediction models could be a basis for clinical decision support systems which often have shown to have added value in clinical practice.[175] When a solid model is developed this could be automatically updated when new information from patients becomes available. By using decision models we could also assess the potential value of the individual prediction models and therewith we can decide whether it is worthwhile to further develop and validate one or multiple prediction models or to start using them in clinical practice. With regard to research, an individual decision model could therewith be used as an informative method before the actual value of a prediction model is evaluated in clinical practice by prognostic impact studies.[176]

In this study we used cost-effectiveness as the criterion for the optimal strategy. It is also possible to use the most effective (highest gain in QALYs) as the optimal strategy. The question is which criteria should be used in personalized decisions in general and in this particular case. Both the added value of the perfectly predicted approach and the prediction model approach would be different when considering the strategy with the highest effect as the optimal strategy. This is an important consideration as we continue developing personalized decision modeling. However, when simulating costs as well as effects, the definition of optimal care can be easily changed.

In conclusion, personalizing the management of the neck in early stage OSCC could improve health outcomes of patients and reduce costs. However, decision-making on a personalized level based on the combination of currently available prediction models does not seem to be sufficiently valuable to be used in clinical practice. Prediction models related to the neck management in OCSCC should be further developed. More generally, incorporating prediction models in decision models appears to be a valuable method to inform decisions on a personalized level and to assess the value of this personalized decision making.

Appendix A: prediction of chance of occult metastases.

A multi variable logistic regression model was developed to predict the presence of occult metastases (i.e. metastases present or metastases not present). This model was based on 235 patients diagnosed with OCSCC in the Radboudumc (Nijmegen, The Netherlands) between January 2000 and January 2013. These patients were treated surgically for the primary tumor and underwent an elective neck dissection. The predicted outcome (presence of occult metastases) was defined as the pN stage (pN_+ or pN_0) after pathological examination of the neck specimen. In this population 58 (25%) patients had occult metastases. The information was obtained from registry data of the Radboudumc. Table 1 shows the variables (predictors) in the prediction model together with the coefficients and levels of significance. Also the concordance index (c-index) is presented. Vasoinvasive growth, spidery growth and tumor depth of invasion was determined on the excision specimen of the primary tumor. Stage was determined after diagnostic work up. Table 2 shows the frequencies of the included variables in the population on which the prediction model was developed.

Table 1 prediction model occult metastases

Variables	Beta-Coefficient	Level of significance (Wald)	Odd ratio (95% CI)
Vasoinvasive growth (reference = no vasoinvasive growth)	0.925	0.006	2.522 (1.312; 4.850)
Spidery growth (reference = no spidery growth)	0.989	0.003	2.690 (1.408; 5.137)
Clinical T classification (reference = T_1)	0.056	0.076	2.115 (0.929; 4.815)
Tumor depth of invasion (per mm)	0.749	0.074	1.058 (0.994; 1.126)
Constant	-2.868	<0.0001	

C-index = 0.725

Range of predicted values: 0.057 to 0.76

Table 2 Frequencies of population of occult metastases prediction model

Variables	Frequencies
Vasoinvasive growth	No = 158 (67%) Yes = 77 (33%)
Spidery growth	No = 162 (69%) Yes = 73 (31%)
Clinical T classification	T_1 = 71 (30%) T_2 = 164 (70%)
Tumor depth of invasion (mm)	Mean = 7.9 Sd = 5.2

Appendix B: Prediction of utility values

Linear multivariable regression models were developed to predict utility values which represent the quality of life in patients with early stage OCSCC. The models were based on patients who were treated in the Radboudumc or the Netherlands cancer institute between 2001 and 2013 with different strategies for the neck (WW, SLN procedure or ND). One model was developed to predict utility for patients with shoulder complaints (n=68) and one model for patients without shoulder complaints (n=83). Information was obtained from a cross-sectional survey study to measure quality of life in these patients. In this study patients were sent questionnaires to measure shoulder complaints (Shoulder disability questionnaire) and health utility (EQ-5D-3L).
Table 1 (patients with shoulder complaints) and Table 2 (patients without shoulder complaints) show the variables (predictors) in the prediction model together with the coefficients and levels of significance. Also the level of explained variation (R-square) of both models is presented. Table 3 and 4 show the frequencies of the included variables in the population on which the prediction models were developed.

Table 1 prediction model of patients with shoulder complaints

Variables	Beta-Coefficient	Level of significance
Gender (reference = male)	-0.085	0.138
Age (per year)	-0.004	0.082
Constant	1.072	<0.0001

R-square 0.075

Range of predicted values: 0.619 to 0.968

Table 2 prediction model of patients without shoulder complaints

Variables	Beta-Coefficient	Level of significance
Gender (reference = male)	-0.047	0.168
Age (per year)	-0.001	0.279
Constant	0.997	0.279

R-square 0.040

Range of predicted values 0.858 to 0.971

Table 3 Frequencies model patients with shoulder complaints

Variables	Frequencies
Gender	Male = 40 (59%) Female = 28 (41%)
Age	Mean = 64 Sd = 12

Table 4 Frequencies model patients without shoulder complaints

Variables	Frequencies
Gender	Male = 57 (69%) Female = 26 (31%)
Age	Mean = 64 Sd = 12

Appendix C: prediction survival model

A Cox proportional hazard model was obtained from a Dutch group.
The model was based on 1369 patients with Head and Neck Squamous Cell Carcinoma (HNSCC) from the Leiden University Medical Center between 1981 and 1999.
We used a subset of this model to predict survival for patients with early stage (cT_1-T_2) N0M0 OCSCC. Survival was predicted for 10 years both for patients with and without regional failure. The model included a factor y, which was built out of 3 predictors: Stage, age and co-morbidity (ACE-stage).

With regional failure: Y=-2.009+0.3229*Stage+0.0359*age+ACE-stage+0.57575

Without regional failure: Y=-2.009+0.3229*Stage+0.0359*age+ACE-stage

With:
Stage: T_1=0 and T_2=1
Age: Age in years
ACE-stage: ACE1=0, ACE2=0.0847, ACE3=0.3577 and ACE4=0.44

Factor Y is translated to 10 year survival probabilities with:

Individual survival = average survival ^ EXP (y-2.499908291)

With average survival (Table 1):

0	1.0000
1	0.8568
2	0.7474
3	0.6895
4	0.6255
5	0.5652
6	0.5211
7	0.4827
8	0.4263
9	0.3856
10	0.3530

Harrell's C: 0.7

Table 2 Frequencies of population of survival

Variables	Frequencies
Gender	Male = 1088 (79%) Female = 283 (20.6%)
Age	Mean = 63 Sd = 12
Tumor Location	Lip = 123 (9.0%) Hypopharynx = 280 (20%) Oral cavity = 152 (11%) Oropharynx = 41 (3%) Glottic larynx = 137 (10%) Supraglottic larynx = 442 (32%) Nasopharynx = 196 (14%)
Clinical T stage	T_1 = 516 (38%) T_2 = 369 (27%) T_3 = 208 (15%) T_4 = 278 (20%)
N classification	N_0 = 964 (70%) N_1 = 145 (11%) N_2 = 180 (13%) N_3 = 82 (6%)
M classification	M_0 = 1354 (99%) M_1 = 17 (1%)
Comorbidity	ACE27 Grade 0 = 845 (62%) ACE27 Grade 1 = 251 (18%) ACE27 Grade 2 = 193 (14%) ACE27 Grade 3 = 82 (6%)
Recurrent disease	Regional = 72 (5%)

Appendix D: input

D1: Transition probabilities

Parameter	Estimate	Source
Percentage occult metastases	0.22	Prediction model of occult metastases*
Sensitivity gene expression profiling	0.86	Van Hooff et al.(116)
Specificity gene expression profiling	0.45	Van Hooff et al.(116)
Sensitivity sentinel lymph node procedure	0.93	Govers et al.(65)
Specificity sentinel lymph node procedure	1	Govers et al.(65)
Shoulder complaints watchful waiting	0.32	Govers et al.(177)
Shoulder complaints sentinel lymph node procedure	0.36	Govers et al.(177)
Shoulder complaints neck dissection	0.43	Govers et al.(177)
Regional failure after neck dissection with occult metastases (TP)	0.42	Empirical study with 96 patients
Regional failure after WW with occult metastases (FN)	1	Empirical study with 69 patients
Survival probability with / without regional failure:		Prediction model survival*
Year 1	0.88 / 0.80	
Year 2	0.89 / 0.82	
Year 3	0.94 / 0.89	
Year 4	0.92 / 0.87	
Year 5	0.92 / 0.86	
Year 6	0.94 / 0.89	
Year 7	0.94 / 0.89	
Year 8	0.90 / 0.83	
Year 9	0.92 / 0.86	
Year 10	0.93 / 0.88	

*Average predictions from prediction models

D2 Costs

Parameter	Costs (€)	Source
Neck dissection	1915	Cost calculation Radboudumc
Salvage therapy	3077	Cost calculation Radboudumc
Gene expression profile	2675	Agendia BV
Sentinel lymph node procedure	1201	Cost calculation Radboudumc
Hospital days (included in each strategy)	6875	Cost calculation Radbodumc

9

D3 Utility values

Parameter	Estimate	Source
Utility shoulder complaints	0.92	Prediction model utility*
Utility no shoulder complaints	0.80	Prediction model utility*

*average predictions from prediction models

Chapter 10

General discussion

As described in chapter 1, the overall aim of this thesis was to inform evidence-based decisions in the management of oral cavity squamous cell carcinoma (OCSCC) patients. To achieve this overall aim, this thesis is structured around three specific objectives: 1) to evaluate the need for evidence-based surgery in general, and more particularly for diagnostic and surgical innovations in OCSCC; 2) to assess the cost-effectiveness of various management strategies for the neck in early-stage OCSCC patients; and 3) to bridge the gap between evidence-based decision-making and personalized healthcare in the management of the neck in OCSCC patients. In this discussion, I will outline the main findings and then put the results into a more general perspective. Furthermore, since the overall aim was to inform evidence-based decisions, I will discuss three aspects related to decision modeling which will be relevant for future incorporation of decision modeling into evidence-based decision-making. Finally, specific recommendations will be given to further enhance these aspects.

Main findings

As to the need for evidence-based surgery in general, and more particularly for innovative surgical procedures and diagnostics in OCSCC (*objective 1*), we showed that health care spending could be reduced by no longer performing lower-value surgery. We estimated that approximately 250 million Euros a year could be saved by no longer performing lower-value surgical procedures in the Netherlands. With regard to OCSCC surgical care, the management of the neck shows a considerable variation between centers and countries. For instance, the relevant parties in OCSCC management did not agree on the necessity to use the sentinel node procedure regularly in OCSCC patients with a clinically negative neck (cN_0). As a result, the use of this procedure varied between centers and countries. Variations in practice and rising health care costs indicate a lack of clear evidence and the need for evidence-based surgery. Decision modeling could provide us with information on the added value and cost-effectiveness of innovations and procedures. As a result, we may reduce variation and health care spending and improve patients' health.
To assess the (cost-)effectiveness of various management strategies for the neck in OCSCC patients (*objective 2*), we performed a diagnostic meta-analysis showing that the sentinel lymph node biopsy (SLNB) procedure has a sensitivity of 93%. This outcome was used as an input value for our decision analytic model, which compared the costs and effects of five available strategies for the cN_0 neck in early-stage OCSCC patients. This evaluation showed that the SLNB procedure was the most effective and cost-effective strategy. However, the model was sensitive to changes in assumed quality of life (expressed in utility values). It showed that further information on quality of life in this population was very useful. We therefore studied the quality of life after different diagnostic and treatment modalities for the neck and we found relatively high utility values after a SLNB procedure was performed. When using the updated utility values for our decision-analytic

model, SLNB remained the most (cost-)effective strategy. Another modeling study showed that selective neck dissection was not cost-effective compared to modified radical neck dissection in early-stage patients with a cN_+ neck and singular node disease. However, sensitivity analysis suggests that there may be subgroups for whom selective neck dissection is cost-effective. Overall, decision-analytic modeling proved to be a valuable method to assess the (cost-)effectiveness of various management strategies for the neck in OCSCC patients.

To bridge the gap between evidence-based decision-making and personalized healthcare in the management of the neck in OCSCC patients *(objective 3),* we incorporated three prediction models into the decision model for the cN_0 neck. Separate prediction models predicted the probability of a patient having occult metastases, survival, and the quality of life in patients with and without shoulder complications. This allowed us to calculate the costs and effects of the different management strategies taking into account the patient and disease characteristics instead of providing average population-level outcomes. In addition, we were able to calculate the value (expressed in health benefits and health cost savings) of making more personalized decisions in the management of the neck in OCSCC patients. Incorporating prediction models into a decision-analytic model seems to be a promising method to bridge the gap between evidence-based decision-making and personalized healthcare.

