

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



**VALUING FOLLOW-UP PROGRAMS IN
HEAD AND NECK CANCER**

MICHELA MEREGAGLIA

**Thesis submitted in accordance with the requirements for the degree
of
Doctor of Philosophy of the
University of London**

JULY 2018

Department of Health Services Research and Policy

Faculty of Public Health and Policy

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

No funding received

DECLARATION BY CANDIDATE

I, Michela Meregaglia, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

SIGNED:



DATE: July 27th, 2018

ABSTRACT

This thesis deals with the topic of head and neck cancer (HNC) follow-up from a health economics perspective. Despite recent advances in treating primary HNC, the long-term prognosis of these patients is still poor due to a high risk of cancer recurrence. Until now, there is no agreement about the best way of monitoring patients after the end of therapies. Moreover, patients' preferences for alternative surveillance schemes are unknown. A multicentre randomized controlled trial comparing two follow-up strategies of different intensity is currently ongoing in Italy. This thesis aims at filling some of the literature "gaps" around HNC surveillance, using Italy as a case study. The first chapter introduces the topic. The second chapter is a systematic literature review and critical appraisal of economic evaluation studies of post-treatment follow-up programs in any cancer type. The third chapter is a systematic literature review and quality appraisal of studies reporting original health state utility values in HNC, with a focus on articles addressing the post-treatment phase. The fourth chapter maps the EuroQol 5-Dimension 5-Level (EQ-5D-5L) utility values from two cancer-specific measures developed by the European Organization for Research and Treatment of Cancer (EORTC) by using a variety of regression techniques (linear, Tobit, mixture models) and several EQ-5D-5L country tariff sets; the developed functions are useful to inform future economic evaluations in HNC. The fifth chapter presents an exploratory model-based economic evaluation of the two follow-up strategies under investigation in the trial, where an intensive program of radiological assessments is compared to a symptom-driven surveillance; the cost analysis is conducted from a regional healthcare system perspective in Italy. Lastly, the sixth chapter presents a discrete choice experiment using best-worst scaling to elicit patients' preferences during follow-up at the National Cancer Institute (Milan, Italy).

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof John Cairns, an excellent Professor but, above all, a great person. Many thanks, John, for your time, patience, and kindness.

I am very thankful to my colleagues from Bocconi University, who trusted me unconditionally during the last years: Giovanni Fattore, Alessia Melegaro, Rosanna Tarricone, and Aleksandra Torbica.

I would like to express my deep gratitude to Dr Paolo Bossi and his colleagues from the National Cancer Institute of Milan for their precious collaboration.

I am grateful to my family who have always supported me. My thoughts go to my beloved grandmother, Agnese, who is not with us anymore: I am sure she can see me and is proud of me.

Finally, I would like to thank my husband Luca, because the 'value' of our joint achievements goes beyond this project.

TABLE OF CONTENTS

DECLARATION BY CANDIDATE	2
ABSTRACT	3
Acknowledgments	4
TABLE OF CONTENTS	5
LIST OF FIGURES	9
LIST OF TABLES	10
LIST OF APPENDICES	12
LIST OF ABBREVIATIONS	13
1 BACKGROUND AND AIMS	15
1.1 Thesis’s overall purposes	15
1.2 Head and neck cancer: basic facts	16
1.3 Head and neck cancer: follow-up	19
1.4 The HETeCo trial	24
1.5 Measuring ‘value’ in oncology: the QALY approach	25
1.6 Measuring ‘value’ in oncology: alternative frameworks	29
1.7 ‘Value’ in head and neck cancer follow-up	33
1.8 Knowledge “gaps”	35
1.9 Thesis overview	38
2 ECONOMIC EVALUATIONS OF FOLLOW-UP STRATEGIES FOR CANCER SURVIVORS: A SYSTEMATIC REVIEW AND QUALITY APPRAISAL OF THE LITERATURE	43
2.1 Introduction	43
2.2. Methods	45
2.2.1 Study identification and selection	45
2.2.2 Study quality assessment	46
2.3 Results	47
2.3.1 Study characteristics	47
2.3.2 Study findings	62
2.3.3 Study quality assessment	65
2.4 Discussion	73
2.4.1 Summary of evidence	73
2.4.2 Critical issues	75

2.4.3 Strengths and limitations	76
2.5 Conclusions	78
3 A SYSTEMATIC LITERATURE REVIEW OF HEALTH STATE UTILITY VALUES IN HEAD AND NECK CANCER	80
3.1 Introduction	80
3.2 Methods	83
3.3 Results	86
3.3.1 Study selection	86
3.3.2 Study characteristics	88
3.3.3 Study quality assessment	95
3.3.4 Overview of HSUVs	96
3.4 Discussion	98
3.5 Conclusions	102
4 MAPPING THE CANCER-SPECIFIC EORTC QLQ-C30 AND QLQ-H&N35 TO THE GENERIC EQ-5D-5L IN HEAD AND NECK CANCER	104
4.1 Introduction	104
4.2 Methods	106
4.2.1 Instruments	107
4.2.2 Sample and data collection	108
4.2.3 Data analysis	109
4.3 Results	112
4.3.1 Descriptive statistics	112
4.3.2 Mapping functions	116
4.4 Discussion	129
4.5 Conclusions	132
5 THE USE OF INTENSIVE RADIOLOGICAL ASSESSMENTS IN ROUTINE SURVEILLANCE AFTER TREATMENT FOR HEAD AND NECK CANCER: AN ECONOMIC EVALUATION	134
5.1 Introduction	134
5.2 Methods	136
5.2.1 Clinical studies	136
5.2.2 Model structure	138
5.2.3 State-transition matrix	140
5.2.4 Health state utility values	144
5.2.5 Cost data	145
5.2.6 Cost-effectiveness analysis	150

5.2.7 One-way sensitivity analysis	150
5.2.8 Two-way sensitivity analysis	151
5.2.9 Probabilistic sensitivity analysis	151
5.2.10 Expected value of perfect information.....	152
5.3 Results	153
5.3.1 Base-case analysis.....	153
5.3.2 Sensitivity analyses	153
5.3.3 Expected value of perfect information.....	157
5.4 Discussion	158
5.4.1 Synthesis of results	158
5.4.2 Comparison with other studies	160
5.4.3 Limitations	161
5.5 Conclusions	164
6 ELICITING PREFERENCES FOR CLINICAL FOLLOW-UP IN HEAD AND NECK CANCER PATIENTS USING BEST-WORST SCALING	165
6.1 Introduction	165
6.2 Methods	166
6.2.1 Experimental design	166
6.2.2 Literature review	167
6.2.3 Questionnaire development	168
6.2.4 Recruitment and setting	172
6.2.5 Statistical analysis.....	173
6.3 Results	175
6.3.1 Sample characteristics.....	175
6.3.2 Best and worst choice counts.....	176
6.3.3 Conditional logistic regression analysis	178
6.3.4 Covariate-adjusted regression analysis.....	179
6.3.5 Scenario's utilities.....	183
6.3.6 Patients' evaluation of the experiment.....	186
6.4 Discussion	187
6.5 The use of best-worst scaling utilities for health economic comparison: fact or fiction?	190
6.6 Conclusions	193
7 DISCUSSION AND CONCLUSIONS.....	195
7.1 Introduction	195

7.2 Main findings	197
7.3 Limitations.....	204
7.4 Contributions to research.....	205
7.5 Policy implications	206
7.6 Future research	208
7.7 Conclusions	210
REFERENCES.....	212
APPENDICES	239
APPENDIX TO CHAPTER 2.....	240
APPENDIX TO CHAPTER 3.....	242
APPENDIX TO CHAPTER 4.....	257
APPENDIX TO CHAPTER 5.....	292
APPENDICES TO CHAPTER 6	294

LIST OF FIGURES

CHAPTER 1

- Figure 1.1** The potential ‘value’ of follow-up programs in improving survival..... 35
Figure 1.2 Data sources and study perspectives in thesis’s chapters.41

CHAPTER 2

- Figure 2.1** Flow diagram.48

CHAPTER 3

- Figure 3.1** Free-text terms for electronic database searching..... 85
Figure 3.2 PRISMA flow chart. 87

CHAPTER 4

- Figure 4.1** Observed EQ-5D-5L utility values. 116
Figure 4.2 Scatterplots of observed versus predicted EQ-5D-5L for linear-mixed models using QLQ-C30. 123
Figure 4.3 Scatterplots of observed versus predicted EQ-5D-5L for linear-mixed models using QLQ-H&N35. 123
Figure 4.4 Scatterplots of observed versus predicted EQ-5D-5L for ALDVM models using QLQ-C30. 128
Figure 4.5 Scatterplots of observed versus predicted EQ-5D-5L for ALDVM models using QLQ-H&N35..... 128

CHAPTER 5

- Figure 5.1** Markov state-transition diagram. 138
Figure 5.2 Two-way sensitivity analysis of the proportion of potentially salvageable recurrences (or second primaries) detected over the 5-year follow-up..... 155
Figure 5.3 Cost-effectiveness plane with Monte Carlo simulations (5,000 runs)..... 156
Figure 5.4 Cost-effectiveness acceptability curve (CEAC). 157
Figure 5.5 Expected value of perfect information (EVPI). 158

LIST OF TABLES

CHAPTER 1

Table 1.1 HNC prevalence (per 100,000) by age, gender, and geographical area in Italy (2010).....	17
Table 1.2 Frequency of clinical/endoscopic evaluations and radiological assessments according to the mostly adopted guidelines in Italy.....	22
Table 1.3 5-year follow-up programs under evaluation in the HETeCo trial.	25
Table 1.4 Knowledge “gaps” and objectives of the thesis.	42

CHAPTER 2

Table 2.1 Search keywords according to PICOS elements.	45
Table 2.2 Key features of economic evaluation studies included in the review (n=53).50	
Table 2.3 Quality assessment of the included studies (n=53) based on the CHEERS checklist (<i>Items: 1-12</i>).....	69
Table 2.4 Quality assessment of the included studies (n=53) based on the CHEERS checklist (<i>Items: 13-24</i>).....	71

CHAPTER 3

Table 3.1 Studies estimating HSUVs in HNC using direct or indirect methods (n=25).	89
Table 3.2 Mapping studies predicting HSUVs in HNC (n=3).	94
Table 3.3 Quality assessment criteria in studies using direct or indirect methods (n=25).	97

CHAPTER 4

Table 4.1 Sample’s characteristics.	113
Table 4.2 Summary statistics of EQ-5D-5L and QLQ-C30/-H&N35 scores.....	114
Table 4.3 Spearman’s rank correlation coefficients between EQ-5D-5L utilities and QLQ-C30/-H&N35 summary scores.	117
Tables 4.4 Spearman’s rank correlation coefficients between QLQ-C30/-H&N35 summary scores.	118
Table 4.5 Synthesis of best performing linear mixed-effects models using QLQ-C30.	121
Table 4.6 Synthesis of best performing linear mixed-effects models using QLQ-H&N35.....	122
Table 4.7 Synthesis of best performing ALDVM models using QLQ-C30.....	124
Table 4.8 Synthesis of best performing ALDVM models using QLQ-H&N35.	126

CHAPTER 5

Table 5.1 Clinical parameters.....	141
Table 5.2 State-transition matrix.....	142
Table 5.3 Utility parameters.....	144
Table 5.4 Unit costs (€, 2016).....	147
Table 5.5 5-year follow-up costs in the HETeCo trial arms (A and B).....	148
Table 5.6 Health state description and monthly costs.....	149
Table 5.7 Cost-effectiveness results (base-case analysis).....	153
Table 5.8 One-way sensitivity analysis.....	154
Table 5.9 Two-way sensitivity analysis of the proportion of potentially salvageable recurrences (or second primaries) detected over the 5-year follow-up.....	155

CHAPTER 6

Table 6.1 Previous studies using stated preference techniques in cancer follow-up....	169
Table 6.2 Attributes and levels (in relation to the HETeCo trial arms).....	171
Table 6.3 Orthogonal scenarios.....	171
Table 6.4 Dataset example.....	174
Table 6.5 Sample's characteristics.....	177
Table 6.6 Frequency of best and worst selections across scenarios.....	178
Table 6.7 Utility coefficients from paired conditional logistic regression analysis.....	179
Table 6.8 Utility coefficients from covariate-adjusted conditional logistic regression analysis.....	180
Table 6.9 Utilities of the survey scenarios (including the HETeCo trial arms).....	185
Table 6.10 Self-reported difficulties in performing the exercise.....	186
Table 6.11 Cost-utility comparison using the survey results.....	192