Decision modeling and evidence-based decisions

The clinical example of the management of the neck in patients with early stage OCSCC showed that decision modeling can be useful in providing a more evidence-based insight into decision-making regarding clinical issues. The decision model offers a framework for linking relevant evidence regarding a specific clinical issue, allowing important assessments of such matters to become transparent. As a result, more specific choices can be made. For instance, a wider use of the sentinel node procedure is currently being discussed by the Dutch Group of Head and Neck Tumors (NWHHT) partly as a result of the findings described in this thesis. Furthermore, these decision models can offer guidance for further research as they provide insight into what additional information is of great value, similar to the information provided on quality of life in OCSCC patients after undergoing various modalities for the neck. I have also shown opportunities to support patient-level decisions by using decision models. Our examples demonstrate that decision models can contribute to evidence-based decision-making within the field of head and neck surgery. Our examples focused on supporting clinicians in evidence-based decision-making regarding new procedures. However, decision modeling in the field of surgery seems to be infrequently used to actually support decisions. In order to develop a more evidence-based approach to surgery we must see to it that methods are being developed further, allowing strategies to be evaluated in an evidence-based manner. We should also see to it that these evaluations are actually used in the decision process. During my doctoral research activities I discovered three aspects which in my opinion play a role in the limited use of decision modeling: the way the models and results are presented to the intended user, the current evaluation phase of the innovations, and the ever-recurring area of tension between average population-based outcomes and the treatment of individual patients. In the following sections, I will examine each of these three aspects further.

Presentation of the model and results

The aim of decision modeling is to provide relevant stakeholders with information regarding the consequences of different options when a decision is being made. In order to achieve this result, it is necessary for them to understand the model and therefore trust the subsequent results produced.[177] However, methodological developments in the field of decision modeling lead to a higher level of complexity. As a result, results are increasingly difficult to interpret.[178,179]

It is important that the model and results are presented well to the people who need to make the ultimate decision. Currently, decision models are most frequently used for economic evaluations to support decisions on reimbursement for medicines. Reasons for this include the way access for different types of care has been regulated. A so-called closed description applies to drugs, which means the minister is allowed to make additional demands for admission regarding efficiency, i.e. manufacturers have to provide information regarding the cost-effectiveness of a new medicine. For other forms of care, including surgery, an open description applies in which no clear demands are being made regarding cost-effectiveness.[180] For this reason, cost-effectiveness does not seem to play an important role in the Netherlands (and most of the other countries) in deciding whether or not to offer reimbursement for these other forms of care.Therefore, clinicians have more influence on the use of medical interventions not involving drugs, such as new surgical procedures and medical devices. In general, they will make decisions in daily practice, for instance when deciding whether a procedure will be actually used. Furthermore, clinicians are the main link in the development and evaluation of new procedures. In the field of surgery, clinicians therefore seem to be a relatively important target group to focus the decision evaluation regarding new procedures on. It should be mentioned that clinicians will initially be interested in the effectiveness of the procedures. However, it is also possible to only study effectiveness by using a decision model. In order to create not only better but also affordable care it is important for clinicians as well to be aware of the necessity to weigh costs and effects against each other and therewith to provide clinicians with information on this. Cost-effectiveness may play a future role in decisions on paying for non-drug-related interventions.[180] Especially with regard to surgical innovations (new techniques, modified strategies or innovative instruments), which can often be used in various ways, it is important for clinicians to stay involved in modeling studies in order to ensure they are relevant and the results are implemented into practice in the right manner.

The target group will have to understand the model and results in order to actually use them. Reporting checklists have been developed to make decision modeling more transparent and reduce ambiguity.[177,181-183] One of these checklists, the joint guideline from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Society for Medical Decision Making (SMDM), is specifically designed to report decision modeling in a transparent manner.[177] This guideline comprises a list of fourteen aspects, which need to be described when presenting model-based economic evaluations. Furthermore, they indicate that in a number of cases the model itself (computer code or program) could be added in order to give more insight into the model. However, the recommendations are rather concise. For instance, although it is mentioned what information should be given (i.e. information on input, output, model validation and effects of insecurity), it is not mentioned how this information should be described. The way in which this information is described, will determine whether relevant stakeholder will truly understand it.

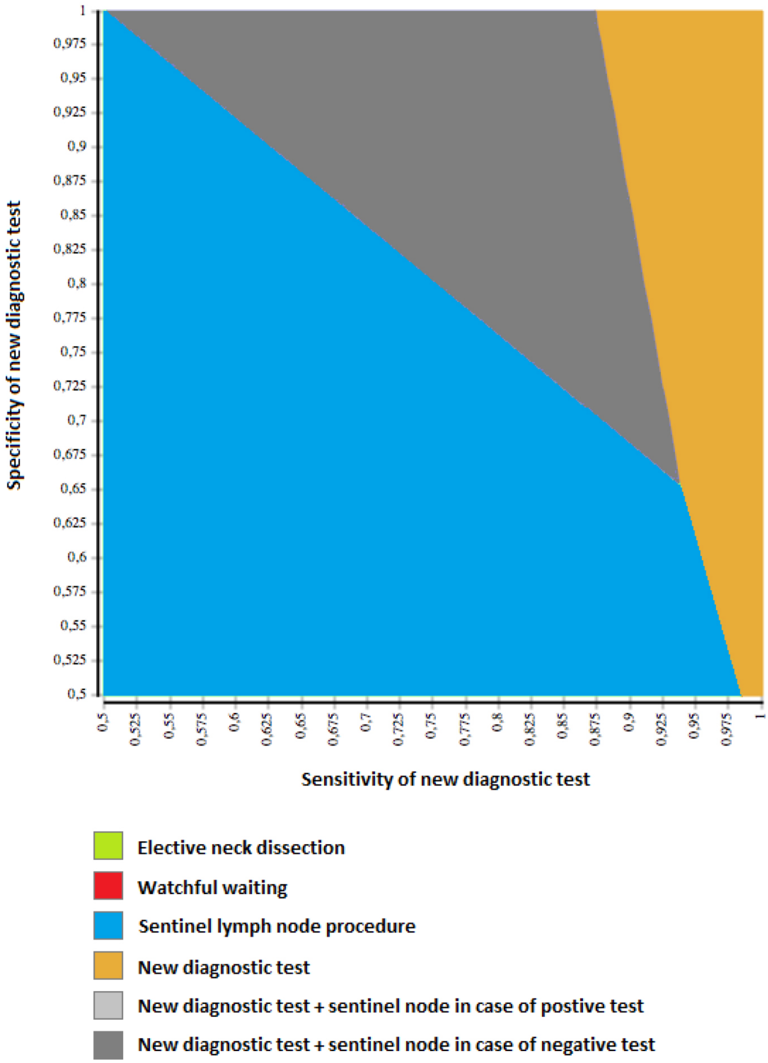
Although we used the ISPOR-SMDM guideline for reporting our articles, it was shown that the information provided did not always fit the clinicians' needs. Clinicians expressed their wish to have a better understanding of how the model works and how certain parameters and assumptions affect the results.

<u>Evaluation phase</u>

Currently, surgical innovations and devices are hardly evaluated. In general, if an evaluation does take place, the innovation under evaluation is already being used frequently.[3,4] The IDEAL collaboration has defined various phases of surgical innovation with recommendations regarding evaluations during those phases. The real evaluations only begin when the technique is already widely used.[6] In the initial phases, a description of the procedure and a report of complications, among other things, are recommended in particular. Potential (cost-)effectiveness of the new procedure is left aside. When new equipment (medical aids) is being purchased, which is usually a hospital-level decision, the (potential) added value from a therapeutic point of view and the effectiveness of the intervention are rarely a point of consideration.[184] For instance, the da Vinci robot has been purchased by many hospitals and has been used for an increasing number of indications without the presence of clear evidence. One of the main reasons for this type of technology being rapidly implemented is its high-tech and innovative character. Clinicians and hospitals want to remain in a leading position. The opportunities provided by the new technique in particular are presented to the public, whereas uncertainties are pushed into the background. As a result, patients tend to prefer innovative technologies as well.[185] The current policy followed by various health insurers is to only pay for robotic surgery unless it is shown to be of added value.[186]

So a new procedure is often evaluated when it has been used frequently, a lot of money and energy has been put into the development and purchase of this intervention technique, and professionals have become familiar with it. Previous studies as well as our study into costs of lower value surgery (chapter 2) have shown that common procedures are not abandoned just like that, even if they seem to be of no added value.[187] Therefore, de-implementation of existing procedures has proven to be very difficult.[187-190] A report written by the Health Council of the Netherlands mentions several reasons for difficulties regarding de-implementation, including resistance to change in general and the occasional effectiveness of a procedure in a small group of patients despite the absence of added value for the majority of patients. Due to the previous-mentioned factor, the public opinion in particular may interfere with de-implementing procedures.[184]
An early evaluation of procedures and devices with regard to whether they have added value may offer a solution and avoid the need for de-implementation. An early evaluation preferably started in the development phase may be able to stop or adjust the development, because there are fewer reasons for resistance if a procedure has not yet been widely implemented. These early forms of evaluations are called early Health Technology Assessments (HTAs).[191] In the literature on this subject, various methods and frameworks for early HTAs are described. These evaluations are based on decision models designed to map out the potential (cost-)effectiveness of new procedures or equipment.[191-195] The headroom method is an important method which examines whether there is room for improvement within a particular indication area in which the innovation technique can be used, showing whether it is useful to invest in (further) development of this particular innovation technique.[194,196] A study conducted by Chapman et al. shows that the headroom method has a predictive value for medical devices regarding future inclusion by Britain's National Health Service, which does consider cost-effectiveness to be an important criterium.[197] When a procedure or innovation is being developed, decision modeling can provide helpful information to make decisions regarding further development, research and implementation, depending on the development phase.[193, 194] An example of the opportunities offered by early HTA can be showed with the case of the gene expression profile, evaluated in chapter 6. This evaluation showed that the gene expression profile with its currents characteristics is not cost-effective. Despite the gene expression profile being in a relatively early stage of development, time and money had already been spent on it. For instance, a validation study had been carried out after its development and initial accuracy tests.[116] After the initial tests a decision model could have examined in an objective manner whether the gene expression profile was of potential value based on the first test results. Next, it could have been decided whether a validation study or further (technical) development of gene expression should have followed, if the gene expression profile was not likely to be cost-effective at that point based on the results.

Figure 1: Minimally required accuracy of a new diagnostic test to find potentially occult metastases in the neck in cN_0 patients (Willingness to pay €80,000/QALY)



An evaluation prior to the development could have determined whether there is room for improvement and if so, how big this room for improvement will be at most. For instance, prior to the development of the gene expression model for cN_0 neck, evaluation could have taken place regarding the minimal diagnostic value of this test in order for it to be or become cost-effective. By carrying out this type of evaluation, it can be decided whether developing such tests provide chances of improving treatment regimens. After the development and after carrying out the initial tests, it can be decided whether this minimal level of accuracy has been or can be reached. Figure 1 shows the combination of sensitivity and specificity in which a new diagnostic test would be cost-effective. In this figure we have estimated the costs of the diagnostic test at 1,000 Euros. This seems to be a realistic cost price for the gene expression profile (because the costs of these tests are going down). The figure shows that the specificity of the diagnostic test must be around 70% at least for it to be cost-effective. In that case, the sensitivity should be high, around 90% at least. Apart from the minimally required diagnostic accuracy, it can also be determined what the effects of an improved accuracy are in terms of health profit and change in costs. This information could enable the developers and clinicians involved to make better informed choices regarding research and development.

Towards personalized healthcare

In general, decision modeling involves making a comparative assessment in a population. However, doctors do not treat an average population, but individual patients to whom they want to give the best possible information on potential results and consequences of different treatment options based on their own patient and disease characteristics.
Therefore, we tried to develop a more evidence-based approach to decision-making in individual patients by using an individual model in chapter 7. The incorporation of prediction models into decision models seems to be a useful method to inform individual-level evidence-based decisions and to evaluate the value of these individual decisions. However, further steps need to be taken and several questions need to be answered before these models will and can actually be used in practice.
It is important to know how an individualized decision model can be used best in practice, if an analysis of the model will show that individualized decisions are indeed useful. It is mentioned before that decision modeling is a method aimed at making comparative assessments between different outcomes (for instance, quality of life and survival). Patients may differ in the way they value these outcomes. Furthermore, we want to support decisions regarding the individual patient. Therefore, the needs of individual patients have to be taken into account when deciding which management strategy fits their preferences.[198,199]
When making preference-sensitive decisions, it is important to have a collaborative process that allows patiens and their care providers to make health care decisions together, taking into account the best scientific evidence, as well as the patient's values. This approach is called shared decision making, and is likely to result in better professional-patient relationships, better decisions and better outcomes.[200] Shared decision-making is often facilitated by decision aids. These are tools designed to help patients make decisions regarding different options. These tools provide information on the different options available and help patients design, clarify and communicate their personal values in relation to the various options.[155,201] A limitation to the currently available decision aids is that they are often not being based on up-to-date information. Furthermore, clinicians may find the format impractical to use in consultations and may be just as unfamiliar with

risk estimates and the inherent uncertainty associated with probabilities as their patients are.[202] Decision models can offer a framework for linking relevant data and therefore seems to be useful for selection tools. Individual decision models provide specific and detailed information on comparative assessments between the likelihood of a particular outcome and the consequences of this outcome for individual patients. The question is whether this can help in the application of shared decision-making and if so, in what way. Although decision modeling offers a framework for linking relevant evidence, it remains difficult and time-consuming to collect all relevant evidence in every aspect. This goes for individual models based on numerous data in particular. If individual modeling is proven to be useful for selection tools, it remains an important challenge to keep these individual models and selection tools based on these up to date.

Shared decision-making has also implications for the evaluation of the value of individual care. In our individual model (described in chapter 9) we have not included patient preferences. However, by applying shared decision-making these preferences would be taken into account in practice when making decisions on a particular management strategy. If we evaluate the value of individual care by using individual decision models, patient preferences should be taken into account in these models as well.

Recommendations

Based on the previous discussion I would like to make a number of recommendations for future research which in my opinion will contribute to the further dissemination and implementation of decision modeling.