LIST OF APPENDICES

CHAPTER 2

Table A2.1 PRISMA 2009 checklist.....	240
--	-----

CHAPTER 3

Table A3.1 Details of HSUVs reported by the studies (n=27).....	242
--	-----

CHAPTER 4

Table A4.1 MAPS checklist of included items.	257
Tables A4.2 Linear mixed-effects and random-effects Tobit models using QLQ-C30.	259
Tables A4.3 Linear mixed-effects and random-effects Tobit models using QLQ- H&N35.....	266
Tables A4.4 ALDVM models using QLQ-C30.	273
Tables A4.5 ALDVM models using QLQ-H&N35.	282

CHAPTER 5

Table A5.1 CHEERS checklist - items to include when reporting economic evaluations of health interventions.....	292
---	-----

CHAPTER 6

Annex A6.1 BWS questionnaire.	294
Annex A6.2 BWS questionnaire (original Italian version).	302
Tables A6.1-9 Utility coefficients from covariate-adjusted univariate conditional logistic regression analyses.	310

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
AIRO	Italian Association of Oncology Radiotherapy
ALDVMM	Adjusted Limited Dependent Variable Mixture Model
AOOI	Italian Hospital Otorhinolaryngologic Association
ASCO	American Society of Clinical Oncology
BIC	Bayesian Information Criterion
BWS	Best-Worst Scaling
CEAC	Cost-Effectiveness Acceptability Curve
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CT	Computed Tomography
DCE	Discrete Choice Experiment
DRGs	Diagnosis-Related Groups
EORTC	European Organization for Research and Treatment of Cancer
EQ-5D-3L	EuroQol Five-Dimension Three-Level
EQ-5D-5L	EuroQol Five-Dimension Five-Level
ESMO	European Society for Medical Oncology
EVPI	Expected Value of Perfect Information
FACT-G	Functional Assessment of Cancer Therapy - General
FACT-H&N	Functional Assessment of Cancer Therapy - Head and Neck Cancer
HERC	Health Economics Research Centre
HETeCo	Health and Economic Outcomes of Two Different Follow-Up Strategies in Effectively Cured Advanced Head and Neck Cancer
HNC	Head and Neck Cancer
HPV	Human Papilloma Virus
HRQoL	Health-related Quality of Life
HSUV	Health State Utility Value
HTA	Health Technology Assessment

ICER	Institute for Clinical and Economic Review
ICER	Incremental Cost-Effectiveness Ratio
ICUR	Incremental Cost-Utility Ratio
ISPOR	International Society of Pharmacoeconomics and Outcomes Research
LYG	Life Year Gained
MAUI	Multi-attribute utility instrument
MRI	Magnetic Resonance Imaging
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NICE	National Institute for Health and Care Excellence
OLS	Ordinary Least Squares
PET	Positron Emission Tomography
PICOS	Population, Intervention, Comparator, Outcome, Study (design)
PRISMA	Preferred Reporting System for Systematic Review and Meta-Analysis
PSA	Probabilistic Sensitivity Analysis
QALY	Quality-Adjusted Life Year
QLQ-C30	30-item Core Quality of Life Questionnaire
QLQ-H&N35	35-item Head and Neck Cancer Quality of Life Questionnaire
RCT	Randomized Controlled Trial
SCCHN	Squamous Cell Carcinoma of the Head and Neck
SG	Standard gamble
TTO	Time trade-off
UWQOL	University of Washington Quality of Life
VAS	Visual Analogue Scale

1 BACKGROUND AND AIMS

1.1 Thesis's overall purposes

This thesis was conceived around the concept of 'value' in the follow-up of head and neck cancer (HNC) patients. Contrary to therapeutic strategies, post-treatment surveillance in HNC has not been deeply studied yet, nor have costs and outcomes been jointly evaluated in any health economic framework. This thesis work was conducted alongside an ongoing randomized controlled trial (RCT), which generated research questions and was the basis for multiple data collections. The trial is going to estimate the 'value', from a health economic perspective, of an intensive follow-up program of radiological assessments, as designed by the trial's investigators, compared to symptom-driven surveillance as recommended by international guidelines.

The thesis contributes in several ways to the clinical study and to current knowledge about post-treatment surveillance in HNC. First, using an exploratory modelling framework, the thesis provides the expected lifetime outcomes and costs arising from the two follow-up interventions investigated in the trial. Second, the thesis generates evidence on aspects not considered within the trial, such as patient's preferences for features of follow-up programs in HNC, including the two under evaluation in the clinical study. Moreover, the thesis provides a systematic review of economic evaluation studies of follow-up programs in oncology to be compared with the modelling study's results. Following a similar systematic approach, the thesis collects health state utility values (HSUVs) from the HNC literature, which are required to calculate quality-adjusted life year (QALY), the main outcome of the cost-effectiveness model. The review provides useful data especially for recurrent and palliative states, which are not considered in the ongoing RCT, since health-related quality of life

(HRQoL) questionnaires are not being administered after the patient's relapse. Lastly, a statistical relationship between generic and HNC-specific HRQoL tools is established, using the most recent recommendations from the mapping literature.

1.2 Head and neck cancer: basic facts

HNC is the sixth most common cancer in the world with around 690,000 incident cases and 375,000 deaths reported in 2012 [1]. In the same year, there were 140,000 new diagnoses and 63,500 deaths in Europe [2]. Squamous cell carcinoma of the head and neck (SCCHN) is the predominant histological type concerning around 91% of all HNC cases. The remaining cancers are sarcomas (2%) and adenocarcinomas, melanomas, or not well specified tumours (7%) [2]. HNC includes a heterogeneous group of malignancies located in the upper respiratory and digestive tract with the most prevalent site being the oral cavity followed by the larynx and the pharynx. Alcohol and tobacco abuse are traditional etiological factors for the disease with a synergistic effect and are estimated to be responsible for at least 75% of all HNC cases [3]; additional risk factors for HNC development are poor dental hygiene, limited consumption of fruit and vegetables and genetic predisposition to cancer [2] [4].

The incidence rate in Italy has recently been assessed at 16 per 100,000 [3], with considerable differences between men (30.2 per 100,000) and women (5.1 per 100,000) as recently estimated for the Sardinia region [4]. The risk of being diagnosed with HNC, indeed, is seven times higher in men [3] although an increasing trend in incidence in women only can be observed in recent decades [1] [4]. In 2010, the number of people living with HNC in Italy was equal to 111,520 (4.3% of the overall cancer survivors); of these, the proportions diagnosed more than 2, 5 and 10 years earlier were 84%, 67% and 45%, respectively [3]. HNC prevalence was higher in Northern regions (\approx 230 per 100,000 inhabitants) compared with Central (164 per 100,000) and Southern ones (154

per 100,000) and in people aged ≥ 75 years (783 per 100,000) compared to the overall population (199 per 100,000) (Table 1.1). No significant differences in survival have been shown by gender and geographical areas; the 5-year survival rate, indeed, stratified by gender, ranges between 54% in women living in Central Italy and 59% in those living in Northern regions [5]. The differences in HNC prevalence across Italy may be explained by different incidence rates [3] [6], likely due to higher consumption of alcoholic drinks in the North and in the Centre compared to the South [7]. There is no difference in the prevalence of smokers between the Northern and Southern regions, but the proportion of smokers is significantly higher in Central Italy [8].

Table 1.1 HNC prevalence (per 100,000) by age, gender, and geographical area in Italy (2010).

		0-44	45-59	60-74	75+	All ages
North-West	Male	17	229	896	1808	372
	Female	13	77	197	298	101
	Total	15	151	520	839	231
North-East	Male	16	231	918	1756	354
	Female	13	79	220	315	103
	Total	15	155	551	843	225
Centre	Male	12	165	628	1381	263
	Female	15	65	121	241	72
	Total	13	114	362	684	164
South and Islands	Male	14	201	741	1346	250
	Female	12	73	128	235	65
	Total	13	135	417	670	154
Italy	Male	15	216	840	1625	318
	Female	13	76	180	283	88
	Total	14	145	490	783	199

Source: AIRTUM [6].

The epidemiology of HNC is rapidly changing. Over the last decade, there has been a shift in the primary site distribution, with a steady increase in oropharyngeal and oral cavity cancers and a slight decline of cases in the larynx, hypopharynx, and nasopharynx [2] [9]. Indeed, an epidemic of oropharyngeal cancers (particularly those of lingual and palatine tonsils) caused by the Human Papilloma Virus (HPV) has recently emerged in younger age groups compared with other HNC types [10] [11] [12] [13] [14]. Therefore, the overall incidence has remained substantially stable over the

past 10 years despite declining smoking habits in Europe [2]. The great majority of HPV-related oropharyngeal cancer is due to the HPV16 serotype [4] [12] [15]. HPV-positive patients are typically middle-age, non-smoking men with high socio-economic status and a history of exposure to multiple sexual partners; their prognosis is substantially better than for HPV-negative tobacco-related cancer patients treated similarly [9]. HNC also presents a marked socio-economic gradient in survival between affluent and deprived patients or geographical areas (e.g. Northern vs. Eastern Europe) [2].

In early stage SCCHN (about 20%) the standard clinical therapy is surgery and/or radiation therapy [13] [16], while another fifth diagnosed with metastatic disease can only aim at systemic chemotherapy [17] [18]. Most patients (around 60%) who present instead with locally advanced disease usually receive multimodality treatments including concurrent radiotherapy and chemotherapy with surgery, if indicated [17]. Despite aggressive multimodal therapy, the 5-year overall survival is 50%-60% [3] [11], mainly due to loco-regional or distant relapses occurring within few years after the end of primary treatments [10] [19]. Cancer recurrence is usually defined as the re-emergence of the disease after a post-treatment six-month period of complete regression; the risk varies from 10% to 50% according to cancer site and stage [20]. Moreover, there is a lifetime risk of developing second primary tumours of around 3% per year [16] [21] [22] [23] and the most frequent localizations are head and neck, lung, and oesophagus [24].

A few patients with loco-regional recurrences or second primaries can be salvaged by surgery or re-irradiation [16] [17]. In patients with resectable cancers and good health status, surgical salvage remains the best option for long-term disease-free and overall survival; however, in recent years, non-surgical therapies such as re-irradiation have also demonstrated significant improvements in loco-regional control [25]. There exist

different types of re-irradiation including conventional techniques, intensity modulated radiotherapy, and stereotactic body radiotherapy [26]. The overall survival of patients treated with re-irradiation at 2 years range from 30% to 67% according to the technique adopted and disease extent [23]. Early diagnosis of HNC relapse or second primary cancer allows these options to be implemented more frequently and with better outcomes. However, most patients with recurrent or metastatic disease only qualify for palliative treatment, comprising supportive care alone, or in addition to, single-agent chemotherapy or chemotherapy combined with biological agents [9] [17]. The administration of cetuximab plus 5-fluoracil and cisplatin is currently the standard of care in recurrent patients not amenable to surgery or re-irradiation [27]. Despite the choice of treatment, the prognosis is poor with a median survival that does not exceed 10-12 months [10].

1.3 Head and neck cancer: follow-up

Post-treatment follow-up is a well-established service in oncology, due to the risk of relapse experienced by treated cancer patients. The primary objective of follow-up, indeed, is the early detection of loco-regional recurrences, metastases and second primaries that are likely to occur in the years following any primary treatment with curative intent [20] [28]. In HNC, follow-up visits should include a physical examination of the head and neck region, an assessment of vocal, breathing, and swallowing functions, and a pain evaluation [23]. As secondary aims, surveillance programs should manage treatment-related side effects and late complications, provide psychological and social support to patients and families, and discourage patients from dangerous habits that contributed to the development of the primary cancer [16] [28]. In HNC, dental care, nutritional counselling, speech, and swallowing rehabilitation are interventions routinely delivered during follow-up [23]. At population level, the follow-

up aims at collecting data on the efficacy of the available cancer therapies and other clinical and epidemiological variables that are useful to improve the future provision of healthcare services in oncology [29].