In order to improve transparency and the stakeholders' insight into decision models, it may be useful to examine whether stakeholders would benefit from having the opportunity to change parameters themselves and being able to see the influence on the results. In this way, they would be provided with more insight into the model, as is the case with prediction models in which apps or nomograms are being used for the presentation. This could be achieved by adding a computer-based overview of the model, for instance, in which the value of parameters can be 'shifted' by using certain panels, after which the influence on the results is directly visible. Cost-effectiveness studies have been carried out in which such overviews are added as web interface.[203,204] This web interface provides the user with an overview of all relevant input. The value of this input can be changed within set ranges in order to see what influence a certain input has on the results. In addition, different scenarios can be looked at such as differences in availability of facilities between hospitals. It should be examined which type of interface with what sort of information improve transparency towards the relevant stakeholders. This may not be the same for the different stakeholders (clinicians, policymakers et cetera). Based on this information, existing guidelines such as the ISPOR-SMDM guideline could be adjusted by giving examples of how to best describe certain information.

Not all innovations entering the market lead to better and affordable care, because most innovations are currently evaluated in a late stage. Therefore, we believe that it is important to carry out and further develop early health technology assessments in order to enable companies/developers, clinicians and other parties involved to make better informed and early choices regarding the development of new procedures.

In order to apply these early evaluations in practice at a larger scale when new surgical procedures are being developed, it would be logical to include early HTAs into the IDEAL framework for developing and evaluating surgical procedures.[6] Our group has already been discussing this with IDEAL. An important issue is legislation on the evaluation of surgical procedures and medical equipment with regard to entering the market. Although this issue has not been arranged yet, efforts are being made at a European level to draw up such legislation.[205] Hopefully, it is possible to develop legislation which allows valuable innovations to enter the market relatively quickly. This can only be achieved by having the courage to diverge from existing legislation on drugs. The incorporation of (early) HTAs into legislation can help ensure that valuable innovations continue to be developed without an increase in development and heathcare costs. For instance, in the field head- and neck oncology the development of new techniques for intraoperative evaluation of the resection margins of the primary tumor would be a candidate for an early HTA. Multiple technologies are suggested.[206,207] However their potential (cost-)effectiveness is unknown.

The claim that early HTAs allow procedures and innovations to be more efficient should be further proved by carrying out proper evaluation. We could achieve this by following the innovations which have been evaluated by carrying out early HTAs. The correspondence between the potential value of innovations and the actual cost-effectiveness after implementation as well as the relationship between the potential value and the extent to which an innovation is being used in the health care sector should be evaluated. We have already started to register the early HTAs conducted by us in order to allow this evaluation to be carried out. If a future evaluation study into early HTAs indeed shows that these kinds of assessments contribute to the development of useful innovations, the result will hopefully be a more frequent use in practice by the stakeholders.

We will also have to examine how individual models can be used in practice, if the analysis shows that individualized care is of value in a certain field. Basically all desired outcomes important to make a comparative assessment can be included into the models, such as expected risk of recurrence, survival prognosis, risk of shoulder complaints, quality of life and costs. The outcomes that are important to the patient (patient reported outcome measures) could be used to help define individual patients' preferences regarding treatment, for instance by incorporating these into a decision aid. It should also be considered how these decision aids can be kept constantly up to date. The emergence of cognitive computing, artificial intelligence and deep learning (IBM's Watson, for instance), seems to lead to many changes in this field as well.[208]

Concluding remarks

By giving a concrete clinical example of treatment of the neck in OCSCC patients, I have shown that decision modeling can be a valuable method to inform evidence-based decisions regarding surgical procedures and innovations. Increased and better use of such decision models in the future will result in better and affordable care.

Summary

Samenvatting

Bibliography

Publicaties

Curriculum Vitae

Dankwoord

Summary

Surgical innovations comprise modified strategies, new techniques, and innovative instruments. Evaluation of these innovations, e.g. in surgical trials, is difficult and rare due to practical and methodological challenges like the right timing of the evaluation, using the right comparator and the difficulties to achieve blinding. Furthermore, the introduction of new surgical procedures and devices is not regulated yet. Hence, these procedures and devices may become widely used with little evidence of their efficacy and (cost-) effectiveness. Since others have shown that only half of all new experimental treatments will prove to be better than established treatments when tested in randomized trials, many used surgical procedures and devices may not have added value for patients. At the same time technological innovation is an important driver of the increasing healthcare cost.

An efficient and evidence based health care system should enhance the ability of medical professionals, their patients, health care insurers, manufacturers, and policy makers to make informed choices about the implementation of new procedures and devices, without jeopardizing innovation. Therefore, there is a need to improve the scientific tools and methods to inform evidence based decision making. Decision analytical modeling can be an ideal starting point for clinical decision making and the development of technologies. It can provide a basis for further research and an evaluative framework for decision making. The aim of this thesis was to inform decisions in the management of early oral cavity squamous cell carcinoma (OCSCC) patients in an evidence based manner. This empirical example offers the opportunity to assess the opportunities and challenges of an evidence based approach for the evaluation of (surgical) innovation.

This thesis consists of three parts. In Part 1 we evaluate the need for evidence based surgery in general and more particularly for OCSCC. In part 2 we assess the cost-effectiveness of various management strategies for the neck in early stage OCSCC patients, and in part 3 we try to bridge the gap between evidence based decision making and personalized healthcare.

Part 1: the need for evidence-based surgery.

Chapter 2 provides an overview of the estimated costs associated with lower value surgery. Several initiatives have focused on defining healthcare services (including surgery) that provide little or no benefit. It is important to translate these lists of lower value services into meaningful metrics of avoidable costs to inform policy makers and providers how to bend the cost curve. To provide these metrics we calculated avoidable costs associated with surgical procedures included in the ‘do not do’ recommendations of the National Institute for Health and Clinical Excellence. We used both indication specific Dutch utilization data and more general utilization data from England in combination with actual costs from cost-(effectiveness) studies and reimbursement tariffs. In addition, we extrapolated the

results to all potential lower value surgical procedures, varying the assumption of lower value surgery to be between 5 and 33% of all surgical procedures. The estimated costs that could be saved by not performing proven lower value surgery in the Netherlands were €11 million and €8 million per year using actual costs and reimbursement tariffs, respectively. An extrapolation of this result to all surgical procedures showed a potential cost saving that varied between €63 million and €419 million per year. Data from England showed cost savings of about 48.5 million using actual costs. Extrapolation of the data from England showed a potential cost saving between €382 million and €2.5 billion. Therewith chapter 2 shows that lower value surgical procedures result in a significant waste in healthcare.

Chapter 3 reports on the variation in practice with regard to management of the neck in OCSCC patients within and between countries. Such practice variation can be an indication of a lack of evidence-based management. The practice information was obtained by questionnaires, which were sent to representatives of head and neck cancer centers (HNC) in the Netherlands, UK, and USA. In total, representatives of 45 HNC centers completed the questionnaire; 10 from the Netherlands, 26 from the UK and 9 from the USA. Our results demonstrate a distinct variation in the management strategy of OCSCC, both within and between countries. Variation was found in the diagnostics, treatment and follow up of patients with OCSCC. Only a small part of this variation could be linked to differences in guidelines and differences in health systems, which indicates that the absence of clear evidence about optimal management strategies in OCSCC plays a role in the variation.

Chapter 4 describes the results of an interactive evaluation in the field of HNC. An interactive evaluation involves anonymously exchanging concerns, claims and issues between all relevant stakeholders during interviews. It is aimed at systematically eliciting the views per persons involved, particularly their norms and values, and to identify shared values. It seeks to develop a shared view on where technology should develop, what are desirable scenarios (for research and development), and what are appropriate outcome measures for assessment. We asked the stakeholders about their expectancies for future OCSCC management options, taking into account the opportunities that the new state of the art operating theatres that are being built in the Radboudumc should offer. Twelve stakeholders in the field of head and neck cancer, i.e. doctors, a nurse, patients and paramedicals were included in the interactive evaluation. Most stakeholders agreed that the time interval between diagnosis and treatment is too large in some cases. Communication and exchange of information across disciplines should be improved. Stakeholders also agreed that research regarding the merits of the sentinel lymph node biopsy (SLNB) procedure and research on resection margins should be performed. Chapter 4 showed important and shared targets for future research and development in the field of head and neck cancer. Interactive evaluation was demonstrated to identify and prioritize viable innovations in head and neck cancer care.

Part 2: cost effectiveness of management strategies for the neck in OCSCC patients.

Chapter 5 shows the results of a diagnostic meta-analysis of the accuracy of a sentinel lymph node biopsy (SLNB) for the prediction of neck node involvement in early stage OCSCC. A systematic search was performed in electronic databases to identify all relevant articles on the diagnostic accuracy of the SLNB. On basis of this search 21 diagnostic accuracy studies (including 847 patients), which used neck dissection or follow up as a reference test could be included. The pooled results of these 21 studies showed an overall sensitivity of 93% and a specificity of 100%. This high sensitivity and specificity supported a role of SLNB in the diagnostic work-up of OCSCC.

Chapter 6 describes the results of a decision analytical model assessing the cost-effectiveness of five management strategies for possible occult lymph node metastases in the clinically negative neck (cN_0) in early stage OCSCC. The five strategies were (1) elective neck dissection (END), (2) watchful waiting (WW), (3) gene expression profiling (GEP) followed by neck dissection (ND) or WW, (4) sentinel lymph node biopsy (SLNB) procedure followed by ND or WW, and (5) GEP and SLNB (for positive GEP) followed by ND or WW. The cost effectiveness of a strategy was expressed in terms of costs, Quality Adjusted Life Years (QALYs) and incremental cost-effectiveness ratios, which show the extra costs of a strategy to gain a QALY as compared to the next best strategy. The model showed that the SLNB procedure followed by ND or WW was the most effective and cost-effective strategy. Compared with END the incremental cost-effectiveness ratio for SLNB was €3356 per QALY gained, which is lower than the often used willingness to pay of €80,000 per QALY. The model was, however, sensitive for changes in quality of life (expressed in utility values), suggesting that further information on quality of life in this population was highly valuable.

In *chapter 7* we measured this quality of life in terms of health utilities in patients with early stage OCSCC following different diagnostic and treatment modalities for the neck .Patients with oral cavity cancer who underwent WW, SLNB, END (dissection of levels I-III), or therapeutic modified radical neck dissection (MRND; dissection of levels I-V) were included in this cross-sectional survey study. Patients received the EuroQol-5D-3L and the shoulder disability questionnaire to measure health utilities and shoulder complaints. A total of 181 patients (62%) returned the questionnaires. Mean health utilities were 0.804, 0.863, 0.834 and 0.794 for the WW, SLNB, END and MRND subgroups, respectively. Mean shoulder disability scores (higher scores mean more shoulder complaints) for these groups were 8.64, 10.57, 18.92 and 33.66. Patients with shoulder complaints had a mean utility of 0.78 while patients without shoulder complaints had a mean utility of 0.90. When we imputed these utilities into our previously constructed decision analytical model (chapter 6), the strategy including SLNB remained the most (cost-)effective strategy and was even more (cost-) effective than calculated in chapter 6.

Chapter 8 presents the results of a study in which we compared the cost-effectiveness of selective neck dissection (SND), consisting of the dissection of levels I-III or I-IV, with modified radical neck dissection (MRND), which consists of dissection of levels I-V, in patients with early stage OCSCC with singular nodal disease confined to level I or II. For this purpose a decision analytical model was developed to model QALYs and costs of SND and MRND in the target population. The SND strategy resulted in an expected health loss of 0.06 QALY and savings of €1351 per patient compared to MRND. Overall, based on currently available evidence the SND was not cost-effective compared to MRND. However, sensitivity analyses revealed that there could be subgroups within the target population for which SND is cost-effective compared to MRND.

Part 3: Evidence based personalized decision making

Chapter 9 shows how prediction models can be incorporated in decision models to allow more personalized evidence based decisions; and to assess the value of this approach in decisions regarding the management of the neck in early stage OCSCC. Using the decision model from chapter 6 three approaches were compared: 1) a population-based approach in which all patients undergo the strategy that is optimal for the average population; 2) a perfectly predicted approach, in which each patient undergoes the optimal strategy for that patient, and 3) a prediction model approach in which each patient undergoes the strategy that is predicted to be optimal based on a combination of predictions models (for the probability of occult metastases, survival and quality of life). The optimal strategy was chosen on the basis of a trade-off between costs and effects (quality-adjusted life years, QALYs). We studied the difference in average costs and effects per patient between these three approaches to assess the value of a personalized approach in OCSCC. The population based approach resulted in 4.9158 QALYs with mean costs of €8,675 per patient. The perfectly personalized approach yielded 0.21 more QALYs and saved €1,024 per patient compared with the population based approach. The prediction model approach yielded 0.0014 more QALYs and saved €152 per patient compared with the population based approach. The results of the perfectly personalized approach show that developing a more personalized approach appears to be promising in early OCSCC. However, the prediction model approach shows that decision-making based on the combination of currently available prediction models has only limited added value, probably due to the limited accuracy and discriminative power of these models. On a more general level, incorporating prediction models in decision models appears to be a valuable method to inform decisions on a personalized level and to assess the value of this personalized decision making.

In *chapter 10* the main findings of this thesis are summarized followed by a discussion regarding the use of decision modeling to inform evidence-based decisions. Three particular aspects were highlighted. First, the transparency of decision modeling is discussed. We developed our modeling studies for and with clinicians since we believe that they will take the important decisions with respect to new surgical procedures and/ or medical devices. Due to new and more sophisticated methodologies, modeling studies are becoming more complex which reduces transparency. To increase the use of decision modeling for evidence based decision making, the transparency of the models should be improved. Second, the assessment of surgical procedures and medical devices is currently often performed when the procedure is already widely used. De-implementation of procedures has proven to be very difficult. We demonstrated that the assessment of potential value of innovations at early stages of development is possible and informative. Third, incorporating prediction models in decision models appears to be a valuable method to inform decisions on a personalized level. We should evaluate how these models could be used in clinical practice, for instance by incorporating them in decision aids.

By giving a concrete clinical example of treatment of the neck in OCSCC patients, I have shown that decision modeling can be a valuable method to inform evidence-based decisions regarding surgical procedures and innovations. Increased and better use of such decision models in the future will result in better and affordable care.

Samenvatting

Chirurgische innovaties omvatten aanpassingen in chirurgische strategieën, nieuwe technieken en innovatieve instrumenten. De evaluatie van deze innovaties, bijvoorbeeld in chirurgische trials, is komt weinig voor en is ingewikkeld. Dit komt met name door praktische en methodologische uitdagingen, zoals het kiezen van het juiste moment van de evaluatie, het gebruiken van de juiste vergelijking en de moeilijkheden rondom blindering. Daarnaast is er nog maar weinig regelgeving rondom de introductie van chirurgische innovaties. Dit betekent dat deze innovaties breed ingezet kunnen worden zonder dat er duidelijk bewijs is over de werkzaamheid en kosten(effectiviteit). Er is aangetoond dat slechts de helft van alle nieuwe experimentele strategieën beter blijken wanneer deze worden getest in gerandomiseerde trials. Daardoor lijkt het waarschijnlijk dat een relevant deel van de gebruikte chirurgische innovaties geen toegevoegde waarde hebben voor de patiënt. Tegelijkertijd spelen deze chirurgische innovaties wel een belangrijke rol in de toename van de kosten in de gezondheidszorg.
Een gezondheidszorgsysteem dat efficiënt en 'evidence-based' is, moet ervoor zorgen dat medische professionals, patiënten, verzekeraars, de ontwikkelaars van innovaties en beleidsmakers goed geïnformeerde beslissingen kunnen maken over de implementatie van nieuwe strategieën en technieken. Aan de andere kant moet het verzamelen van deze informatie innovatie niet in de weg staan. Daarom is het belangrijk om de wetenschappelijke methoden rondom chirurgische innovaties te verbeteren. Besliskundig modelleren kan een goed startpunt zijn voor het nemen van beslissingen rondom nieuwe strategieën en de ontwikkeling van nieuwe technieken. Het kan een raamwerk zijn voor het maken van keuzes en de basis voor vervolgonderzoek.
Het doel van dit proefschrift was om de besluitvorming rondom de behandeling van vroeg stadium plaveiselcelcarcinomen (PCC) van de mondholte te informeren op basis van wetenschappelijk bewijs. Deze casus geeft ons de mogelijkheid om de kansen en uitdagingen van het modelleren als basis voor de evaluatie van chirurgische innovaties te onderzoeken.
Dit proefschrift bevat drie delen. In deel 1 evalueren we de behoefte aan chirurgie op basis van wetenschappelijk bewijs, in het algemeen, en meer specifiek in de behandeling van mondholte tumoren. In deel 2 evalueren we de kosteneffectiviteit van verschillende strategieën voor behandeling van de hals bij vroeg stadium PCC van de mondholte. Tot slot proberen we in deel 3 de brug tussen het maken van wetenschappelijk onderlegde keuzes voor de populatie aan de ene kant, en individuele zorg aan de andere kant, te overbruggen.

Deel 1: De behoefte aan chirurgie op basis van wetenschappelijk bewijs

Hoofdstuk 2 geeft een overzicht van de geschatte kosten die gerelateerd zijn aan het uitvoeren van chirurgische procedures die weinig toegevoegde waarde lijken te hebben voor de patiënt. Er zijn verschillende initiatieven die strategieën binnen de gezondheidszorg in kaart brengen die weinig of geen waarde toevoegen. Het is belangrijk om deze inzichten te vertalen naar daadwerkelijke kosten die mogelijk voorkomen kunnen worden door deze strategieën niet langer uit te voeren. Dit geeft beleidsmakers en aanbieders van zorg mogelijkheden om de stijgende kosten te reduceren. We hebben daarom de kosten berekend die gerelateerd zijn aan chirurgische procedures die opgenomen zijn in de 'do not do recommendations' van het 'National Insitute for Health and Clinical Excellence' (NICE). Om te bepalen hoe vaak deze procedures nog worden uitgevoerd is gebruik gemaakt van zowel indicatie specifieke volume data uit Nederland en van meer algemene volume data uit Engeland. Deze volumes zijn gecombineerd met cijfers over zowel de daadwerkelijke kosten van deze procedures (vanuit kosteneffectiviteitanalyses) als de vergoedingen (vanuit zorgverzekeraars). Daarnaast hebben we de potentiële kostenbesparingen gerelateerd aan deze procedures vertaald naar de potentiële kostenbesparingen gerelateerd aan alle procedures die mogelijk weinig toegevoegde waarde hebben. Hiervoor hebben we aangenomen dat tussen de 5% en 33% van alle chirurgische procedures van weinig toegevoegde waarde zijn. De kosten die bespaard kunnen worden met het niet langer uitvoeren van de procedures die zijn opgenomen in de 'do not do recommendations' zijn geschat op €11 miljoen en €8 miljoen per jaar in respectievelijk daadwerkelijke kosten en vergoedingen. De extrapolatie naar alle procedures kwam uit tussen de €63 miljoen en €419 miljoen per jaar. Met de Engelse data kwamen we op een schatting van een potentiële besparing van ongeveer €48,5 miljoen per jaar in daadwerkelijke kosten. Extrapolatie van deze data liet een potentiële kostenbesparing zien van tussen de €382 miljoen en €2,5 miljard euro. Daarmee hebben we in hoofdstuk 2 laten zien dat er waarschijnlijk een significante verspilling ontstaat door het uitvoeren van chirurgische procedures met weinig toegevoegde waarde.

Hoofdstuk 3 geeft een overzicht van de praktijkvariatie in de behandeling van de hals in patiënten met mondholtetumoren, binnen landen en tussen landen. Deze praktijkvariatie kan wijzen op een praktijkvoering die niet gebaseerd is op wetenschappelijk bewijs. Om deze informatie te verzamelen zijn vragenlijsten gestuurd naar vertegenwoordigers van hoofd-halskanker centra (HNC) in Nederland, het Verenigd Koninkrijk (VK) en de Verenigde Staten (VS). In totaal hebben 45 vertegenwoordigers van HNC de vragenlijsten beantwoord; 10 uit Nederland, 26 vanuit het VK, en 9 vanuit de VS. De resultaten laten een duidelijke variatie zien in de behandeling van de hals in patiënten met mondholtetumoren, zowel binnen landen als tussen landen. Er is variatie in de gebruikte diagnostiek, de daadwerkelijke behandeling en het traject na de behandeling. Slechts een klein gedeelte van deze variatie kon worden verklaard door verschillen in richtlijnen en verschillen in de gezondheidssystemen van de geïncludeerde landen. Dit suggereert dat de afwezigheid van duidelijke informatie over optimale strategieën in de behandeling van mondholtetumoren een rol speelt in de gevonden variatie.

Hoofdstuk 4 beschrijft de resultaten van een interactieve evaluatie van de zorg voor

patiënten met hoofd-halskanker. Een interactieve evaluatie omvat de uitwisseling van zorgen, beweringen en kwesties tussen relevante belanghebbenden rondom een bepaald onderwerp. Het heeft als doel om de inzichten van de belanghebbenden, met name hun normen waarden, te achterhalen en om hieruit gezamenlijke waarden te identificeren. Het beoogt een gezamenlijk inzicht te vormen over hoe nieuwe technologie zich zou moeten ontwikkelen, wat daarin de gewenste scenario's zijn (qua onderzoek en ontwikkeling) en wat geschikte uitkomstmaten zijn bij een evaluatie van de nieuwe technologie.
Middels interviews hebben we de belanghebbenden binnen de zorg voor hoofd-halskanker gevraagd naar hun verwachtingen omtrent toekomstige behandelmogelijkheden. Daarbij hebben we meegenomen dat er in het Radboudumc nieuwe operatiekamers worden gebouwd waarin deze mogelijkheden uitgewerkt zouden kunnen worden. Twaalf belanghebbenden binnen de zorg voor hoofd-halskanker (artsen, verpleegkundigen, patiënten en paramedici) zijn geïncludeerd in de interactieve evaluatie. De meeste van deze belanghebbenden vonden dat in bepaalde gevallen de tijd tussen de diagnose en de behandeling op dit moment te lang is. Ze vonden dat communicatie en uitwisseling van informatie tussen de betrokken disciplines verbeterd zou moeten worden. Daarnaast waren ze het er over eens dat er onderzoek gedaan zou moeten worden naar de schildwachtklier oftewel de 'sentinel node' procedure en naar verbetering van snijranden. Hoofdstuk 4 laat hiermee belangrijke doelen zien voor toekomstig onderzoek binnen de hoofd-halskanker, die gedeeld worden door de belanghebbenden.

Deel 2: Kosteneffectiviteit van strategieën voor behandeling van de hals in patiënten met plaveiselcelcarcinomen van de mondholte.

Hoofdstuk 5 laat de resultaten zien van een diagnostische meta-analyse naar de nauwkeurigheid van de sentinel node (schildwachtklier) procedure (SNP) voor het voorspellen van metastasen in de hals bij vroeg stadium PCC van de mond- en keelholte. Relevante studies naar de nauwkeurigheid van de SNP werden op een systematische manier verzameld uit elektronische databases. Er werden 21 studies gevonden (met in totaal 847 patiënten) die een halsklierdissectie of actief volgen ('watchful waiting') gebruikten als de referentietest. De gepoolde resultaten van deze studie lieten een sensitiviteit zien van 93% en een specificiteit van 100%. De hoge sensitiviteit en specificiteit van de SNP wijzen richting een rol van deze procedure binnen de diagnostiek van vroeg stadium PCC van de mond- en keelholte.
Hoofdstuk 6 beschrijft de resultaten van de evaluatie van de kosteneffectiviteit van vijf verschillende strategieën voor de diagnostiek en behandeling van mogelijke verborgen metastasen in de klinisch negatieve hals (cN0) bij vroeg stadium PCC van de mondholte. De vijf strategieën waren: (1) electieve halsklierdissectie, level I-III (HKD), (2) 'watchful waiting' (WW), (3) genexpressie profilering (GEP) gevolgd door HKD of WW, (4) sentinel node procedure (SNP) gevolgd door HKD of WW, en (5) GEP en SNP (SNP bij positieve GEP) gevolgd door HKD of WW. De strategieën zijn vergeleken op kosten, voor kwaliteit

gecorrigeerde levensjaren (zogenaamde QALYs) en incrementele kosteneffectiveitsratio's. Een incrementele kosteneffectiviteitsratio laat zien wat het bij een bepaalde strategie kost om een QALY te winnen ten opzichte van een andere strategie. Uit het model bleek de SNP procedure, gevolgd door HKD of WW, de meest effectieve en kosteneffectieve strategie. De incrementele kosteneffectiviteitsratio van de SNP strategie ten opzichte van een strategie met direct een electieve HKD was €3.356 per gewonnen QALY. Dit is lager dan de veelgebruikte drempelwaarde van €80.000 per QALY. Het model was met name gevoelig voor veranderingen in kwaliteit van leven (uitgedrukt in utiliteiten) en liet zien dat meer informatie over kwaliteit van leven in deze populatie van grote waarde is voor het maken van de juiste keuzes.
In hoofdstuk 7 hebben we daarom de kwaliteit van leven (utiliteiten) gemeten na verschillende diagnostische en behandelingsmethoden voor de hals in patiënten met vroeg stadium PCC van de mondholte. Patiënten met PCC van de mondholte die oftewel WW, SNP, electieve HKD (levels I-III) of therapeutische gemodificeerd radicale HKD (level I-IV) ondergingen werden geïncludeerd in deze cross-sectionele studie. Patiënten ontvingen de EuroQoL-5D-3L vragenlijst en de 'shoulder disability' vragenlijst om utiliteiten en schouderklachten te meten. In totaal beantwoordden 181 patiënten (62%) de vragenlijsten. De gemiddelde utiliteiten na WW, SNP, electieve HKD en gemodificeerde radicale HKD waren respectievelijk 0,804; 0,863; 0,834 en 0,794. De gemiddelde scores van de 'schouder disability' vragenlijst (hogere scores betekenen meer schouderklachten) waren voor deze groepen respectievelijk 8,64; 10,57, 18,92 en 33,66. Patiënten met schouderklachten hadden een gemiddelde utiliteit van 0,78, terwijl patiënten zonder schouderklachten gemiddeld een utiliteit lieten zien van 0,90. Wanneer we de utiliteiten uit deze studie terugbrachten in ons besliskundig model rondom de cN0 hals (hoofdstuk 6) bleef de SNP de meest effectieve en de meest kosteneffectieve strategie, en bleek zelfs kosteneffectiever dan eerder berekend in hoofdstuk 6.
Hoofdstuk 8 laat de resultaten zien van een studie waarin de kosteneffectiviteit van een selectieve HKD (level I-III of level I-IV) is afgezet tegen de gemodificeerde radicale HKD (level I-V), in patiënten met een vroeg stadium PCC met enkelvoudige lymfekliermetastasen die zich beperken tot niveau I of II in de hals. Om de kosten en effecten van deze strategieën tegen elkaar af te wegen is een besliskundig model ontwikkeld. De selectieve HKD strategie resulteerde in een verwacht verlies in QALYs van 0,06 en een besparing van €1.351 per patiënt ten opzichte van de gemodificeerde radicale HKD. Gebaseerd op de beschikbare informatie was de selectieve HKD strategie daarmee niet kosteneffectief ten opzichte van de gemodificeerd radicale HKD. Echter lieten de sensitiviteitsanalyses zien dat er mogelijk subgroepen bestaan waarvoor dit wel het geval is.