Several modalities for monitoring patients treated for HNC have been proposed in clinical practice to fulfil the primary aims of follow-up [16] [20] [23]. Overall, the published guidelines are concordant in recommending a 5-year hospital-based program with frequency of visits decreasing over time (although the time interval between clinical appointments is controversial). From the 6th year onwards, follow-up visit is advised every year, especially for patients at high risk of second primaries [16] [20]; in some cases, the general practitioners conduct a minimal surveillance, with referral to the specialist doctors for doubtful cases. There are less homogeneous indications instead about the number and type of diagnostic tests to be prescribed during follow-up [29]. Overall, an imaging test should be administered between 2 and 6 months since the end of primary therapies to assess the treatment response and provide a benchmark for future evaluations. Subsequently, routine imaging is advisable for patients in whom there is a clinical suspicion of cancer relapse and/or clinical assessment and endoscopic visualization are not reliable or sufficient [16] [29]. Computed tomography (CT), magnetic resonance imaging (MRI), and 18-F-fluorodeoxyglucose positron emission tomography (18-F-FDG-PET, thereafter PET) are standard diagnostic techniques in HNC. CT is routinely used to assess initial response to treatment, MRI has the advantage of providing a better soft-tissue differentiation, while PET should be helpful in distinguishing tissue necrosis (following primary surgery or radiotherapy) from recurrent tumour [20]. PET is also recommended for restaging patients who are being considered for major salvage surgery [16]. Clinical guidelines do not usually provide specific indications regarding the best radiological technique to adopt during follow-up, although MRI and PET are usually preferred over CT in identifying early recurrences;

PET, however, is usually recommended as a “second level” examination due to its high costs [29].

Since there are no standardized and universally accepted guidelines, the contents of follow-up in terms of frequency of clinical and radiological investigations are often at discretion of local centres in Italy [23] [30]. The main clinical guidelines to which Italian oncologists refer are summarized in Table 1.2. Among them, the National Comprehensive Cancer Network (NCCN) recommends a program of outpatient visits according to cancer subsite and a baseline radiological assessment within 6 months of the treatment ending; further imaging is recommended based on signs and symptoms over the course of follow-up. Similarly, the European Society for Medical Oncology (ESMO) suggests the adoption of clinical evaluations and imaging to monitor HNC survivors, but without specifying a timetable of visits; a baseline radiological assessment should be performed once after primary treatment, and subsequently, only when recurrence is suspected. The guidelines developed by Italian scientific associations are generally more intensive. Among them, the Italian Association of Oncology Radiotherapy (AIRO) prescribes imaging at fix time points over the 5-year period irrespective of the symptoms reported by the patient. Similarly, the Italian Hospital Otorhinolaryngologic Association (AOOI) suggests performing a yearly MRI or CT scan for high-risk patients and PET scans during the first two years (subsequently only for high-risk patients) [29]. More recently, a follow-up program proposal distinguishing between clinically evaluable and not evaluable primary tumours has been published by a research group in Italy [30]. Moreover, a chest CT or x-ray can be prescribed to heavy smokers (≥ 20 pack/years) or patients aged above 50, who are at increased risk of developing second lung primaries [3].

Table 1.2 Frequency of clinical/endoscopic evaluations and radiological assessments according to the mostly adopted guidelines in Italy.

	NCCN		BAHNO		AIOM		AIRO		AOOI		
	Clinical exam	Imaging	Clinical exam	Imaging	Clinical exam	Imaging	Clinical exam	Imaging	Clinical exam	Imaging MRI/CT	Imaging PET
Year 1	1-3 months	Once within 6 months	4-6 weeks	Once within 3 months	3 months	Once within 3 months	1-2 months	6 months	1-3 months	Once within 6 months	Once within 4-6 months
Year 2	2-6 months	-	4-6 weeks	-	3 months		2-3 months	6 months	2-4 months	Only HR	6 months
Year 3	4-8 months	-	3 months	-	3-6 months		4-6 months	6 months*	4-8 months	Only HR	Only HR
Year 4	4-8 months	-	6 months	-	6 months		4-6 months	6 months*	4-8 Months	Only HR	Only HR
Year 5	4-8 months	-	6 months	-	6 months		4-6 months	-	4-8 months	Only HR	
Year 6 onwards	Annually	-	Annually	-	Annually		Annually	-	Annually	Only HR	

NCCN: National Comprehensive Cancer Network; BAHNO: British Association of Head and Neck Oncologists; AIOM: Italian Association of Medical Oncology; AIRO: Italian Association of Oncology Radiotherapy; AOOI: Italian Hospital Otorhinolaryngologic Association; HR: high-risk.

* Only nasopharynx

Sources: [3] [20] [29].

Best practice in HNC follow-up remains uncertain. The primary aim of a surveillance program is to achieve earlier detection of recurrent cancer compared to patient self-identification through frequent clinical assessments at regular intervals. This task is justified by the high-risk of relapsing experienced by HNC patients. Early diagnosis is beneficial because it increases the opportunity for potentially salvageable treatment instead of palliative care. However, there are currently limited data regarding the survival benefits achievable by different follow-up programs [30] [31]. In particular, there is no RCT comparing a structured follow-up program with a patient self-referral policy in the literature [20] [29]. Some observational studies demonstrate a survival benefit in patients diagnosed at routine follow-up, compared to those who present spontaneously with symptoms. Conversely, other studies do not find any survival gains from detecting asymptomatic recurrences [23]. However, most clinical studies only report intermediate outcomes, i.e. the proportion of recurrences detected during scheduled appointments versus those identified thanks to symptoms referred by the patient, and dissenting conclusions are obtained across studies [29]. Additionally, the administration of intensive follow-up programs may encounter feasibility constraints, cause unnecessary discomfort to the patients, and have serious cost implications for healthcare systems [32]. In recent years, the oncological community has been arguing whether intensive follow-up assessments might pose a relevant economic burden without significant clinical benefits [23]. Therefore, a cost-effectiveness evaluation of alternative follow-up programs is recommended to avoid over-treatment and waste of scarce healthcare resources.

1.4 The HETeCo trial

The first RCT comparing alternative follow-up strategies in HNC entitled: “Health and Economic Outcomes of Two Different Follow-Up Strategies in Effectively Cured Advanced Head and Neck Cancer (HETeCo)” started in Italy in 2014. Full details are available in the trial protocol (clinicaltrials.gov identifier NCT02262221) [33]. In brief, 330 patients are being enrolled and followed-up for five years in several health centres across Italy and Switzerland. The centre leading the trial is the National Cancer Institute (NCI) located in Milan (Italy). Inclusion criteria are as follows: diagnosis of stage III-IV squamous HNC located in the oral cavity, oropharynx, hypopharynx, or larynx; if oropharyngeal cancer, HPV negative or HPV positive with a smoking history (i.e. more than 10 pack/years); radiotherapy (as a minimum) already administered as curative treatment or in post-operative setting; no evidence of disease for six months after treatment’s end; age ≥ 18 years. The patients enrolled in the study do not have to be potential candidates for salvage surgery, in case of recurrence or a second primary. Patients are excluded if they have a diagnosis of HNC in any site different from those reported in the inclusion criteria (e.g. nasopharynx, paranasal sinus, salivary glands, or unknown site), or a diagnosis of any other malignancies unless free of disease for at least five years or are unable to comply with the study requirements in the opinion of the investigators. Patients are randomized to symptom-driven surveillance (arm A) according to NCCN guidelines [34] or to more intensive follow-up (arm B), as described in Table 1.3. In both study groups, HRQoL questionnaires including the EuroQol five-dimensional five-level questionnaire (EQ-5D-5L) and the European Organization for Research and Treatment of Cancer (EORTC) Core Quality of Life Questionnaire (QLQ-C30), supplemented with the HNC module (QLQ-H&N35), are administered to patients at every other visit in the first two years and then at each visit, until the patient is free of disease. The primary study objective is to evaluate the two-

year cost-effectiveness of the two alternative follow-up programs. As secondary objectives, the trial aims to estimate the proportion of potentially salvageable loco-regional recurrences or second primaries, the cause-specific survival, and the overall survival of relapsing patients in both arms.

Table 1.3 5-year follow-up programs under evaluation in the HETeCo trial.

Arm A (non-intensive follow-up) – NCCN guidelines	Arm B (intensive follow-up)
Outpatient visits every 2-3 months in the first two years, every 4-6 months in the last three years (according to cancer subsite)	Outpatient visits every 2-3 months in the first two years, every 4-6 months in the last three years (according to cancer subsite)
Physical and fibre optic endoscopic head and neck examination at each visit	Physical and fibre optic endoscopic head and neck examination at each visit
Laboratory tests (complete blood count, renal, hepatic, and thyroid function) performed once a year	Laboratory tests (complete blood count, renal, hepatic, and thyroid function) performed once a year
Loco-regional imaging (MRI/CT scan) performed once at the beginning of follow-up and then recommended only at the occurrence of new signs or symptoms	Loco-regional imaging (MRI/CT scan) requested twice/year in the first two years and once/year in the third and fourth years
Patients instructed how to recognize signs or symptoms of recurrence	PET scan requested yearly in the first three years only in patients ≥ 50 years and with smoking history of ≥ 20 pack/years
Patients contacted by a phone call between visits to monitor potential recurrence symptoms	

CT: computed tomography; MRI: magnetic resonance imaging; HETeCo: Health and Economic Outcomes of Two Different Follow Up Strategies in Effectively Cured Advanced Head and Neck Cancer; NCCN: National Comprehensive Cancer Network; PET: positron emission tomography.

1.5 Measuring ‘value’ in oncology: the QALY approach

The concept of ‘value’, traditionally defined as the health outcome achieved per unit of cost, has recently become a central topic in cancer care due to rising cancer costs in all developed countries. Specifically, value should be measured by costing the full range of products and services delivered over the full cycle of care to achieve a pre-defined outcome. Whilst costs are relatively easy to identify and measure, what constitutes

meaningful patient outcome measures is still debated in the scientific community. Survival, also defined as life years gained (LYG), is undoubtedly the primary outcome to achieve (or to maximize) in most medical conditions [35]. However, even in a life-threatening disease such as cancer, survival is not the only measure of effectiveness and additional elements should be considered when evaluating a new treatment. In oncology, clinicians are particularly required to balance any survival gains with drug-related toxicities. As a composite measure combining length of life with the quality of the life itself, there is consensus that the QALY represents a good “proxy” for value, which has been increasingly recognized as a multidimensional concept. The theory underlying the QALY and synthesized in the paper by Weinstein et al [36] is that *individuals move across health states over time and attach a value to them*, which is defined indeed as HSUV; *a health state that is more desirable is also more valuable*. Thus, value is equated with “preference” or “desirability” or “utility”. HSUVs usually range between 0, which is the value attached to the state of death, and 1 corresponding to perfect health, although some methods allow for below zero utilities in case of states valued as worse than being dead. Accordingly, *the per patient outcome to maximize is defined as the value-weighted time accumulated over lifetime to yield QALYs*. The QALY is currently the standard outcome metric used in cost-effectiveness analysis to inform reimbursement decisions of cancer treatments in several countries.

In recent years, a discussion arose in the literature regarding the ability of the QALY to summarize appropriately the experience of cancer patients, and the use of alternative outcome metrics in the assessment of value in oncology [37]. Indeed, there are a few challenges in estimating HSUVs and, consequently, QALYs in cancer patients. The article from Devlin and Lorgelly [38] and from Garau et al [39] highlighted most of them. First, overall survival is required for the estimation of QALYs, whilst RCTs in oncology rarely are long enough to cover the full patient’s life span and tend to focus

instead on progression-free survival or other intermediate outcomes, which may not capture the whole treatment effect and be a valid surrogate for overall survival [40] [41] [42]. Moreover, post-progression therapies, either protocolled or prescribed once the patient has left the trial, may affect overall survival estimates [43].