Deel 3: Geïndividualiseerde beslissingen op basis van wetenschappelijk bewijs
Hoofdstuk 9 laat zien hoe predictiemodellen kunnen worden toegevoegd aan beslismodellen om zo geïndividualiseerde beslissingen te kunnen nemen op basis van

wetenschappelijk bewijs, en om de waarde hiervan te bepalen. Hiervoor hebben we de casus van de behandeling van de hals bij patiënten met vroeg stadium PCC als voorbeeld gebruikt. Met het besliskundig model uit hoofdstuk 6 als basis, zijn drie benaderingen met elkaar vergeleken: 1) Een populatie benadering, waarin alle patiënten de strategie krijgen die het meest optimaal is voor de populatie; 2) een 'perfecte voorspelling' benadering, waarin alle patiënten de strategie ondergaan die optimaal is voor die patiënt; 3) een predictiemodel benadering, waarin alle patiënten de strategie ondergaan waarvan op basis van een combinatie van predictiemodellen (voor de kans op occulte metastasen, overleving en kwaliteit van leven) is voorspeld dat deze het meest optimaal is. De optimale strategie in de verschillende benaderingen werd gekozen op basis van een afweging tussen kosten en effecten (in QALYs). De verschillen tussen de drie benaderingen werden uitgedrukt in het gemiddelde verschil in kosten en effecten per patiënt. Dit laat de potentiële waarde van het individualiseren van de zorg voor de hals bij PCC van de mondholte zien. De populatie benadering resulteerde in 4,9158 QALYs en kostte gemiddeld €8.675 per patiënt. De perfecte voorspelling benadering resulteerde in een winst in QALYs van 0,21 en een besparing van kosten van €1.024 per patiënt ten opzichte van de populatiebenadering. De predictiemodel benadering zorgde voor een winst van 0,0014 QALYs en een kostenbesparing van €152 per patiënt in vergelijking met de populatiebenadering. De resultaten van de perfecte voorspelling benadering laten zien dat er winst te behalen is door het individualiseren van de zorg bij vroeg stadium PCC van de mondholte. Echter liet de predictiemodel benadering zien dat het maken van keuzes op basis van de momenteel beschikbare predictiemodellen slechts weinig toegevoegde waarde heeft. Dit komt waarschijnlijk door de beperkte nauwkeurigheid en een gebrek aan discriminerend vermogen van de predictiemodellen. Meer algemeen lijkt het toevoegen van predictiemodellen aan besliskundige modellen een waardevolle methode om beslissingen rondom het individualiseren van zorg te evalueren.

In hoofdstuk 10 worden de belangrijkste bevindingen uit dit proefschrift samengevat. Dit wordt gevolgd door een discussie die met name gaat over het gebruik van besliskundig modelleren voor het maken van beslissingen op basis van wetenschappelijk bewijs. In deze discussie worden drie aspecten benadrukt. Ten eerste de transparantie bij het besliskundig modelleren. De besliskundige modellen in dit proefschrift zijn ontwikkeld voor en met de artsen aangezien zij uiteindelijk leidend zijn in de beslissingen rondom nieuwe procedures en medische apparatuur. Door de ontwikkelingen in modelstudies worden deze steeds complexer maar daarmee ook minder transparant voor de uiteindelijke beslissers. Om ervoor te zorgen dat besliskundige modellen meer worden gebruikt zal de transparantie ervan moeten worden verbeterd. Ten tweede wordt de evaluatie van chirurgische procedures en medische apparatuur op dit moment pas gedaan wanneer een procedure al vrij uitgebreid wordt gebruikt. Het blijkt dat het de-implementeren van procedures lastig is. Besliskundig modelleren maakt het mogelijk om innovaties al in een vroeg stadium, voordat ze gebruikt worden, te evalueren. Het derde aspect ligt in het individualiseren van zorg. We hebben laten zien dat het toevoegen van predictiemodellen aan besliskundige modellen een waardevolle methode kan zijn om het individualiseren van zorg te evalueren. We moeten echter onderzoeken hoe we deze modellen daadwerkelijk in de praktijk kunnen gebruiken, bijvoorbeeld door ze te verwerken in keuzehulpen.

Met dit voorbeeld over de behandeling van de hals in patiënten met PCC van de mondholte hebben we laten zien dat besliskundig modelleren een waardevolle methode kan zijn om de besluitvorming rondom chirurgische procedures en innovaties te informeren op basis van wetenschappelijk bewijs. Meer en beter gebruik van deze modellen in de toekomst kan zorgen voor betere en betaalbaardere zorg.

Bibliography

1. Djulbegovic B, Kumar A, Glasziou PP, Perera R, Reljic T, Dent L, et al. New treatments compared to established treatments in randomized trials. The Cochrane database of systematic reviews. 2012;10:MR000024.

2. Barbash GI, Glied SA. New technology and health care costs--the case of robot-assisted surgery. The New England journal of medicine. 2010;363(8):701-4.

3. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. Trials. 2009;10:9.

4. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Balliol C, et al. Evaluation and stages of surgical innovations. Lancet. 2009;374(9695):1089-96.

5. Adams CP, Brantner VV. Estimating the cost of new drug development: is it really 802 million dollars? Health affairs. 2006;25(2):420-8.

6. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374(9695):1105-12.

7. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance G. Evidence based medicine: a movement in crisis? Bmj. 2014;348:g3725.

8. Haddad RI, Shin DM. Recent advances in head and neck cancer. The New England journal of medicine. 2008;359(11):1143-54.

9. La Vecchia C, Tavani A, Franceschi S, Levi F, Corrao G, Negri E. Epidemiology and prevention of oral cancer. Oral Oncol. 1997;33(5):302-12.

10. Macfarlane GJ, Zheng T, Marshall JR, Boffetta P, Niu S, Brasure J, et al. Alcohol, tobacco, diet and the risk of oral cancer: a pooled analysis of three case-control studies. European journal of cancer Part B, Oral oncology. 1995;31B(3):181-7.

11. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol. 2009;45(4-5):309-16.

12. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA: a cancer journal for clinicians. 2005;55(2):74-108.

13. Bessell A, Glenny AM, Furness S, Clarkson JE, Oliver R, Conway DI, et al. Interventions for the treatment of oral and oropharyngeal cancers: surgical treatment. Cochrane Database Syst Rev. 2011(9):CD006205.

14. Capote A, Escorial V, Munoz-Guerra MF, Rodriguez-Campo FJ, Gamallo C, Naval L. Elective neck dissection in early-stage oral squamous cell carcinoma--does it influence recurrence and survival? Head Neck. 2007;29(1):3-11.

15. Ferlito A, Rinaldo A, Silver CE, Shah JP, Suarez C, Medina JE, et al. Neck dissection: then and now. Auris, nasus, larynx. 2006;33(4):365-74.

16. Robbins KT, Medina JE, Wolfe GT, Levine PA, Sessions RB, Pruet CW. Standardizing neck dissection terminology. Official report of the Academy's Committee for Head and Neck Surgery and Oncology. Arch Otolaryngol Head Neck Surg. 1991;117(6):601-5.

17. Ferlito A, Robbins KT, Shah JP, Medina JE, Silver CE, Al-Tamimi S, et al. Proposal for a rational classification of neck dissections. Head Neck. 2011;33(3):445-50.

18. Bradley PJ, Ferlito A, Silver CE, Takes RP, Woolgar JA, Strojan P, et al. Neck treatment and shoulder morbidity: still a challenge. Head Neck. 2011;33(7):1060-7.

19. Rogers SN, Ferlito A, Pellitteri PK, Shaha AR, Rinaldo A. Quality of life following neck dissections. Acta Otolaryngol. 2004;124(3):231-6.

20. Kuntz AL, Weymuller EA, Jr. Impact of neck dissection on quality of life. Laryngoscope. 1999;109(8):1334-8.

21. Leipzig B, Suen JY, English JL, Barnes J, Hooper M. Functional evaluation of the spinal accessory nerve after neck dissection. American journal of surgery. 1983;146(4):526-30.

22. Remmler D, Byers R, Scheetz J, Shell B, White G, Zimmerman S, et al. A prospective study of shoulder disability resulting from radical and modified neck dissections. Head & neck surgery. 1986;8(4):280-6.

23. Stuiver MM, van Wilgen CP, de Boer EM, de Goede CJ, Koolstra M, van Opzeeland A, et al. Impact of shoulder complaints after neck dissection on shoulder disability and quality of life. Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery. 2008;139(1):32-9.

24. van Wilgen CP, Dijkstra PU, van der Laan BF, Plukker JT, Roodenburg JL. Shoulder complaints after nerve sparing neck dissections. Int J Oral Maxillofac Surg. 2004;33(3):253-7.

25. Speksnijder CM, van der Bilt A, Slappendel M, de Wijer A, Merkx MA, Koole R. Neck and shoulder function in patients treated for oral malignancies: A 1-year prospective cohort study. Head Neck. 2013;35(9):1303-13.

26. Coskun HH, Medina JE, Robbins KT, Silver CE, Strojan P, Teymoortash A, et al. Current philosophy in the surgical management of neck metastases for head and neck squamous cell carcinoma. Head Neck. 2014.

27. Robbins KT, Ferlito A, Shah JP, Hamoir M, Takes RP, Strojan P, et al. The evolving role of selective neck dissection for head and neck squamous cell carcinoma. Eur Arch Otorhinolaryngol. 2013;270(4):1195-202.

28. Fasunla AJ, Greene BH, Timmesfeld N, Wiegand S, Werner JA, Sesterhenn AM. A meta-analysis of the randomized controlled trials on elective neck dissection versus therapeutic neck dissection in oral cavity cancers with clinically node-negative neck. Oral Oncol. 2011;47(5):320-4.

29. Yuen AP, Ho CM, Chow TL, Tang LC, Cheung WY, Ng RW, et al. Prospective randomized study of selective neck dissection versus observation for N0 neck of early tongue carcinoma. Head Neck. 2009;31(6):765-72.

30. Borgemeester MC, van den Brekel MW, van Tinteren H, Smeele LE, Pameijer FA, van Velthuysen ML, et al. Ultrasound-guided aspiration cytology for the assessment of the clinically N0 neck: factors influencing its accuracy. Head Neck. 2008;30(11):1505-13.

31. Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. Jama. 2012;307(17):1801-2.

32. National Institute for Health and Care Excellence: do not do recommendations 2014. Available from: http://www.nice.org.uk/usingguidance/donotdorecommendations/search.jsp.

33. ZonMw. Verstandige zorg, kiezen voor kwaliteit en kostenbesparing 2014. Available from: http://mediator.zonmw.nl/mediator-special-april-2014/.

34. Elshaug AG, Watt AM, Mundy L, Willis CD. Over 150 potentially low-value health care practices: an Australian study. The Medical journal of Australia. 2012;197(10):556-60.

35. Dutch Hospital Data: National Registration Hospital Care. 2012.

36. Health & Social Care Information Centre: Hospital Episode Statistics, Admitted Patient Care, Engeland [Internet]. 2012-13.

37. Statistics Netherlands (CBS) Statline: Consumer price index 2014. Available from:://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71311ned&D1=0&D2=0&D3=12,25,38,51,64,77,90,103,116,129,142,155,168,181,194,219,232&HDR=T&ST-B=G2,G1&VW=T.

38. Achmea. Health Database. Available from: http://www.achmeazorg.nl/overige-zorg/overige-zorg/initiatievenindezorg/Achmea-Health-Database/Pages/default.aspx.

39. van den Brekel MW, Castelijns JA. What the clinician wants to know: surgical perspective and ultrasound for lymph node imaging of the neck. Cancer Imaging. 2005;5 Spec No A:S41-9.

40. Office for National Statistics: Population Estimates for UK, England and Wales, Scotland and Northern Ireland 2012. Available from: http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk--england-and-wales--scotland-and-northern-ireland/mid-2011-and-mid-2012/index.html.

41. Fergusson RJ, Lethaby A, Shepperd S, Farquhar C. Endometrial resection and ablation versus hysterectomy for heavy menstrual bleeding. Cochrane Database Syst Rev. 2013;11:CD000329.

42. Hagemann AR, Zighelboim I, Odibo AO, Rader JS, Mutch DG, Powell MA. Cost-benefit of laparoscopic versus medical ovarian suppression in premenopausal breast cancer. The breast journal. 2011;17(1):103-5.

43. Lubowitz JH, Appleby D. Cost-effectiveness analysis of the most common orthopaedic surgery procedures: knee arthroscopy and knee anterior cruciate ligament reconstruction. Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association. 2011;27(10):1317-22.

44. Verry H, Lord SJ, Martin A, Gill G, Lee CK, Howard K, et al. Effectiveness and cost-effectiveness of sentinel lymph node biopsy compared with axillary node dissection in patients with early-stage breast cancer: a decision model analysis. Br J Cancer. 2012;106(6):1045-52.