Second, the measurement of HSUVs is not straightforward. There are a range of methods that can be used to estimate preference weights encompassing direct techniques (i.e. visual analogue scale, VAS; standard gamble, SG; time trade-off, TTO) and generic HRQoL preference-based tools, among which the most common one is EQ-5D. The TTO method is mostly used to generate EQ-5D preference weights and is accordingly the most preferred by the National Institute for Health and Care Excellence (NICE) among the direct techniques. Some argue that the ‘constant proportional trade-off’ assumption in TTO might be violated in the end-of-life state, since people with a short life expectancy are less likely to give up any of the remaining time to improve their health status [44] [45]. However, the current evidence is conflicting and not strong enough to support the hypothesis that HSUVs are higher for patients closer to death, but rather for those with a short time from terminal diagnosis, and any way this effect is likely to be small [46]. Additionally, the question of whether patients or the public should value health states (through direct methods or generating preference-based scoring algorithms for generic HRQoL measures) is debated in the literature, although many scholars argue that weights should come from the general population as taxpayers and potential patients. Moreover, in advanced stages, cancer patients are too sick to fill in questionnaires or take part in interviews, thus limiting the possibility of self-valuing their health state. According to others, however, the general population is at risk of underestimating the impact of living with cancer. Finally, some generic preference-based tools used to elicit HSUVs might not be the best way of assessing HRQoL in cancer, since they have been shown to not be sensitive enough to changes in patient’s

health status and to not cover relevant aspects in cancer, such as vitality, energy, or fatigue [47]. For this reason, most RCTs adopt cancer-specific questionnaires such as the EORTC QLQ-C30 or the Functional Assessment of Cancer Therapy – General (FACT-G), which are not provided with a preference-based algorithm or, if they have one (e.g. EORTC 8D), generate HSUVs that are hardly comparable across other disease areas.

Third, there is the question if the cost-effectiveness threshold should be different for cancer, especially at advanced and final stages. There is some evidence that individuals put a greater value on cancer therapies, and even on the “hope” of new therapies, since cancer is a frightening and common disease. In a recent survey [48], more than two thirds of cancer patients preferred “hopeful gambles” (i.e. treatments with a variety of outcomes but potentially offering a longer survival time) to “safe bets” (i.e. treatments providing similar average survival but less chance of large gain). These results suggest that current health technology assessment (HTA) techniques, which traditionally focus on average survival gains, might omit a relevant value element (i.e. hope) and should find a way to incorporate it. In England and Wales, higher cost-effectiveness thresholds have been adopted by NICE for end-of-life care since 2009; however, this might cause substantial financial pressure on healthcare services and potential QALY losses, unless a societal preference for end-of-life gains is clearly shown [49].

Fourth, the standard definition of ‘value’ only depends on the measures of outcome (i.e. QALY) and not on inputs or the process of care used [35]. There is growing interest in patient’s preferences regarding the process of care, such as the value attached to a home-based palliative care service versus a hospital-based one. These preferences have been shown to impact directly on the clinical outcome through psychological factors or indirectly via compliance and adherence [50]; however, despite their increasing use in treatment decision-making, there is little consensus on how cost-effectiveness analyses

should account for individual preferences to inform decisions on allocation of healthcare resources.

1.6 Measuring ‘value’ in oncology: alternative frameworks

In 2015, five health-related organizations developed frameworks to assess the value of oncological drugs. These new tools share with the traditional “cost per QALY” approach the overall idea that ‘value’ is determined by the amount of a treatment’s clinical benefit balanced against costs. However, they vary considerably in the definition of outcome, in the types of costs addressed, and in the purpose of the evaluation. A brief description of each tool follows.

1. The American Society of Clinical Oncology (ASCO) launched in 2015 a first version of its frameworks to assess the value of medical cancer therapies; an updated version appeared in 2016 in response to comments received via a web-based survey [51]. One framework deals with advanced, non-curative disease (scored out of 130), while the latter with potentially curative disease (i.e. adjuvant therapy, scored out of 100). Each incorporates three essential elements including clinical benefit, toxicity, and costs. In the advanced disease framework, the relative weights are modifiable according to patient’s preferences; for example, the toxicity score may have a higher weight than survival. The tool grants bonus points for some health gains including symptom palliation, HRQoL improvement, and survival increase at the end of the curve. The sum of clinical benefit, bonus points, and toxicity scores generates a Net Health Benefit (NHB), which represents a measure of relative improvement of a new treatment compared with the standard of care within a clinical trial. The net health benefit is reported with the direct treatment cost including the patient’s co-payment.

2. In the same year, ESMO developed the Magnitude of Clinical Benefit Scale (MCBS) as a framework to define the value of new anti-cancer drugs approved in Europe [52]. Drugs are assigned a clinical benefit grade, ranging from A to C in the curative (adjuvant) setting, and 1 to 5 in the non-curative setting. Overall survival and disease-free survival are the outcomes considered as primary efficacy measures, while lower scores are assigned to progression-free survival, response rate, time to response and HRQoL. Scores are awarded based on the level of clinical improvements in the relevant variables obtained within an RCT or cohort study. Costs are not included in this framework since European countries have their own pricing strategies.
3. The NCCN has incorporated ‘evidence blocks’ into their clinical guidelines [53]. Five domains are considered in this value framework including efficacy, toxicity, quality/quantity of evidence, consistency of evidence, and affordability; each dimension is scored between 1 (least favourable) and 5 (most favourable). The affordability domain includes drug costs, supportive care, administration costs, and monitoring/management of toxicities. The evidence blocks are currently available for ten cancer types within the NCCN guidelines.
4. The Memorial Sloan Kettering Cancer Centre in New York developed an interactive drug tool (Drug Abacus) to inform physician and policy-makers about the value of cancer drugs and to establish fair prices [54]. Based on clinical data and expert opinion, the tool measures six attributes including efficacy, tolerability, novelty, research and development (R&D) costs, rarity, and population health burden. Contrary to other value frameworks, the output of Drug Abacus is not a “value score” but rather a “price” that should be assigned to the drug.

5. The Institute for Clinical and Economic Review (ICER), which provides independent HTA reports on drugs, devices, and other medical services in the United States, introduced an ICER Value Assessment Framework to calculate a reasonable price for treatments (not only in oncology). The framework includes two broad components: “care value”, considering clinical effectiveness and incremental cost per clinical outcome achieved, and “health system value”, assessing the short-term budget impact. A cost-effectiveness threshold of \$100,000-150,000 is used to calculate a reasonable price range. Moreover, the drug’s potential budget impact should stay below \$904 million per year, above which a new treatment is likely to increase healthcare costs excessively [55].

All these “new” value frameworks have shown several limitations that make them difficult to overcome the QALY approach, at least in their current formulation. First, ‘value’ is still an indefinable concept and there is no consensus across the frameworks on what domains should it comprise. Second, many frameworks rely on evidence from RCTs only, disregarding any information coming from observational studies or other types of real-world data (e.g. cancer registries) [56]. Third, most of them focus on clinical efficacy defined as survival without incorporating HRQoL, which is only partially captured by toxicity domains in some frameworks. Fourth, in determining “value for money” most tools consider the drug acquisition cost only without assessing other relevant medical costs (e.g. cost saving from avoidance of surgical interventions), non-medical costs and productivity losses in the societal perspective. Fifth, the new frameworks explicitly address drug therapies in cancer, but hardly fit with the value assessment of non-drug interventions, such as screening programs or surgical procedures; consequently, these value assessment results are not comparable with other types of interventions or diseases from a health economics perspective.

Therefore, the QALY remains the standard of outcome measurement for economic evaluation. The QALY approach is well established internationally and increasingly accepted by most HTA agencies worldwide including emerging economies (e.g. China, Brazil) and developed ones (e.g. United States, where researcher and policy-makers are beginning to use QALYs) [38]. The cost-utility framework using QALYs captures most of the assessment elements (i.e. survival, HRQoL, costs) considered by alternative value frameworks suggested in the literature; moreover, it covers the full range of interventions provided in healthcare (and not drugs only). The ICER Value Assessment Framework only explicitly mentions QALYs and cost-effectiveness ratios for the price calculation, which is not the original purpose of the cost-utility approach. As a multidimensional concept, the QALY already provides a metric to incorporate survival and HRQoL within a unique value framework. Moreover, cost-utility analyses can be characterized according to the source of the clinical information. Model-based economic evaluations, unlike trial-based ones, synthesize all the available clinical evidence deriving from various studies and not limited to a single RCT; when adopting a lifetime horizon, these studies provide a framework to estimate all medical and non-medical costs arising from the alternative interventions. Additionally, the use of QALYs facilitates comparison of outcomes across different disease areas, which is relevant to achieve an efficient allocation of the healthcare budget. Thus, despite the challenges posed by its measurement, the QALY remains a powerful conceptual tool that has given a relevant contribution to decision-making in healthcare over the last few decades [36]. At the same time, the above-mentioned, newly developed approaches have the merit of enriching the debate about ways to incorporate patient's wishes into value definition within an emerging vision of personalized medicine. However, the enlargement of the definition of 'value' to include additional elements, such as patient's preferences or the quality of scientific evidence, still not considered by QALYs, requires a careful

consideration of the feasibility of measuring these aspects and systematically incorporating them into the HTA decision-making process.

1.7 ‘Value’ in head and neck cancer follow-up

The ‘value’ of non-drug interventions is less clearly identifiable compared to pharmaceuticals. Most economic evaluations in cancer care have been in the treatment phase, such as cost-effectiveness analyses of chemotherapeutics or radiotherapy techniques [57]. In 2013, in the United States, only 15% of the research portfolio in HNC was devoted to the cancer control, survivorship, and outcomes research area (<https://www.cancer.gov>). This is also due to more difficulties in performing RCTs for non-pharmaceutical interventions such as follow-up programs, where blinding is difficult to achieve (and to maintain), and its absence can lead to biased outcome assessment [58] [59]. Moreover, any survival gains of surveillance programs may only become evident after decades and hardly captured by time-limited RCTs. In HNC, this is particularly true after the recent epidemiological changes and the rising number of HPV-related cases with younger age at diagnosis, better prognosis, and longer life expectancy [1]. In the post-treatment-phase, the primary objective of follow-up is not the direct achievement of improved health status, but rather the avoidance of a worst one through the early identification of cancer relapses. Secondary functions such as psychological support, the management of side effects of primary treatment, and education for healthy lifestyles impact instead directly on HRQoL. Thus, even a program that was not able to show any survival benefits could be “valuable” in enhancing HRQoL, speeding-up the process of recovery and return to normal life. An increase in overall survival, however, remains the primary goal of follow-up, although only achievable through intermediate outcomes, i.e. the early identification of potentially curable recurrences and second primaries and timely administration of

salvage treatments. The measurement of intermediate endpoints, such as the rate of detection of potentially curable recurrences, represents useful clinical information routinely reported by studies in the field. However, if not supplemented with an extrapolation of final outcomes (i.e. survival) through statistical techniques [60], these results pose a problem of comparability with other healthcare interventions and disease areas from a health economics perspective. For this reason, model-based economic evaluations are particularly useful in assessing the cost-effectiveness of surveillance programs by projecting future costs and QALY over longer time horizons.

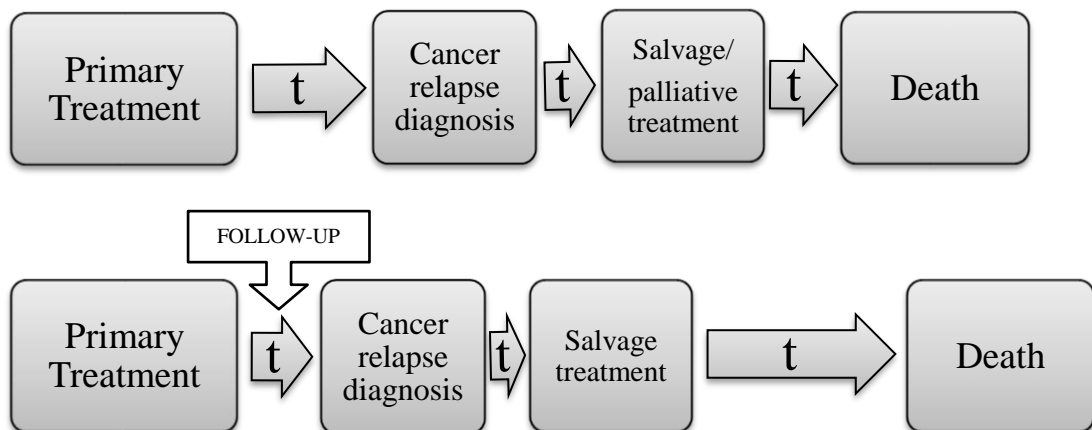
As well as cancer treatments, follow-up programs might have also some ‘side effects’. In particular, anatomical imaging (i.e. PET-CT) using radioactive substances might cause discomfort, fear, and long-term toxicities to patients, besides posing a substantial burden to healthcare systems and societies. Accordingly, any survival gains should be weighed against potential adverse effects and costs (i.e. medical and non-medical costs, productivity losses) arising from an intense use of these diagnostic tools. The long-term toxicities arising from prolonged exposure to PET-CT radiations might be important for younger, HPV-positive patients with a good prognosis, but be irrelevant for other HNC patients with poor life expectancy; thus, sub-groups analyses may be helpful in measuring value for this cancer population. Additionally, patients’ preferences regarding the healthcare services delivered in the post-treatment phase should be considered, although not explicitly addressed within the QALY approach. Lastly, the quantity and quality of evidence (pointed out by the NCCN in their Evidence Blocks) is another important aspect when evaluating post-treatment programs in oncology since, unlike drug interventions, their definition is frequently based on current practice or retrospective data analyses in the absence of prospective comparative trials.

1.8 Knowledge “gaps”

‘Value’ in the post-treatment phases of HNC remains largely unmeasured and follow-up services are often delivered without a sound scientific base. Within this thesis, a few “gaps” in the literature have been identified with respect to four main topics: (1) survival, (2) quality-adjusted survival using HSUVs, (3) costs and cost-effectiveness, (4) patient’s preferences (Table 1.4).

1. The first unsolved question is with respect to the primary outcome represented by overall survival, which is used in the QALY calculation (Figure 1.1). The current available evidence, indeed, is not able to state definitely whether an intensive follow-up program with frequent radiological examinations can detect recurrences and second primaries at a time when it is more feasible to administer potentially salvage treatments compared to a symptom-driven surveillance. Additionally, whether an early diagnosis of cancer relapse with subsequent administration of salvage treatments (i.e. surgery or re-irradiation) leads to significant survival gains remains even more controversial. So far, the available re-treatment options have shown a small possibility of long-term survival but a very high chance of treatment-related toxicity including death [61].

Figure 1.1 The potential ‘value’ of follow-up programs in improving survival.



2. As stated above, it is now commonly accepted that ‘value’ is a multi-dimensional concept implying that quantity of life should be weighed against the quality of life to evaluate any treatment’s effects. The QALY meets this requirement but its calculation needs utility values for each health state experienced by the patient during the process of care. At present, very few clinical studies report HSUVs for post-treatment, salvage treatments and palliative stages in HNC, thus limiting the possibility to assess properly the value of follow-up programs. This is mainly due to the poor health of cancer patients with advanced/terminal disease, which makes it difficult to collect patient-reported outcomes at these stages. Additionally, as in other disease areas, most clinical studies prefer to adopt cancer-specific questionnaires to evaluate HRQoL in oncological patients, since they target specific health issues and are more sensitive to changes in disease symptoms [62] [63]. In HNC, the most common tools are the EORTC QLQ-C30 (supplemented with the H&N35 module) and the Functional Assessment of Cancer Therapy - Head and Neck (FACT-H&N). These questionnaires were not originally provided with a scoring system for HSUVs calculation. The EORTC-8D was subsequently developed using the 30-item EORTC QLQ-C30 for use as a preference-based measure in cancer [64]; however, being a disease-specific tool, utility values from EORTC-8D cannot be compared with those derived from generic preference-based instruments (e.g. EQ-5D) and thus are not transferable to other therapeutic areas. In recent years, “mapping” or “cross-walking” has become a common technique to transform disease-specific scores onto preference-based utilities, but a function relating EORTC QLQ-C30 to EQ-5D has yet to be developed for HNC. Moreover, the techniques adopted so far to predict HSUVs from disease-specific scores have shown a poor fit with the typical EQ-5D data distribution.

3. Third, long-term cancer survivorship, although of unquestionable value for patients and societies, imposes a substantial financial burden on healthcare systems. The overall prevalence of people living after a diagnosis of HNC is increasing in all developed countries, with millions more patients who are going to be classified as survivors in the next few years [23]. Follow-up programs can consume a large amount of scarce healthcare resources and might cause delays in the diagnosis and treatment of new cancer patients. At the same time, the best way to save money is often *to spend more on some services to reduce the needs for others* [35], which might be translated, in this case, as investing resources on effective follow-up to avoid extra-morbidity in patients, such as metastases, with subsequent increase of medical and non-medical costs. Few attempts have been made so far to measure the healthcare costs associated with routine surveillance and secondary treatments in HNC, and no data are available for Italy. Moreover, no studies have tried to combine costs and outcomes (i.e. survival or QALYs) of alternative follow-up programs for HNC survivors to inform about an efficient allocation of scarce resources in oncology.
4. Fourth, the strength of preferences for alternative ways of delivering follow-up after the end of primary treatments for HNC has never been investigated with stated preference methods such as discrete choice experiments (DCEs). Most studies investigating patient's preferences in HNC are priority scales [18] not even focussed on follow-up services. Moreover, in tax-based healthcare systems, such as Italy, there is limited interest in estimating the willingness-to-pay for additional or improved services, thus reducing the amount of research devoted to preference elicitation. Although 'value' in healthcare is not simply a 'proxy' of the quality (or quantity) of care delivered [35], the incorporation of individual expectations within a value framework beyond the main clinical outcome(s)

might improve adherence to care, contain medical and non-medical costs and personalize value for each patient [18]. Research on preference elicitation with clear trade-offs among surveillance options is thus needed in HNC.

1.9 Thesis overview

This PhD thesis addresses some of the knowledge “gaps” existing around the ‘value’ assessment of follow-up in HNC (Table 1.4). The thesis is closely linked to the ongoing HETeCo trial in a bidirectional way; on one side, it relies on the clinical study to generate research questions, collect preliminary data, and gain clinical knowledge on the topic, whilst on the other it contributes to the trial by synthesizing evidence from the literature, testing hypotheses, and providing information on additional aspects not considered within the main clinical study.

In Chapter II, the published economic evaluations of cancer follow-up (in general) are reviewed systematically. This work has multiple purposes. First, it aims to identify any potential studies relating costs and outcomes in HNC follow-up. Second, it collects and synthesizes the available economic evaluation studies in the field of cancer surveillance. Third, the collection of these studies provided an opportunity to critically evaluate their quality and to identify any methodological weaknesses from the perspective of health economics research. For example, outcomes reported in the studies are identified as either intermediate (e.g. number of recurrences detected) or final (i.e. survival or quality-adjusted survival). Overall, the way of assessing the ‘value’ of follow-up services within an economic evaluation is investigated and commented, mainly to inform the modelling study presented in Chapter V and future data analyses within the HETeCo trial.

The work presented in Chapter III involves a systematic search of the published literature to identify all the available HSUVs in HNC, which can inform future economic evaluations in this area. Additionally, a quality evaluation of studies reporting original HSUVs is performed by referring to recent recommendations on the topic. This review is of interest on its own but also provides some parameter values for the model-based cost-utility analysis described in Chapter V, where results are expressed as incremental costs per QALY gained. Since the EQ-5D-5L is not collected after a patient's relapse in the HETeCo trial, HSUVs for recurrent and palliative states are retrieved from the literature to populate the model. Moreover, a broader discussion regarding the best techniques to measure HSUVs in HNC is provided within this chapter, to contribute to the current debate on QALY measurement in oncology.

Chapter IV presents a 'mapping work' aiming to derive HSUVs from two non-preference-based instruments frequently adopted to measure HRQoL in HNC patients. Using a sample of questionnaires collected from the patients in the HETeCo trial, a set of mapping functions is developed to predict EQ-5D-5L utility values from EORTC QLQ-C30 and H&N35 responses. A few EQ-5D-5L tariff sets (available at the time of the analysis) are used to generate HSUVs. The most recent techniques suggested by good practices and previous studies on mapping are adopted to contribute to knowledge advancements in the field. The developed functions might be useful for future researchers in studies not collecting EQ-5D data but planning to calculate QALYs.

Chapter V represents the 'core' of the research project and reports the first economic evaluation comparing follow-up programs of different intensity in HNC. An exploratory Markov model is developed to compare the same two surveillance strategies (arm A and B) under evaluation in the HETeCo trial and is populated with data from a variety of sources, such as the trial protocol, trial preliminary data, published literature, and expert opinion. Results are expressed as incremental cost-effectiveness ratios (ICERs) and

incremental cost-utility ratios (ICURs) and tested in extended deterministic and probabilistic sensitivity analyses. Moreover, because the uncertainties related to the efficacy parameters will be reduced only once the trial is completed, value of information methods are applied to estimate the ‘value’ of completing the ongoing clinical study.

The work presented in Chapter VI, using a best-worst scaling (BWS) technique, aims to elicit patients’ preferences for alternative ways of delivering follow-up in HNC. Since the clinical outcomes are still uncertain in the literature and under evaluation in the ongoing HETeCo trial, the survey is limited to analyse the preferences for process-related aspects of follow-up. Four attributes, each with three levels are identified to describe hypothetical follow-up programs in HNC; two of each attribute’s levels correspond to key features distinguishing the alternative programs (arm A and B) compared within the trial to provide the clinicians with additional information about the patient’s preferences for the same interventions evaluated in terms of clinical outcomes and cost-effectiveness.

Finally, a conclusion chapter (Chapter VII) summarizes the main thesis’s finding, and discusses any limitations, contributions to research, policy implications and future areas of study.

Since the HETeCo trial inspiring this thesis is a multicentre study based in Italy, some findings in this thesis are inevitably related to the Italian context; particularly, in the model-based economic evaluation (Chapter V), the cost analysis is conducted from the perspective of a major Italian region (i.e. Lombardy). In Chapter VI, for feasibility reasons, recruitment to the survey was limited to a sample of HNC patients in follow-up at the NCI, based in Milan (Lombardy). Conversely, many other results including the synthesis of economic evaluation studies in cancer surveillance (Chapter II), the

overview of HSUVs in HNC (Chapter III), the mapping functions converting cancer-specific HRQoL into EQ-5D-5L utilities (Chapter IV), and the survival/QALY predictions of patients receiving alternative follow-up programmes (Chapter V) are not country-specific and could be generalizable to other contexts. However, even chapters presenting more local findings should be easily adapted within research projects conducted in other contexts; for example, the questionnaire developed for the BWS survey could be modified to reflect the elements characterizing surveillance schemes routinely offered by different hospitals, regions, or countries. A synthesis of the geographical perspectives and data sources adopted in each chapter are reported in Figure 1.2. Overall, the thesis aims to provide an HNC follow-up value framework that can be populated and tested with data collected anywhere.

Figure 1.2 Data sources and study perspectives in thesis's chapters.

Chapter II	<ul style="list-style-type: none"> • Data source(s): published literature • Perspective(s): international
Chapter III	<ul style="list-style-type: none"> • Data source(s): published literature • Perspective(s): international
Chapter IV	<ul style="list-style-type: none"> • Data source(s): multicentre HETeCo trial, published EQ-5D-5L value sets • Perspective(s): national (HRQoL data); international (mapping algorithms)
Chapter V	<ul style="list-style-type: none"> • Data source(s): multicentre HETeCo trial, published and unpublished literature, clinical opinion, regional tariffs • Perspective(s): international (outcomes); regional (costs), or even national, since differences across Regions are minimal
Chapter VI	<ul style="list-style-type: none"> • Data source(s): cross-sectional survey at the NCI (Milan) • Perspective(s): local (or even regional/national, depending on the type of follow-up administered by other hospitals in Italy)

EQ-5D-5L: EuroQol Five-Dimension Five-Level; HETeCo: Health and Economic Outcomes of Two Different Follow Up Strategies in Effectively Cured Advanced Head and Neck Cancer; HRQoL: health-related quality of life.

Table 1.4 Knowledge “gaps” and objectives of the thesis.