45. Stovsky MD, Griffiths RI, Duff SB. A clinical outcomes and cost analysis comparing photoselective vaporization of the prostate to alternative minimally invasive therapies and transurethral prostate resection for the treatment of benign prostatic hyperplasia. The Journal of urology. 2006;176(4 Pt 1):1500-6.

46. Faria R, Bojke L, Epstein D, Corbacho B, Sculpher M, group Rt. Cost-effectiveness of laparoscopic fundoplication versus continued medical management for the treatment of gastro-oesophageal reflux disease based on long-term follow-up of the REFLUX trial. The British journal of surgery. 2013;100(9):1205-13.

47. Ogan K, Lotan Y, Koeneman K, Pearle MS, Cadeddu JA, Rassweiler J. Laparoscopic versus open retroperitoneal lymph node dissection: a cost analysis. The Journal of urology. 2002;168(5):1945-9; discussion 9.

48. Simpson EL, Kearns B, Stevenson MD, Cantrell AJ, Littlewood C, Michaels JA. Enhancements to angioplasty for peripheral arterial occlusive disease: systematic review, cost-effectiveness assessment and expected value of information analysis. Health technology assessment. 2014;18(10):1-252.

49. Lee R, Li PS, Goldstein M, Schattman G, Schlegel PN. A decision analysis of treatments for nonobstructive azoospermia associated with varicocele. Fertility and sterility. 2009;92(1):188-96.

50. Verhoeve HR, Moolenaar LM, Hompes P, van der Veen F, Mol BW. Cost-effectiveness of tubal patency tests. BJOG : an international journal of obstetrics and gynaecology. 2013;120(5):583-93.

51. Isla-Tejera B, Ruano J, Alvarez MA, Brieva T, Cardenas M, Baamonde C, et al. Economic evaluation of botulinum toxin versus thoracic sympathectomy for palmar hyperhidrosis: data from a real-world scenario. Dermatology and therapy. 2013;3(1):63-72.

52. Statistics Netherlands (CBS): Operaties in het ziekenhuis 2010. Available from: http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=80386NED&LA=NL.

53. Royal College of Surgeons of England: Surgery and the NHS in numbers. Available from: https://www.rcseng.ac.uk/media/media-background-briefings-and-statistics/surgery-and-the-nhs-in-numbers.

54. Berwick DM, Hackbarth AD. Eliminating waste in US health care. Jama. 2012;307(14):1513-6.

55. Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring Low-Value Care in Medicare. JAMA internal medicine. 2014.

56. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. Bmj. 2010;340:c221.

57. Grutters JP, Sculpher M, Briggs AH, Severens JL, Candel MJ, Stahl JE, et al. Acknowledging patient heterogeneity in economic evaluation : a systematic literature review. PharmacoEconomics. 2013;31(2):111-23.

58. Morden NE, Colla CH, Sequist TD, Rosenthal MB. Choosing wisely--the politics and economics of labeling low-value services. The New England journal of medicine. 2014;370(7):589-92.

59. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. Bmj. 2002;324(7351):1448-51.

60. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, et al. Challenges in evaluating surgical innovation. Lancet. 2009;374(9695):1097-104.

61. Lambert R, Sauvaget C, de Camargo Cancela M, Sankaranarayanan R. Epidemiology of cancer from the oral cavity and oropharynx. Eur J Gastroenterol Hepatol. 2011;23(8):633-41.

62. Rodrigo JP, Shah JP, Silver CE, Medina JE, Takes RP, Robbins KT, et al. Management of the clinically negative neck in early-stage head and neck cancers after transoral resection. Head & neck. 2011;33(8):1210-9.

63. Schiefke F, Akdemir M, Weber A, Akdemir D, Singer S, Frerich B. Function, postoperative morbidity, and quality of life after cervical sentinel node biopsy and after selective neck dissection. Head & neck. 2009;31(4):503-12.

64. Murer K, Huber GF, Haile SR, Stoeckli SJ. Comparison of morbidity between sentinel node biopsy and elective neck dissection for treatment of the n0 neck in patients with oral squamous cell carcinoma. Head Neck. 2011;33(9):1260-4.

65. Govers TM, Hannink G, Merkx MA, Takes RP, Rovers MM. Sentinel node biopsy for squamous cell carcinoma of the oral cavity and oropharynx: A diagnostic meta-analysis. Oral Oncol. 2013;49(8):726-32.

66. Wensing BM, Merkx MA, Krabbe PF, Marres HA, Van den Hoogen FJ. Oral squamous cell carcinoma and a clinically negative neck: the value of follow-up. Head Neck. 2011;33(10):1400-5.

67. SurveyMonkey, https://nl.surveymonkey.com/mp/aboutus/ 2014.

68. NWHHT (Nederlandse Werkgroep Hoofd Hals Tumoren): Landelijke richtlijnen Mondholte- en orofarynxcarcinoom 2004. Available from: http://www.oncoline.nl/index.php?pagina=/richtlijn/item/pagina.php&id=21230&richtlijn_id=383.

69. ENT UK: Head and Neck Caner: Multidisciplinary Management Guidelines 2011. Available from: http://bahno.org.uk/docs/BAHNO%20STANDARDS%20DOC09.pdf.

70. NCCN (National Comprehensive Cancer Network): Head and Neck cancers: NCCN CLinical Practice Guidelines in Oncology 2014. Available from: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site.

71. Werning JW, Heard D, Pagano C, Khuder S. Elective management of the clinically negative neck by otolaryngologists in patients with oral tongue cancer. Arch Otolaryngol Head Neck Surg. 2003;129(1):83-8.

72. de Bree R, van der Waal I, Doornaert P, Werner JA, Castelijns JA, Leemans CR. Indications and extent of elective neck dissection in patients with early stage oral and oropharyngeal carcinoma: nationwide survey in The Netherlands. The Journal of laryngology and otology. 2009;123(8):889-98.

73. Galbiatti AL, Padovani-Junior JA, Maniglia JV, Rodrigues CD, Pavarino EC, Goloni-Bertollo EM. Head and neck cancer: causes, prevention and treatment. Brazilian journal of otorhinolaryngology. 2013;79(2):239-47.

74. Gibson MK, Forastiere AA. Multidisciplinary approaches in the management of advanced head and neck tumors: state of the art. Current opinion in oncology. 2004;16(3):220-4.

75. Ferlito A, Rinaldo A, Robbins KT, Leemans CR, Shah JP, Shaha AR, et al. Changing concepts in the surgical management of the cervical node metastasis. Oral Oncol. 2003;39(5):429-35.

76. SamenOK. Available from: www.samenok.nl.

77. Barello S, Graffigna G, Vegni E. Patient engagement as an emerging challenge for healthcare services: mapping the literature. Nursing research and practice. 2012;2012:905934.

78. Chen SC, Lai YH, Liao CT, Chang JT, Lin CC. Unmet information needs and preferences in newly diagnosed and surgically treated oral cavity cancer patients. Oral Oncol. 2009;45(11):946-52.

79. Noy C. Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research. International Journal of Social Research Methodology. 2008;11(4):327-44.

80. Atlas.ti. Atlas.ti qualitative data analysis. 7.0 ed.

81. Burnard P. A method of analysing interview transcripts in qualitative research. 1991;11(6):461-6.

82. Moret-Hartman M, Reuzel R, Grin J, van der Wilt GJ. Participatory workshops are not enough to prevent policy implementation failures: an example of a policy development process concerning the drug interferon-beta for multiple sclerosis. Health care analysis : HCA : journal of health philosophy and policy. 2008;16(2):161-75.

83. SONCOS. Stichting Oncologische Samenwerking. Available from: www.soncos.nl.

84. Reuzel R, Grin J, Akkerman T. The role of the evaluator in an interactive evaluation of cochlear implantation: shaping power, trust and deliberation. International Journal of Foresight and Innovation Policy. 2007;3(1):76-94.

85. Hornstra MT, Alkureishi LW, Ross GL, Shoaib T, Soutar DS. Predictive factors for failure to identify sentinel nodes in head and neck squamous cell carcinoma. Head Neck. 2008;30(7):858-62.

86. Ebrahimi A, Ashford BG, Clark JR. Improved survival with elective neck dissection in thick early-stage oral squamous cell carcinoma. Head Neck. 2012;34(5):709-16.

87. Vermeeren L, Valdes Olmos RA, Klop WM, Balm AJ, van den Brekel MW. A portable gamma-camera for intraoperative detection of sentinel nodes in the head and neck region. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2010;51(5):700-3.

88. Civantos FJ, Zitsch RP, Schuller DE, Agrawal A, Smith RB, Nason R, et al. Sentinel lymph node biopsy accurately stages the regional lymph nodes for T1-T2 oral squamous cell carcinomas: results of a prospective multi-institutional trial. J Clin Oncol. 2010;28(8):1395-400.

89. Reuters T. Endnote v5, Inc., Philadelphia, PA.

90. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.

91. Review Manager (RevMan). Version 5.1. Copenhagen: The Nordic Cochrane Centre, The Cochrane collaboration, 2011.

92. Takwoingi Y, Deeks J. MetaDAS: A SAS macro for meta- analysis of diagnostic accuracy studies. User Guide Version 1.3. 2010 July. Available from: http://srdta.cochrane.org/.

93. The SAS system, 8.2. SAS Institute Inc, Cary, NC, USA.

94. Alkureishi LW, Ross GL, Shoaib T, Soutar DS, Robertson AG, Thompson R, et al. Sentinel node biopsy in head and neck squamous cell cancer: 5-year follow-up of a European multicenter trial. Ann Surg Oncol. 2010;17(9):2459-64.

95. Alvarez Amezaga J, Barbier Herrero L, Pijoan del Barrio JI, Martin Rodriguez JC, Romo Simon L, Genolla Subirats J, et al. Diagnostic efficacy of sentinel node biopsy in oral squamous cell carcinoma. Cohort study and meta-analysis. Med Oral Patol Oral Cir Bucal. 2007;12(3):E235-43.

96. Barzan L, Sulfaro S, Alberti F, Politi D, Marus W, Pin M, et al. Gamma probe accuracy in detecting the sentinel lymph node in clinically N0 squamous cell carcinoma of the head and neck. The Annals of otology, rhinology, and laryngology. 2002;111(9):794-8.

97. Burcia V, Costes V, Faillie JL, Gardiner Q, de Verbizier D, Cartier C, et al. Neck restaging with sentinel node biopsy in T1-T2N0 oral and oropharyngeal cancer: Why and how? Otolaryngol Head Neck Surg. 2010;142(4):592-7 e1.

98. Civantos FJ, Moffat FL, Goodwin WJ. Lymphatic mapping and sentinel lymphadenectomy for 106 head and neck lesions: contrasts between oral cavity and cutaneous malignancy. Laryngoscope. 2006;112(3 Pt 2 Suppl 109):1-15.

99. Hart RD, Nasser JG, Trites JR, Taylor SM, Bullock M, Barnes D. Sentinel lymph node biopsy in N0 squamous cell carcinoma of the oral cavity and oropharynx. Arch Otolaryngol Head Neck Surg. 2005;131(1):34-8.

100. Hoft S, Maune S, Muhle C, Brenner W, Czech N, Kampen WU, et al. Sentinel lymph-node biopsy in head and neck cancer. Br J Cancer. 2004;91(1):124-8.

101. Jeong HS, Baek CH, Son YI, Cho DY, Chung MK, Min JY, et al. Sentinel lymph node radiolocalization with 99mTc filtered tin colloid in clinically node-negative squamous cell carcinomas of the oral cavity. J Korean Med Sci. 2006;21(5):865-70.

102. Kontio R, Leivo I, Leppanen E, Atula T. Sentinel lymph node biopsy in oral cavity squamous cell carcinoma without clinically evident metastasis. Head Neck. 2004;26(1):16-21.

103. Minamikawa T, Umeda M, Komori T. Reliability of sentinel lymph node biopsy with squamous cell carcinoma of the oral cavity. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2005;99(5):532-8; discussion 9-41.

104. Pitman KT, Johnson JT, Brown ML, Myers EN. Sentinel lymph node biopsy in head and neck squamous cell carcinoma. Laryngoscope. 2002;112(12):2101-13.

105. Rigual N, Douglas W, Lamonica D, Wiseman S, Cheney R, Hicks W, Jr., et al. Sentinel lymph node biopsy: a rational approach for staging T2N0 oral cancer. Laryngoscope. 2005;115(12):2217-20.

106. Taylor RJ, Wahl RL, Sharma PK, Bradford CR, Terrell JE, Teknos TN, et al. Sentinel node localization in oral cavity and oropharynx squamous cell cancer. Arch Otolaryngol Head Neck Surg. 2001;127(8):970-4.

107. Stoeckli SJ. Sentinel node biopsy for oral and oropharyngeal squamous cell carcinoma of the head and neck. Laryngoscope. 2007;117(9):1539-51.

108. Melkane AE, Mamelle G, Wycisk G, Temam S, Janot F, Casiraghi O, et al. Sentinel node biopsy in early oral squamous cell carcinomas: a 10-year experience. Laryngoscope. 2012;122(8):1782-8.

109. Vorburger MS, Broglie MA, Soltermann A, Haerle SK, Haile SR, Huber GF, et al. Validity of frozen section in sentinel lymph node biopsy for the staging in oral and oropharyngeal squamous cell carcinoma. J Surg Oncol. 2012;106(7):816-9.