Clinical outcomes	<p>Knowledge "gap": The overall survival gains achievable by an intensive follow-up program of radiological investigations compared to a less intensive surveillance based on symptom reporting in HNC.</p>
	<p>Thesis's aim: To obtain model predictions of LYG and QALYs gained by intensive and non-intensive follow-up using preliminary data from the HETeCo trial and published literature (<i>Chapter V</i>).</p>
Health state utility values	<p>Knowledge "gap": The measurement of HSUVs in HNC during the post-treatment phase including recurrent and palliative states; a mapping algorithm predicting HSUVs from two widely used HRQoL tools (i.e. EORTC QLQ-C30 and QLQ-H&N35).</p>
	<p>Thesis's aim: To perform a systematic literature review of HSUVs in HNC (<i>Chapter III</i>); to develop a set of mapping functions to derive HSUVs from HNC-specific HRQoL questionnaires using preliminary data from the HETeCo trial (<i>Chapter IV</i>).</p>
Costs and cost-effectiveness	<p>Knowledge "gap": The economic burden of follow-up in HNC and the cost-effectiveness of alternative programs with different intensity from a healthcare system perspective.</p>
	<p>Thesis's aim: To perform a systematic literature review of economic evaluations of follow-up in any cancer type (<i>Chapter II</i>). To develop an exploratory Markov model comparing the HETeCo trial arms in terms of incremental costs per LYG and QALY gained (<i>Chapter V</i>).</p>
Patient's preferences	<p>Knowledge "gap": The strength of preferences for alternative ways of delivering follow-up in HNC and any heterogeneity in preferences according to individual characteristics.</p>
	<p>Thesis's aim: To elicit patients' preferences for several features describing follow-up using a stated preference approach (i.e. BWS); to calculate the overall utility for a range of hypothetical follow-up programs including the two under evaluation in the HETeCo trial (<i>Chapter VI</i>).</p>

BWS: best-worst scaling; EORTC: European Organization for Research and Treatment of Cancer; HETeCo: Health and Economic Outcomes of Two Different Follow Up Strategies in Effectively Cured Advanced Head and Neck Cancer; HNC: head and neck cancer; HRQoL: health-related quality of life; HSUV: health state utility value; LYG: life year gained; QALY: quality-adjusted life year; QLQ-C30: 30-item Core Quality of Life Questionnaire; QLQ-H&N35: 35-item Head and Neck Cancer Quality of Life Questionnaire.

2 ECONOMIC EVALUATIONS OF FOLLOW-UP STRATEGIES FOR CANCER SURVIVORS: A SYSTEMATIC REVIEW AND QUALITY APPRAISAL OF THE LITERATURE

2.1 Introduction

In recent years, oncological services worldwide have experienced an increasing number of cancer survivors due to advances in diagnostic tools, curative treatments, and prevention campaigns. As already described for HNC (Chapter I), after being treated with curative intent, all cancer patients usually enter a program of post-treatment follow-up which may last for several years [65]. These programs usually involve hospital-based consultations with specialist cancer physicians, but the frequency of visits and the healthcare settings and professionals involved may vary according to geographical contexts and tumour sites [66]. Routine surveillance is primarily aimed at detecting loco-regional recurrences of cancer, metastases or second primaries at the earliest opportunity in order to administer potentially salvage treatments [67] [68]. Secondary aims include addressing treatment-related side effects, managing the rehabilitation process, and providing psychological and social support to patients and caregivers [66] [69].

A variety of recommendations have been provided at national and international level to guide clinicians in the follow-up process of cancer care; however, the most efficient scheme for monitoring patients after the end of primary treatment is still under debate for most malignancies [70]. Firstly, whether repeated investigations can improve long-term clinical outcomes in cancer survivors remains controversial in oncology [67].

Secondly, the clinical benefit of early detection of a cancer relapse strictly depends on the availability of secondary treatments able to extend survival. Thirdly, cancer patients who are monitored intensively after the end of primary treatment may experience either positive (reassurance, relief) or negative (discomfort, anxiety) feelings [67] [71].

Cancer surveillance schemes extended in time also raise economic considerations. The opportunity cost of delivering post-treatment services is significantly high [67] and the long-term sustainability of these programs must be carefully evaluated. The assessment of the effectiveness and the cost-effectiveness of follow-up services in oncology may not be straightforward; any health benefits may become evident long after the interventions, which often involve different medical specialties and consume a variety of healthcare resources.

This chapter presents the results of a systematic literature search undertaken to review published economic evaluations of post-treatment interventions in any cancer population. Given the focus of the thesis, the first aim of this work is to establish whether there are any economic evaluations of follow-up programs in HNC, to identify any literature “gap” that the thesis can fill. The second aim is to retrieve and narratively describe all the available health economic studies on cancer follow-up in general to identify the recent common and conflicting issues around this topic. The third purpose is to appraise the quality of the studies to understand the methodological limitations of the existing literature and test the suitability of a newly developed checklist. Overall, this work informs the development of the cost-effectiveness model presented in Chapter V.

2.2. Methods

2.2.1 Study identification and selection

A systematic literature review was undertaken searching three major electronic databases (i.e. PubMed, EMBASE, and the Cochrane Library). The Preferred Reporting System for Systematic Review and Meta-Analysis (PRISMA) strategies were used to ensure systematic selection of studies and the corresponding checklist completed (Table A2.1) [72]. Keywords were defined according to PICOS (population, intervention/comparator, outcome, study design) elements (Table 2.1). Economic evaluations comparing (two or more) follow-up interventions for adult patients (i.e. ≥ 18 years) after curative treatment for any cancer were included; both health and economic outcomes (i.e. costs) had to be reported. Childhood malignancies were excluded due to different outcomes and costs trajectories and longer time horizon. Other reasons for exclusion were: follow-up strategies for premalignant lesions not yet treated; screening programs for high-risk populations; clinical studies not reporting cost data; cost analyses focusing on only one alternative.

Table 2.1 Search keywords according to PICOS elements.

PICOS	Inclusion Criteria	Keywords
Population	Adults patients after any cancer treatment	cancer OR carcinoma OR tumor OR neoplasm OR neoplasia
Intervention/Comparator	Post-treatment follow-up strategies	'follow up' OR surveillance
Outcomes	Any health outcomes; costs	<i>Not specified</i>
Study design	Economic evaluations	'economic evaluation' OR 'cost effectiveness' OR 'cost utility' OR 'cost benefit' OR 'cost minimization' OR 'cost consequences'

An initial search was made for studies published in the period 2000-2014; an updated search was performed for studies published between 2015 and June 2017. The reference

lists of relevant articles were searched to avoid missing other pertinent studies. Only original full-text articles (i.e. not conference abstracts or editorial comments) were selected; no language restriction was applied to the search. The candidate reviewed the retrieved studies in close consultation with her supervisor and, in case of disagreement, issues were resolved by discussion. A data extraction template was designed to include all relevant information from the studies identified including country, setting, patient population, number of patients, intervention and comparator, type of economic analysis (e.g. modelling vs. clinical study-based), health and economic outcomes, time horizon, cost perspective, currency, conversion and discounting, uncertainty analysis, data sources, study results, and conclusions.

2.2.2 Study quality assessment

The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement was adopted for the critical appraisal of the studies. The CHEERS checklist was recently developed by the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) and jointly endorsed by the BMJ and nine other journals. Previous health economic evaluation guidelines were aggregated into a single standard to help authors report studies or reviewers assess them for publications. The CHEERS tool consists of a 24-item checklist composed by five broad categories: title and abstract (2 items); introduction (1 item); methods (14 items); results (4 items); discussion (3 items) [73]. The 24-item checklist was completed for each study included in the review, indicating “yes” when the criteria were met, “no” when they were unfulfilled and “not applicable” when they were not required for that type of study. Although the CHEERS checklist is not a scoring instrument, papers were divided into three quality categories according to the proportion of items achieved: high ($\geq 75\%$), average (50%-75%) and poor ($< 50\%$) based on other review studies adopting the same tool [74] [75] [76] [77].

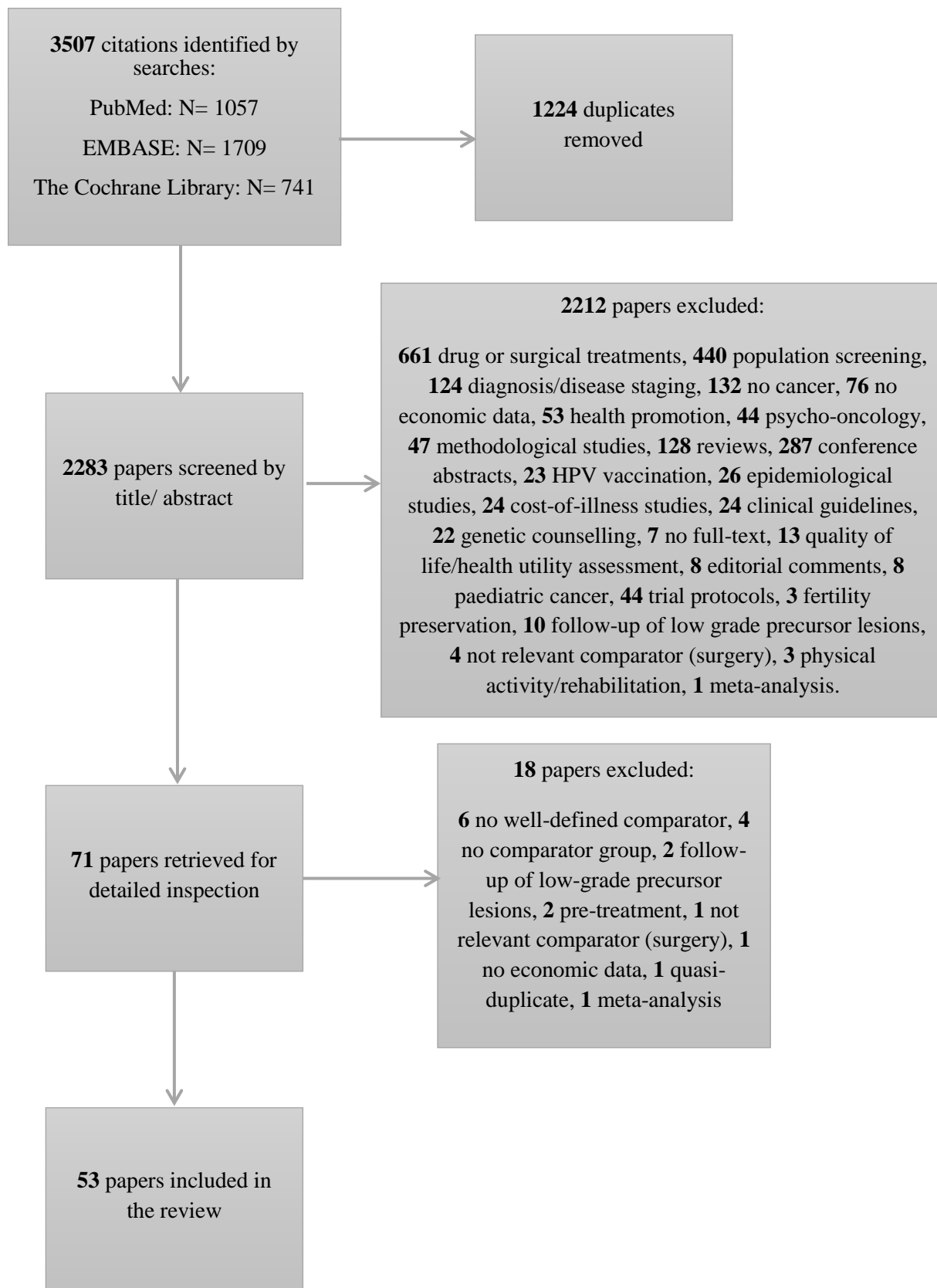
2.3 Results

Figure 2.1, a PRISMA diagram, displays the data for the number of titles initially identified (n=3507), 1224 of them were duplicates. After title and/or abstract screening of the remaining 2283 records, 2212 publications were excluded for a variety of criteria (mainly studies on cancer treatment and cancer prevention/screening in high-risk populations). 71 full-text articles were assessed for eligibility in the study, but only 53 finally met all inclusion criteria and were included in the review.

2.3.1 Study characteristics

Table 2.2. provides a synthesis of the characteristics of the 53 included papers [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104] [105] [106] [107] [108] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [121] [122] [123] [124] [125] [126] [127] [128] [129] [130]. Of these, twelve studies were conducted in the US, ten in the Netherlands, seven in the UK, four each in Canada and Sweden, three in Australia, two each in Italy, France and Spain; the remaining studies, one per country, were carried out in Germany, Finland, Norway, Israel, Iran, Korea, and China.

Figure 2.1 Flow diagram.



The range of tumour sites was quite wide. A quarter of the articles (n=13) were related to breast cancer, followed by cervical (n=9; 17%), colorectal (n=9; 17%), lung (n=5; 9%), and bladder (n=4; 7%); finally, two studies each dealt with oesophageal cancer, prostate cancer, Hodgkin's lymphoma, and melanoma, and one each with anal, pancreatic, ovarian, head and neck, and thyroid cancer. The age range of patients was not systematically recorded, but all were adult subjects according to the review inclusion criteria.