110. Terada A, Hasegawa Y, Yatabe Y, Hanai N, Ozawa T, Hirakawa H, et al. Follow-up after intraoperative sentinel node biopsy of N0 neck oral cancer patients. European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies. 2011;268(3):429-35.

111. Pezier T, Nixon IJ, Gurney B, Schilling C, Hussain K, Lyons AJ, et al. Sentinel lymph node biopsy for T1/T2 oral cavity squamous cell carcinoma--a prospective case series. Ann Surg Oncol. 2012;19(11):3528-33.

112. Burns P, Foster A, Walshe P, O'Dwyer T. Sentinel lymph node biopsy in node-negative squamous cell carcinoma of the oral cavity and oropharynx. The Journal of laryngology and otology. 2009;123(4):439-43.

113. Broglie MA, Haerle SK, Huber GF, Haile SR, Stoeckli SJ. Occult metastases detected by sentinel node biopsy in patients with early oral and oropharyngeal squamous cell carcinomas: Impact on survival. Head Neck. 2012.

114. Weiss MH, Harrison LB, Isaacs RS. Use of decision analysis in planning a management strategy for the stage N0 neck. Arch otolaryngol Head neck surg. 1994;120:699-702.

115. Okura M, Aikawa T, Sawai NY, Iida S, Kogo M. Decision analysis and treatment threshold in a management for the N0 neck of the oral cavity carcinoma. Oral Oncol. 2009;45(10):908-11.

116. van Hooff SR, Leusink FK, Roepman P, Baatenburg de Jong RJ, Speel EJ, van den Brekel MW, et al. Validation of a Gene Expression Signature for Assessment of Lymph Node Metastasis in Oral Squamous Cell Carcinoma. J Clin Oncol. 2012.

117. Kosuda S, Kusano S, Kohno N, Ohno Y, Tanabe T, Kitahara S, et al. Feasibility and cost-effectiveness of sentinel lymph node radiolocalization in stage N0 head and neck cancer. Arch Otolaryngol Head Neck Surg. 2003;129(10):1105-9.

118. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, et al. Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):804-11.

119. Ferlito A, Silver CE, Rinaldo A. Elective management of the neck in oral cavity squamous carcinoma: current concepts supported by prospective studies. Brit J Oral Max Surg. 2009;47(1):5-9.

120. Hakkaart-van Roijen L TS, Bouwmans CAM. Handleiding voor kostenonderzoek, methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. College voor zorgverzekeringen. Geactualiseerde versie 2010.

121. Richtlijnen voor farmaco-economisch onderzoek. College voor zorgverzekeringen. Geactualiseerde versie 2006.

122. Drummond MF, Sculpher MJ, Torrance GW, J. OBB, L. SG. Methods for the economic evaluation of health care programmes. 3rd ed Oxford: Oxford University Press. 2005.

123. Dutch council for Public Health and Health Care. Sensible and sustainable care. http://wwwrvznet. 2009 (15 January 2009).

124. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Econ. 1996;5(6):513-24.

125. Grutters JP, Joore MA, Wiegman EM, Langendijk JA, de Ruysscher D, Hochstenbag M, et al. Health-related quality of life in patients surviving non-small cell lung cancer. Thorax. 2010;65(10):903-7.

126. van Wilgen CP, Dijkstra PU, Nauta JM, Vermey A, Roodenburg JL. Shoulder pain and disability in daily life, following supraomohyoid neck dissection: a pilot study. Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery. 2003;31(3):183-6.

127. Govers TM, Takes RP, Baris Karakullukcu M, Hannink G, Merkx MA, Grutters JP, et al. Management of the N0 neck in early stage oral squamous cell cancer: a modeling study of the cost-effectiveness. Oral Oncol. 2013;49(8):771-7.

128. Sculpher M. NICE's 2008 Methods Guide: sensible consolidation or opportunities missed? PharmacoEconomics. 2008;26(9):721-4.

129. Group TE. EuroQol--a new facility for the measurement of health-related quality of life. . Health policy. 1990;16(3):199-208.

130. van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. Journal of clinical epidemiology. 2000;53(1):29-38.

131. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ. 2006;15(10):1121-32.

132. Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. Medical care. 2007;45(12):1162-70.

133. Hanmer J, Hays RD, Fryback DG. Mode of administration is important in US national estimates of health-related quality of life. Medical care. 2007;45(12):1171-9.

134. Kind PHGMS. UK population norms for EQ-5D. http://ideasrepecorg/p/chy/respap/172chedphtml. 1999.

135. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and quality of life outcomes. 2007;5:70.

136. Glenny AM, Furness S, Worthington HV, Conway DI, Oliver R, Clarkson JE, et al. Interventions for the treatment of oral cavity and oropharyngeal cancer: radiotherapy. Cochrane Database Syst Rev. 2010(12):CD006387.

137. Ferlito A, Silver CE, Rinaldo A. Elective management of the neck in oral cavity squamous carcinoma: current concepts supported by prospective studies. Br J Oral Maxillofac Surg. 2009;47(1):5-9.

138. Schiff BA, Roberts DB, El-Naggar A, Garden AS, Myers JN. Selective vs modified radical neck dissection and postoperative radiotherapy vs observation in the treatment of squamous cell carcinoma of the oral tongue. Arch Otolaryngol Head Neck Surg. 2005;131(10):874-8.

139. Statistics Netherlands, Statline: mortality; gender, age (on December 31) and marital status [Internet]. 2012.

140. Healthcare insurance board: Guidelines for pharmacoeconomic research. Diemen: updated version 2006.

141. Dutch Head and Neck Society: Guideline oral cavity and oropharyngeal cancer. Alphen aan den Rijn2004.

142. 'Stichting Klankbord' (Dutch patient association): For patients with head and neck cancer 2013. Available from: http://www.stichtingklankbord.nl/klankbord/.

143. Hakkaart-van Roijen L TS BC. Instructions for cost studies; methods and standard cost for economic evaluations in health care Institute for Medical Technology Assessment. Erasmus University Rotterdam: Updated version 2010.

144. Ramaekers BL, Grutters JP, Pijls-Johannesma M, Lambin P, Joore MA, Langendijk JA. Protons in head-and-neck cancer: bridging the gap of evidence. International journal of radiation oncology, biology, physics. 2013;85(5):1282-8.

145. Statistics Netherlands, Statline, table: Cao wages, contractual wage costs and working hours; indexfig. (2000=100) [Internet]. 2013.

146. Statistics Netherlands, Statline, table: Consumer prices; price index 2006 = 100 [Internet]. 2013.

147. McNeely ML, Parliament M, Courneya KS, Seikaly H, Jha N, Scrimger R, et al. A pilot study of a randomized controlled trial to evaluate the effects of progressive resistance exercise training on shoulder dysfunction caused by spinal accessory neurapraxia/neurectomy in head and neck cancer survivors. Head Neck. 2004;26(6):518-30.

148. Dutch Council for Public Health and Health Care: Sensible and sustainable care. Zoetermeer, the Nederlands: 2006.

149. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):835-42.

150. O'Brien BJ, Gertsen K, Willan AR, Faulkner LA. Is there a kink in consumers' threshold value for cost-effectiveness in health care? Health Econ. 2002;11(2):175-80.

151. Severens JL, Brunenberg DE, Fenwick EA, O'Brien B, Joore MA. Cost-effectiveness acceptability curves and a reluctance to lose. PharmacoEconomics. 2005;23(12):1207-14.

152. U.S. Food and Drug Administration: Personalized Medicine. Available from: http://www.fda.gov/scienceresearch/specialtopics/personalizedmedicine/default.htm.

153. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. PharmacoEconomics. 2008;26(9):799-806.

154. McNutt R, Hadler NM. Scientific American: How clinical guidelines can fail both doctors and patients 2013. Available from: http://blogs.scientificamerican.com/guest-blog/2013/11/22/how-clinical-guidelines-can-fail-both-doctors-and-patients/.

155. van der Weijden T, Boivin A, Burgers J, Schunemann HJ, Elwyn G. Clinical practice guidelines and patient decision aids. An inevitable relationship. Journal of clinical epidemiology. 2012;65(6):584-9.

156. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. Bmj. 2013;346:e5793.

157. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj. 2015;350:g7594.

158. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS medicine. 2013;10(2):e1001381.

159. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. Journal of thrombosis and haemostasis : JTH. 2013;11 Suppl 1:129-41.

160. Goerkem M, Braun J, Stoeckli SJ. Evaluation of clinical and histomorphological parameters as potential predictors of occult metastases in sentinel lymph nodes of early squamous cell carcinoma of the oral cavity. Annals of surgical oncology. 2010;17(2):527-35.

161. Osaki T, Kimura T, Tatemoto Y, Yamamoto T, Yoneda K. Risk factors of metastasis in oral squamous cell carcinomas. Oncology. 2000;58(2):137-43.

162. Clark JR, Naranjo N, Franklin JH, de Almeida J, Gullane PJ. Established prognostic variables in N0 oral carcinoma. Otolaryngol Head Neck Surg. 2006;135(5):748-53.

163. Hiratsuka H, Miyakawa A, Nakamori K, Kido Y, Sunakawa H, Kohama G. Multivariate analysis of occult lymph node metastasis as a prognostic indicator for patients with squamous cell carcinoma of the oral cavity. Cancer. 1997;80(3):351-6.

164. Woolgar JA. Micrometastasis in oral/oropharyngeal squamous cell carcinoma: incidence, histopathological features and clinical implications. The British journal of oral & maxillofacial surgery. 1999;37(3):181-6.

165. Larsen SR, Johansen J, Sorensen JA, Krogdahl A. The prognostic significance of histological features in oral squamous cell carcinoma. Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology. 2009;38(8):657-62.

166. Ross GL, Soutar DS, MacDonald DG, Shoaib T, Camilleri IG, Robertson AG. Improved staging of cervical metastases in clinically node-negative patients with head and neck squamous cell carcinoma. Annals of surgical oncology. 2004;11(2):213-8.

167. Govers TM, Schreuder WH, Klop WM, Grutters JP, Rovers MM, Merkx MA, et al. Quality of life after different procedures for regional control in oral cancer patients: Cross-sectional survey. Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery. 2015.

168. Baatenburg de Jong RJ, Hermans J, Molenaar J, Briaire JJ, le Cessie S. Prediction of survival in patients with head and neck cancer. Head Neck. 2001;23(9):718-24.

169. van der Schroeff MP, Steyerberg EW, Wieringa MH, Langeveld TP, Molenaar J, Baatenburg de Jong RJ. Prognosis: a variable parameter: dynamic prognostic modeling in head and neck squamous cell carcinoma. Head Neck. 2012;34(1):34-41.

170. Datema FR, Ferrier MB, Vergouwe Y, Moya A, Molenaar J, Piccirillo JF, et al. Update and external validation of a head and neck cancer prognostic model. Head Neck. 2013;35(9):1232-7.

171. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. Medical decision making : an international journal of the Society for Medical Decision Making. 2007;27(2):112-27.

172. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. Health Econ. 2003;12(5):421-7.

173. Basu A, Carlson JJ, Veenstra DL. A Framework for Prioritizing Research Investments in Precision Medicine. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2015;18(7):A729.

174 Halpern EF, Weinstein MC, Hunink MG, Gazelle GS. Representing both first- and second-order uncertainties by Monte Carlo simulation for groups of patients. Medical decision making : an international journal of the Society for Medical Decision Making. 2000;20(3):314-22.

175. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Bmj. 2005;330(7494):765.

176. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. Bmj. 2009;338:b606.

177. Govers TM, Schreuder W, Klop W, Grutters JP, Rovers MM, Merkx MA, et al. Health-related quality of life after different procedures for regional control in early-stage oral cavity squamous cell cancer patients. Recently submitted. 2014.

178. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):843-50.

179. Brousselle A, Lessard C. Economic evaluation to inform health care decision-making: promise, pitfalls and a proposal for an alternative path. Social science & medicine. 2011;72(6):832-9.

180. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. PharmacoEconomics. 2006;24(11):1043-53.

181. Zorginstituut Nederland: Kosteneffectiviteit in de praktijk. www.zorginstituutnederland.nl2015.

182. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. PharmacoEconomics. 2013;31(5):361-7.

183. Caro JJ, Briggs AH, Siebert U, Kuntz KM, Force I-SMGRPT. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):796-803.

184. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. Bmj. 1996;313(7052):275-83.

185. Gezondheidsraad: De kunst van het nee zeggen. Factoren die bijdragen aan onnodige zorg. Den Haag: 2014.

186. Abrishami P, Boer A, Horstman K. Understanding the adoption dynamics of medical innovations: affordances of the da Vinci robot in the Netherlands. Social science & medicine. 2014;117:125-33.

187. College voor Zorgverzekeringen: Kosteneffectiviteit in de zorg; Op weg naar een genuanceerd en geaccepteerd gebruik van kosteneffectiviteitsgegevens in de zorg. 2013.

188. Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring low-value care in Medicare. JAMA internal medicine. 2014;174(7):1067-76.

189 Elshaug AG, Hiller JE, Tunis SR, Moss JR. Challenges in Australian policy processes for disinvestment from existing, ineffective health care practices. Australia and New Zealand health policy. 2007;4:23.

190. Haas M, Hall J, Viney R, Gallego G. Breaking up is hard to do: why disinvestment in medical technology is harder than investment. Australian health review : a publication of the Australian Hospital Association. 2012;36(2):148-52.

191. Hughes DA, Wood EM, Tuersley L. NICE recommendations: why no disinvestment recommendations to offset investment decisions? Bmj. 2015;350:h2656.