Twenty-one studies explicitly adopted a modelling framework (fourteen Markov models; three discrete event simulations (DES); two decision trees; one semi-Markov model, one decision tree followed by Markov model) involving data extrapolation and/or evidence synthesis. Two papers [79] [113] were based on unspecified modelling. Twenty studies (38%) were analyses of empirical data from clinical trials. Among the remaining articles, four each were (non-randomized) prospective [78] [97] [104] [122] and retrospective studies [83] [98] [101] [117]; one paper [124] was classified as a retrospective study plus (unspecified) modelling and another one [86] as retrospective study plus DES. The number of participants recruited in non-modelling studies (n=28) ranged between 69 [115] [129] and 3223 [127], averaging at 360 patients.

Table 2.2 Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic Outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons of follow-up strategies of different intensity</i>										
Auguste (2014)	UK	PET-CT imaging plus current practice	Current practice alone	Markov model	Healthcare system	5 years	QALYs gained: 4.1096	Δ Cost/ Δ QALY: £1 million	96%	[82]
	Cervical			CUA		Costs: 3.5%				
Bessen (2014)	Australia	(1) Current annual mammography FUP (2) Mixed FUP	Less intensive FUP (2-year mammography)	DES model	Healthcare system	10 years	QALYs gained: 0.002-0.006 (50-69 y old); 0.000-0.003 (70-79 y old)	Δ Cost/ Δ QALY: AU\$21,481-AU\$133,525 (50-69 y old); AU\$40,706-AU\$413,230 (70-79 y old)	83%	[85]
	Breast			CUA		No				
Bessen (2015)	Australia	(1) Current annual mammography FUP (2) Mixed FUP (annual mammography for 5 years and 2 yearly thereafter)	Less intensive FUP (2-year mammography)	Retrospective cohort + DES	Healthcare system	Lifetime	QALY gained (mixed vs. 2 yearly): 0.000-0.010; QALY gained (annual vs. 2 yearly): 0.000-0.006.	ICER (mixed vs. 2 yearly): \$14,676-\$327,898; ICER (annual vs. 2 yearly): \$40,381-dominated.	79%	[86]
	Breast			CUA		No				
Borie (2004)	France	CEA-based standard FUP	Simplified FUP	Markov model	Healthcare system	7 years	QALYs gained: 0.25	Δ Cost/ Δ QALY: €3,114	58%	[87]
	Colorectal			CUA		No				
Damude (2016)	Netherlands	Experimental schedule group (ESG, 1-3 visits first year)	Conventional schedule group (CSG, 4 visits first year)	RCT	Healthcare system	12 months	Recurrence rate: 8.6% in the CSG vs. 8.0% in the ESG (p = 0.89). Less cancer-related stress in ESG than in CSG (p = 0.01).	Cost/patient: €418 (ESG) vs. €762 (CSG) (p=0.01)	71%	[91]
	Melanoma			CCA		No				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic Outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons of follow-up strategies of different intensity (cont.)</i>										
Erenay (2016)	US	US guidelines	Other less intensive guidelines and hypothetical more intensive options	DES	Societal	20 years	US guidelines: MCRC incidence of 1.7%-12.8%; 7.84-8.18 discounted LYs per patient	US guidelines: total cost of \$49,101-\$55,162 per patient; ICER (compared to less intensive guidelines): \$140,000/LYG. More intensive options: ICER (compared to US guidelines): ≤\$63,822/LYG	91%	[96]
	Colorectal			CEA		3%				
Forni (2007)	Italy	Simplified FUP (SCC antigen plus gynecologic examination)	Complete FUP	Prospective cohort	Healthcare system	5 years	Rate of missed recurrences: 2.2%	Cost/patient: €298.5 vs. €3,653.4	52%	[97]
	Cervical			CCA		No				
Guadagnolo (2006)	US	Annual CT for 5 or 10 years	FUP with non-CT modalities only	Markov model	Societal (modified)	Lifetime	QALY gained: 0.0005	ΔCost/ΔQALY: \$9,042,300	87%	[99]
	Hodgkin's lymphoma			CEA; CUA		3.0%				
Hatam (2016)	Iran	Intensive FUP	Standard FUP	Retrospective cohort	Health insurance (or patient)	6 years	Early detection rate (before appearance of clinical signs): 0.137 vs. 0.018 (p<0.001); mortality rate: 17% vs. 22% (p=NS)	Cost/patient: \$311.5 vs. \$112.9 (p<0.001). ICER (incremental cost per early case detected): \$148,196.2	95%	[101]
	Breast			CCA		No				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic Outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons of follow-up strategies of different intensity (cont.)</i>										
Hengge (2007)	Germany	Established FUP practice	Less intensive FUP	Markov model	Healthcare system	5 years	No difference in survival	Cost/QALY: €63,252 vs. €42,433	50%	[103]
	Melanoma			CEA; CUA		No				
Kokko (2005)	Finland	Four strategies combining different visit timing and diagnostic tools		RCT	Healthcare system	5 years	Recurrences detected: 28-35 (range); no difference in SDF and OS	Cost/recurrence detected: €4,166 - €9,149 (range)	57%	[108]
	Breast			CEA		No				
Lu (2012)	Netherlands	5-year FUP with annual mammography	Three less intensive strategies	DES model	Healthcare system	5 years	No difference in recurrences detected	Cost (x1000)/1% increase in recurrences detected: range: €62.1 – €83.1 (current strategy)	65%	[111]
	Breast			CEA		No				
MacAfee (2007)	UK	Intensive FUP	Standard FUP	Model (NS)	Healthcare system	5 years	Additional recurrences detected: 853	ΔCost/Δrecurrence detected: £18,077	78%	[113]
	Colorectal			CEA		Costs: 3.5%				
Oltra (2007)	Spain	FUP with annual mammography (n=63)	A more intensive FUP (n=58)	RCT	Healthcare system	3 years	Recurrences detected: 11 (17.5%; 95% CI: 9.6%-25.3%) vs. 13 (22.4%; 95% CI: 13.4%-31.4%)	Cost/patient: €390 vs. €1,278. Total cost: €24,567 vs. €74,171	33%	[118]
	Breast			CCA		No				
Phippen (2016)	US	Routine surveillance plus one PET/CT scan	Routine surveillance	Decision tree	Medicare	3 years	Recurrence rate: 26% vs. 32%	Cost/patient: \$16,579 vs. \$15,450. ICER: \$20,761 per recurrence prevented	83%	[119]
	Cervical			CEA		No				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic Outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons of follow-up strategies of different intensity (cont.)</i>										
Secco (2002)	Italy	Risk-adapted FUP (n=192)	Minimal surveillance (n=145)	RCT	Healthcare system	5 years	Risk of recurrence: 52.6% vs. 57.2% (p<0.05)	No difference in cost	48%	[121]
	Colorectal			CCA		No				
Tzeng (2013)	US	Four strategies of increasing intensity	No scheduled FUP	Markov model	Medicare	Lifetime	OS (months): 24.6 (no surveillance) vs. 32.8 (surveillance)	ΔCost/LYG: US\$5364 - US\$294,696 (range)	96%	[125]
	Pancreatic			CEA		3.0%				
Verberne (2016)	Netherlands	Intensive FUP with more frequent CEA measurements	Usual FUP (Dutch guidelines)	RCT	Societal	2 years	The intensive FUP was proved to detect cancer recurrences earlier than usual FUP	Medical cost/patient (yearly): €548 vs. €497. Non-medical cost/patient (yearly): €509 vs. €488. ICER: €94/additional 1% of recurrences detected and €607/additional 1% of curable recurrences detected	86%	[127]
	Colorectal			CEA		No				
Wu (2015)	US	Annual surveillance (5 years) plus 3-year surveillance (thereafter)	Perpetual annual surveillance	Markov model	Medicare	38 years	QALY gained (perpetual annual surveillance): 0.01	Cost/patient (yearly): \$5,239 (perpetual annual surveillance) vs. \$2,638 (3-year strategy). ICER (perpetual annual surveillance): \$260,100.	87%	[130]
	Thyroid			CUA		Costs: 3%				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons among different diagnostic tools</i>										
Assoumou (2013)	US	Five strategies using high resolution anoscopy (HRA) and/or cytology		Markov model	Medicare	Lifetime	QALYs gained: 0.0723-0.1061 (range)	Cost/QALY: US\$4,446-US\$17,373 (range)	96%	[80]
	Anal			CUA		3.0%				
Dansk (2016)	Sweden	Hexaminolevulinate hydrochloride-guided blue-light flexible cystoscopy (HAL BLFC) plus white-light flexible cystoscopy (WLFC)	White-light flexible cystoscopy (WLFC) alone	Decision tree + Markov model	Hospital (and other healthcare providers)	5 years	HAL BLFC improved recurrence detection and reduced transurethral resection of the bladder tumours, cystectomies, bed days and operating room time	HAL BLFC resulted in minimal budget impact (+1.6% total cost/5 years, or 189 SEK per patient/year), and translated to cost savings from year 2	96%	[92]
	Bladder			CCA		Costs: 3%				
de Bekker-Grob (2008)	Netherlands	Semi-automated MA plus cystoscopy	Cystoscopy alone	Semi-Markov model	Societal	2 years	Probability (no recurrence after 2 years): 86.3% vs. 86.6%	Cost/patient: €4,104 vs. €3,433	78%	[94]
	Bladder			CCA		No				
Kamat (2011)	US	Five strategies combining cystoscopy, cytology, NMP22 and FISH		Prospective cohort	Medicare	≅4 months	Detection rate: 52%-72% (range)	Cost/recurrence detected: US\$7,692 - US\$26,462 (range)	67%	[104]
	Bladder			CEA		No				
Monteil (2010)	France	CDET imaging with 18-FDG (n=36)	Conventional imaging (n=33)	RCT	Healthcare/Societal	2 years	Recurrences detected: 16 (44.4%) vs. 9 (27.3%) (p=0.14). Time to recurrences detection (months): 12±9.9 vs. 18±11.8	Cost/patient: €1,104.96 vs. €755.47 (p<0.001)	67%	[115]
	Lung			CCA		No				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic Outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons among different diagnostic tools (cont.)</i>										
Ok (2014)	Korea	Seven strategies combining CT, cytology, and urinalysis		Retrospective cohort	Health insurance	6 months	Rate of recurrences detected: 24.5% - 77.6% (range)	Cost/recurrence detected: range: KRW11,049 - KRW100,647	67%	[117]
	Bladder			CEA		No				
Van Loon (2010)	Netherlands	(1) PET-CT scan; (2) chest-CT scan	Conventional chest X-ray scan	Markov model	Healthcare system	5 years	QALYs: (1) 1.30 (2) 1.28; OS (months): (1) 25; (2) 24	Δ Cost/ Δ QALY: (1) €69.086; (2) €264.033	92%	[126]
	Lung			CUA		Costs: 4.0%; effects: 1.5%				
<i>HPV testing versus conventional cytology in cervical cancer follow-up</i>										
Almog (2003)	Israel	HPV testing (n=67)	Conventional cytology (n=63)	Prospective cohort	Health insurance	\cong 53 months	No difference in recurrences detected	Cost/recurrence detected: US\$3,485 vs. US\$3,573	57%	[78]
	Cervical			CEA		No				
Coupé (2007)	Netherlands	Six strategies with adjunct HPV testing	Current cytological FUP	Markov model	Societal	5 years	Reduction in missed cases: 32%-77% (range)	Cost/patient: €178-€351 (range)	70%	[88]
	Cervical			CCA		Costs: 3.0%				
Legood (2012)	UK	(1) Sentinel sites HPV test; (2) Extended HPV test	Conventional cytology FUP	Markov model	Healthcare system	10 years	Case averted: \cong 8	Δ Cost/ Δ Case averted: (1) -£1,120; (2) £6,474	87%	[110]
	Cervical			CEA		Costs: 3.5%				
Melnikow (2010)	US	Twelve strategies combining cytology, colposcopy, and HPV testing		Markov model	Medicare	Lifetime	LYG: 0.001 vs. 0.108; QALY gained: 0.153 – 0.363 (range)	Cost/LYG: US\$4,083 - US\$1,160,000 Cost/QALY: US\$54 - US\$5,246 (range)	100%	[114]
	Cervical			CEA; CUA		3.0%				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons between follow-up programs and 'do nothing' options</i>										
Cromwell (2016)	Canada	Anal cytology	No anal screening	Markov model	Healthcare system	50 years	LYG: 0.004; QALY gained: zero.	ICER: \$20,561/LYG	96%	[90]
	Cervical			CEA; CUA		5%				
Das (2006)	US	Annual low-dose CT screening	No screening	Markov model	Societal (modified)	Lifetime	LYG: 0.64 and 0.16; QALYs gained: 0.58 and 0.14 (smokers and non-smokers)	Δ Cost/ Δ QALY: US\$34,100 (smokers); US\$125,400 (non-smokers)	92%	[93]
	Hodgkin's lymphoma			CEA; CUA		3.0%				
Hassan (2009)	US	1-year endoscopy surveillance	No early endoscopy	Decision tree	Societal	Lifetime	LYG: 2,653	Δ Cost/ Δ LYG: US\$40,313	87%	[100]
	Colorectal			CEA		No				
Kent (2005)	US	Annual CT-based FUP	No annual CT-based FUP	Markov model	Medicare	5 years	QALYs gained: 0.16	Δ Cost/ Δ QALY: US\$47,676	75%	[105]
	Lung			CUA		3.0%				
Tergas (2013)	US	Colposcopy (n=27 low-grade Pap; n=60 high-grade Pap)	No colposcopy (n=23 low-grade Pap; n=18 high-grade Pap)	Retrospective cohort and model (NS)	Medicare	\approx 34 months	Rate of recurrences detected: 8.3% vs. 0.0%	Cost/recurrence detected: US\$7481	71%	[124]
	Cervical			CEA		No				
<i>Comparisons among different organizational aspects of post-treatment follow-up</i>										
Armstrong (2014)	Canada	Mobile-app FUP care	Conventional, in-person FUP care	Model (NS)	Societal	30 days	None	Cost/patient: C\$136 vs. C\$381	61%	[79]
	Breast			CMA		No				
Augestad (2013)	Norway	GP-organized FUP (n=55)	Hospital surgeon-based FUP (n=55)	RCT	Societal	2 years	No difference in QoL and time to recurrences detection	Cost/patient: £8,233 vs. £9,889 ($p<0.001$)	95%	[81]
	Colorectal			CMA		3.0%				
Baena-Canada (2013)	Spain	Primary care FUP (n=60)	Hospital specialist care FUP (n=38)	Retrospective cohort	Healthcare system	5 years	No difference in recurrences detected or QoL	Cost/patient: €112.86 vs. €184.61 ($p=0.0001$)	62%	[83]
	Breast			CCA		No				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic Outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons among different organizational aspects of post-treatment follow-up (cont.)</i>										
Beaver (2009)	UK	Hospital-based FUP (n=183)	Nurse-led telephone FUP (n=186)	RCT	Healthcare/Societal	2 years	No difference in psychological morbidity (STAI), recurrence rate or time to recurrence	NHS FUP cost/patient: £124 vs. £179. Recurrences treatment cost/patient: £143 vs. £182. Transport and productivity cost/patient: £67 vs. £19	95%	[84]
	Breast			CMA		Costs: 3.5%				
Emery (2017)	Australia	Shared care (mixed hospital/GP-based)	Usual care (hospital-based)	RCT	Healthcare system	12 months	No differences in psychological distress, unmet needs, HRQoL, patient's satisfaction, and patient's preferences	Cost savings: \$323 (range: \$91-\$554)	76%	[95]
	Prostate			CMA		No				
Gilbert (2000)	Canada	FP-led FUP	Surgeon-led FUP	Retrospective cohort	Healthcare system	5 years	Recurrences detected: 78 (70.3%) vs. 26 (23.4%)	Cost/recurrence detected: C\$1,105 vs. C\$4,387	48%	[98]
	Lung			CEA		No				
Helgesen (2000)	Sweden	On-demand nurse-led FUP (n=200)	Urologist-led FUP (n=200)	RCT	Healthcare system	3 years	No differences in medical safety and HAD scale	Cost/patient: SEK17,033 vs. SEK19,454	71%	[102]
	Prostate			CCA		No				
Kimman (2011)	Netherlands	Four strategies combining hospital-based or telephone nurse-led FUP with or without EGP		RCT	Societal	12 months	QALYs: 0.776 vs. 0.772	ΔCost/ΔQALY: €235.750 (hospital + EGP vs. telephone + EGP)	91%	[106]
	Breast			CUA		No				
Koinberg (2009)	Sweden	Physician-led FUP (n=131)	On-demand nurse-led FUP (n=133)	RCT	Healthcare system	5 years	No difference in HADS, patient satisfaction, recurrences, and mortality	Cost/patient/year: €630 (95% CI: €557-€1,055) vs. €495(95% CI: €410-€797)	81%	[107]
	Breast			CMA		3.0%				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic Outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons among different organizational aspects of post-treatment follow-up (cont.)</i>										
Lanceley (2017)	UK	Individualized nurse-led FUP	Conventional FUP	RCT	Healthcare system	2 years	Improved QoL (QLQ-C30, p=0.013 and QLQ-OV28, p=0.14) and patient's satisfaction (PSQ III, p=0.002). No difference in HADS (p=0.42).	Cost savings: £700 (p=0.07)	81%	[109]
	Ovarian			CCA		No				
Lyu (2017)	China	FUP on WeChat (WFU)	FUP on telephone (TFU)	RCT	Healthcare system	6 months	Lost to FUP rate: 7.0% vs. 9.8% (p=0.732); patient's satisfaction: 94.3% vs. 80.4% (p=0.034)	Time consumption (per patient): 23.4±6.2 min vs. 42.9±7.1 min (p<0.001). Cost/patient: 90 Yuan vs. 196 Yuan (p=NS).	76%	[112]
	Head and neck			CCA		No				
Moore (2002)	UK	Nurse-led FUP (n=100)	Conventional medical FUP (n=103)	RCT	Health and social care system	12 months	No difference in survival or disease progression	Cost/patient: £696.50 vs. £744.50 (p=0.66)	71%	[116]
	Lung			CCA		No				
Polinder (2009)	Netherlands	Surgeon-led FUP (n=55)	Home-based nurse-led FUP (n=54)	RCT	Societal	12 months	No difference in recurrences detected, patient satisfaction and QoL	Cost/patient: €3,798 vs. €2,592 (p=0.11)	86%	[120]
	Esophageal			CMA		No				
Siddika (2015)	UK	Remote surveillance	(Hypothetical) hospital-based surveillance	Prospective cohort	Healthcare system	5 years	Local recurrence/distant metastasis rate (remote surveillance): 4%/10.3%. Survival rate: 88.8% (comparable with national statistics). Patient's satisfaction: 97% (overall).	Cost savings: £52,593.	43%	[122]
	Colorectal			CMA		No				