192. Pietzsch JB, Pate-Cornell ME. Early technology assessment of new medical devices. International journal of technology assessment in health care. 2008;24(1):36-44.

193. Wild C, Langer T. Emerging health technologies: informing and supporting health policy early. Health policy. 2008;87(2):160-71.

194. Ijzerman MJ, Steuten LM. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. Applied health economics and health policy. 2011;9(5):331-47.

195. Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: A decision framework. Journal of commercial Biotechnology. 2007;13:263-71.

196. Retel VP, Grutters JP, van Harten WH, Joore MA. Value of research and value of development in early assessments of new medical technologies. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(5):720-8.

197. McAteer H, Cosh E, Freeman G, Pandit A, Wood P, Lilford R. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. Journal of tissue engineering and regenerative medicine. 2007;1(5):343-9.

198. Chapman AM, Taylor CA, Girling AJ. Early HTA to inform Medical Device Development Decisions - The Headroom Method. Mediterranean Conference on Medical and Biological Engineering and Computing 2013 IFMBE Proceedings; Seville, Spain: Springer; 2014. p. 1151-4.

199. Eddy DM. Clinical decision making: from theory to practice. Designing a practice policy. Standards, guidelines, and options. Jama. 1990;263(22):3077, 81, 84.

200. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. The New England journal of medicine. 2012;366(9):780-1.

201. Stiggelbout AM, Pieterse AH, De Haes JC. Shared decision making: Concepts, evidence, and practice. Patient education and counseling. 2015.

202. Trenaman L, Bryan S, Bansback N. The cost-effectiveness of patient decision aids: A systematic review. Healthcare. 2014;2(4):251-7.

203. Agoritsas T, Heen AF, Brandt L, Alonso-Coello P, Kristiansen A, Akl EA, et al. Decision aids that really promote shared decision making: the pace quickens. Bmj. 2015;350:g7624.

204. Hubben GA, Bos JM, Glynn DM, van der Ende A, van Alphen L, Postma MJ. Enhanced decision support for policy makers using a web interface to health-economic models--illustrated with a cost-effectiveness analysis of nation-wide infant vaccination with the 7-valent pneumococcal conjugate vaccine in the Netherlands. Vaccine. 2007;25(18):3669-78.

205. van Hulst M, Hubben GA, Sagoe KW, Promwong C, Permpikul P, Fongsatitkul L, et al. Web interface-supported transmission risk assessment and cost-effectiveness analysis of postdonation screening: a global model applied to Ghana, Thailand, and the Netherlands. Transfusion. 2009;49(12):2729-42.

206. European Commission: Safer, more effective and innovative medical devices [press release]. Brussels2012.

207. Hinni ML, Ferlito A, Brandwein-Gensler MS, Takes RP, Silver CE, Westra WH, et al. Surgical margins in head and neck cancer: a contemporary review. Head Neck. 2013;35(9):1362-70.

208. Feichtinger M, Pau M, Zemann W, Aigner RM, Karcher H. Intraoperative control of resection margins in advanced head and neck cancer using a 3D-navigation system based on PET/CT image fusion. Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery. 2010;38(8):589-94.

209. Kohn MS, Sun J, Knoop S, Shabo A, Carmeli B, Sow D, et al. IBM's Health Analytics and Clinical Decision Support. Yearbook of medical informatics. 2014;9:154-62.

Publicaties:

Govers TM, de Kort TBH, Merkx MAW, Steens S, Rovers MM, de Bree R, Takes RP. An international comparison of the management of the neck in early squamous cell carcinoma in the Netherlands, UK, and USA. Journal of Cranio-Maxillofacial Surgery 2016, Jan; 44 (1): 62-9

Govers TM, Hannink G, Merkx MAW, Takes RP, Rovers MM. Sentinel node biopsy for squamous cell carcinoma of the oral cavity and oropharynx: A diagnostic meta-analysis. Oral Oncology 2013, Aug; 49 (8): 726-32

Govers TM, Takes RP, Karakullukcu MB, Hannink G, Merkx MAW, Grutters JPC, Rovers MM. Management of the cN0 neck in early stage oral squamous cell cancer: A modeling study of the cost-effectiveness. Oral Oncology 2013, Aug; 49 (8): 771-7

Govers TM, Schreuder WH, Klop WM, Grutters JPC, Rovers MM, Merkx MAW, Takes RP. Quality of life after different procedures for regional control in oral cancer patients: Cross-sectional survey. Clinical Otolaryngology 2015, July

Govers TM, Patel S, Takes RP, Merkx MAW, Rovers MM, Grutters JPC. Cost-effectiveness of selective neck dissection versus modified radical neck dissection for treating metastases in oral cavity cancer patients; a modeling study. Head & Neck 2014, July

Rongen JJ, Govers TM, Buma P, Grutters JPC, Hannink G. Societal and economic impact of meniscus scaffold procedures for irreparable meniscus injuries. The American Journal of Sports Medicine, accepted for publication.

Curriculum Vitae

Tim Govers werd geboren op 25 juli 1989 in Heesch. Hij ging naar basisschool de Toermalijn in Heesch en later volgde hij VWO(+) op het Mondriaan college in Oss. Na het behalen van zijn diploma studeerde hij Biomedische Wetenschappen, met als specialisatie Health Technology Assessment, aan de Radboud Universiteit in Nijmegen. In 2013 ronde hij de studie Biomedische Wetenschappen af met een afsluitende masterstage bij de afdeling operatiekamers. Dit onderzoek zette hij voort als promovendus bij de afdeling Operatiekamers, de afdeling Keel-Neus-Oorheelkunde en de afdeling Mond-Kaak en Aangezichtschirurgie. Dit heeft geresulteerd in de onderzoeken beschreven in dit boekje. Tijdens zijn promotie was hij betrokken bij de oprichting van MedValue. Hier voerde hij verschillende onderzoeken uit naar de (potentiële) toegevoegde waarde van medische innovaties, met name in de vroege fases van ontwikkelling. Vanaf 2015 werkt hij als wetenschappelijk onderzoeker bij de afdeling operatiekamers waar vanuit hij onderzoeken uitvoert voor MedValue en daarnaast ook betrokken is bij de verdere ontwikkeling van MedValue. Daarnaast is hij betrokken bij wetenschappelijk onderzoek, met name op gebied van besliskundig modelleren.

Dankwoord

Graag wil ik iedereen bedanken die heeft bijgedragen aan dit proefschrift. Ik wil een aantal mensen in het bijzonder bedanken die ervoor hebben gezorgd dat ik tijdens mijn promotie veel heb geleerd en dit bovenal met veel plezier heb kunnen doen.

Beste Maroeska, als ik aan iemand veel dank ben verschuldigd, dan is het aan jou. Je gaf mij de mogelijkheid om met dit onderzoek te starten tijdens mijn masterstage en om dit een vervolg te geven in een promotietraject. Wat ik vooral bijzonder vind, is dat je me daarbij veel mogelijkheden hebt gegeven om mezelf te ontwikkelen, zowel binnen als naast mijn promotie. Soms brengt dat ons in lastige situaties, maar uiteindelijk zorgt dat er juist voor dat ik als onderzoeker maar zeker ook als persoon kan groeien. Ik wil je ook bedanken voor de leuke sfeer die jij mogelijk maakt in de groep.

Beste Thijs, bedankt voor je frisse ideeën bij nieuwe onderzoeken en bij het schrijven van de artikelen. Je was altijd bereid om mee te zoeken naar de benodigde data of om de manuscripten door te nemen. Soms vroeg ik jou om meerdere manuscripten van mij tegelijk te beoordelen (met een beetje haast…), waarbij je dan zelfs in je vakantie daar tijd voor maakte. Dat heeft me een aantal keer wat stress bespaard! Bedankt ook voor de prettige sfeer waarin ik onder jou mocht promoveren.

Beste Janneke, ik vind het super om met jou te kunnen sparren over modellen. Ik bewonder je scherpe blik en de manier waarop jij ingewikkelde dingen simpel kunt verwoorden. Daarnaast sta jij altijd voor mij klaar om ergens over mee te denken of om mij advies te geven. Bedankt voor de dingen die je mij hebt geleerd op het gebied van modellen maar bovenal voor de dingen die je me hebt geleerd over het proces rond het doen van onderzoek. Ik hoop nog veel van je te mogen leren de komende jaren.

Beste Robert, ik wil je bedanken voor het vertrouwen dat je me altijd hebt gegeven. Je liet vaak blijken dat wat ik deed een wezenlijke bijdrage leverde aan de zorg. Je gaf me ook de kans om deel te nemen aan verschillende initiatieven, bijvoorbeeld rondom de N0 hals. Dit heeft het plezier en enthousiasme waarmee ik aan mijn proefschrift heb gewerkt erg vergroot. Ook kon ik altijd bij je terecht voor vragen, waarbij ik het bewonderenswaardig vond hoe snel je altijd kon schakelen en meteen begreep hoe de modellen waar ik mee bezig was in elkaar zaten.

Bij het modelleren gaat het om het verbinden van puzzelstukjes, maar dan moet je die puzzelstukjes wel hebben. Ik wil iedereen bedanken die me heeft geholpen bij het verzamelen van informatie. De artsen van de hoofd-halswerkgroep en vele anderen binnen en buiten het Radboudumc waren altijd bereid om met mij te spreken of om te helpen bij het verkrijgen van de benodigde informatie.

Ik wil graag de co-auteurs van de artikelen in dit proefschrift bedanken voor hun onmisbare bijdrage.

Tijdens mijn promotie heb ik een aantal studenten begeleid tijdens hun bachelorstage. Een aantal daarvan heeft ook meegeschreven aan een artikel. Ik wil jullie bedanken voor jullie inzet, gezelligheid en de dingen die ik van jullie heb geleerd. Ook de andere stagiaires op de afdeling wil ik bedanken voor hun bijdrage aan de leuke sfeer op de afdeling.

De mensen van de afdeling operatiekamers, bedankt voor de hulp bij het uitzoeken van allerlei vragen rondom chirurgische procedures. Ook bedankt voor de leuke momenten tijdens verschillende evenementen, zoals de opening van MITeC.

De andere promovendi van de evidence-based surgery groep wil ik ook graag bedanken. Ondanks dat we verdeeld zitten over verschillende afdelingen, vind ik het erg fijn om onderdeel te zijn van deze groep. Bedankt voor de hulp bij uiteenlopende vragen. Bedankt ook voor de mooie momenten tijdens de vis vrijdagen, de LVVDMBs, de EBS-lunches, de heidagen en natuurlijk tijdens de Vierdaagse.

Beste Gerjon, jij was mijn stagebegeleidertijdens mijn masterstage, die tevens het begin van dit proefschrift betekende. Bedankt voor de vliegende start die je me hebt gegeven en voor alle hulp in de jaren daarna.

Ik wil ook de mensen bij MedValue bedanken. Allereerst Richard, super dat ik vanaf het begin betrokken ben bij MedValue. Samen met Maroeska heb jij me de kans gegeven om naast mijn promotie MedValue mee te ontwikkelen. Bedankt voor de dingen die ik daarmee, en van jou persoonlijk, heb mogen leren. Ook bedankt voor de leuke autoritten. Miranda, bedankt vooral alle hulp, onder andere bij het organiseren van veel dingen bij MedValue die naast mijn promotie liepen. Sabine bedankt voor de fijne samenwerking de laatste tijd.

Beste Maarten, jij nam me meteen op sleeptouw toen ik startte met mijn stage. Ik heb ontzettend veel aan je gehad tijdens mijn promotie. We hebben ook heel veel gelachen en leuke dingen naast het werk gedaan. Dit heeft eraan bijgedragen dat ik mijn promotie met veel plezier heb kunnen doen. Heel erg bedankt daarvoor.

Beste Jorte, al tijdens de studie heb je me veel geholpen en hebben we veel en fijn samengewerkt. Ik was dan ook erg blij dat we weer samen konden werken. Niet alleen bij MedValue heb ik veel aan je , maar ook bij de laatste loodjes van mijn promotie heb je me weer ontzettend geholpen. Bovenal ben je gewoon een hele fijne collega en een fijn mens. Bedankt!

Beste Casper, altijd sta je meteen voor me klaar als ik een vraag heb over de medische praktijk of eigenlijk wat dan ook. Jouw inbreng is daarmee van grote waarde geweest voor de artikelen in dit proefschrift, maar zeker ook voor wat ik nu allemaal doe. Sorry voor alle ..on the road agains. Ik hoop dat we nog lang samen kunnen optrekken binnen en buiten het werk.

Mijn familie en vrienden wil ik natuurlijk ook bedanken. Het plezier dat ik met jullie heb is heel belangrijk voor mij. Oma, bedankt dat je me af en toe zei dat ik wel een beetje 'vort moest doen' met promoveren. Rick bedankt voor de mooie cover van mijn proefschrift. Nick, bedankt voor alle 'slappe klets' waar vaak alleen wij om kunnen lachen. Mam en Jos, bedankt dat jullie alles voor mij mogelijk hebben gemaakt en dat ik altijd op jullie terug kan vallen.

Lieve Lisanne, dat wij het zo leuk hebben samen zorgt ervoor dat veel vanzelf gaat. Jij zegt altijd dat ik het nooit heel zwaar leek te hebben tijdens mijn promotie, maar dat komt ook omdat ik dat zo vergeten ben als we thuis weer samen aan het lachen zijn. Daarmee heb je een hele grote bijdrage geleverd aan dit proefschrift (ook al vond je zelf van niet ;)).