Table 2.2 (cont.) Key features of economic evaluation studies included in the review (n=53).

First author (Year)	Country	Intervention	Comparator	Study design	Cost perspective	Time horizon	Health outcomes	Economic outcomes	Quality score	Ref.
	Cancer			Economic evaluation		Discount rate				
<i>Comparisons among different organizational aspects of post-treatment follow-up (cont.)</i>										
Strand (2011)	Sweden	Nurse-led FUP (n=54)	Surgeon-led FUP (n=56)	RCT	Healthcare system	NS	Metastases detected: 8 vs. 7 ($p=0.953$)	Cost/patient: €51 vs. €55 ($p=0.779$)	57%	[123]
	Colorectal			CCA		No				
Verschuur (2009)	Netherlands	Home-based, nurse-led FUP (n=54)	Standard outpatient clinic-based FUP (n=55)	RCT	Healthcare system	13 months	No difference in QoL and patient satisfaction	Cost/patient: €2600 vs. €3800 ($p=0.11$)	71%	[128]
	Esophageal			CCA		No				
Visser (2015)	Netherlands	Group medical consultations	Individual outpatient visits	RCT	Societal	3 months	No difference in patient's satisfaction, efficacy outcomes (i.e. psychological distress and empowerment).	Cost/patient: €53 vs. €35; $p<0.001$.	81%	[129]
	Breast			CMA		No				
<i>Follow-up with education program versus follow-up without educational program</i>										
Coyle (2013)	Canada	Survivorship care plane (SCP)	Current practice (no SCP)	RCT	Healthcare/Societal	2 years	No difference in QALYs	Cost/QALY: SCP is dominated	86%	[89]
	Breast			CUA		5.0%				

CCA: cost-consequences analysis; CDET: Coincidence detection system; CEA: carcinoembryonic antigen; CEA: cost-effectiveness analysis; CMA: cost-minimization analysis; CSG: conventional schedule group; CT: computed tomography; CUA: cost-utility analysis; DES: discrete event simulation; DFS: disease-free survival; EGP: educational group program; ESG: experimental schedule group; FDG: fluorodeoxyglucose; FISH: fluorescence *in situ* hybridization; FP: family physician; FUP: follow-up; GP: general practitioner; HADS: hospital anxiety and depression scale; HPV: human papilloma virus; ICER: incremental cost-effectiveness ratio; KRW: Korean Won; LYG: life year gained; MA: microsatellite-analysis; MCRC: metachronous colorectal cancer; NMP22: BladderCheck®; NS: not specified; NS: not significant; OS: overall survival; PET: positron emission tomography; PSQ: patient satisfaction questionnaire; QALY: quality-adjusted life year; QLQ-C30: 30-item Core Quality of Life Questionnaire; QLQ-OV28: 28-item Ovarian Cancer Quality of Life Questionnaire; QoL: quality of life; RCT: randomized controlled trial; SCC: squamous cell carcinoma; SCP: survivorship care plane; SEK: Swedish crowns; STAI: State-Trait Anxiety Inventory; TFU: telephone follow-up; WFU: web-follow-up.

In Table 2.2, studies were clustered according to different types of follow-up programs compared. More than one third of the studies (19 out of 53) were classified as comparisons of follow-up strategies of different ‘intensity’; among them were included studies addressing follow-up schemes with different timing of controls (e.g. annual or 2-year mammography) or with novel diagnostic tests (e.g. PET-CT imaging) added to routine investigations. Seven studies [80] [92] [94] [104] [115] [117] [126] examined outcomes and costs of a variety of diagnostic tools (e.g. PET vs. CT), while four studies [78] [88] [110] [114] related to cervical cancer compared HPV test versus cytology as potential instruments to detect new lesions. Five studies [90] [93] [100] [105] [124] compared costs and health gains arising from a surveillance program versus a ‘do nothing’ strategy; thirteen studies examined traditional hospital-based follow-up programs in comparison with programs led by other healthcare professionals (i.e. in nine cases [84] [102] [106] [107] [109] [116] [120] [123] [128] the nurse and in four [81] [83] [95] [98] the family physician). Three studies compared a mobile-app follow-up versus traditional in-person consultations [79] [122] or versus a follow-up using telephone [112]; lastly, one study [89] described two programs with (and without) an educational session and another one [129] group medical consultations with individual outpatient visits.

Almost half of the papers (26 out of 53) adopted a limited healthcare perspective where only direct medical costs were considered; one study [116] from UK included the costs borne by social services as well. Nine US-based studies [80] [99] [104] [105] [114] [119] [124] [125] [130] carried out the analysis from the national social insurance program (i.e. Medicare) perspective. Three authors [78] [101] [117] calculated the costs borne by the health insurance companies in their countries (i.e. Israel, Iran, and Korea), while a Swedish study [92] adopted the hospital’s (or other healthcare providers’) perspective. Ten studies embraced a societal perspective estimating broader costs to

society irrespective of the payer and, thus, including also out-of-pocket costs, informal care, and productivity losses. The remaining three articles [84] [89] [115] presented study results according to both healthcare and societal perspectives.

More than half of the included papers (30/53) compared post-treatment surveillance outcomes and costs over a period between 1 and 5 years. Eight studies [79] [104] [106] [112] [116] [117] [120] [129] adopted a shorter timeframe (≤ 1 year), and five studies [85] [87] [96] [101] [110] a longer one (6-20 years). For nine studies [80] [86] [90] [93] [99] [100] [114] [125] [130] the model was run over a lifetime horizon. In one case [123] the time horizon was not specified. The fourteen studies adopting a longer timeframe (i.e. >5 years) were all modelling studies but one that was retrospective [101]; conversely, when primary data collection was performed (i.e. RCTs and cohort studies) the time horizon did not exceed 5 years.

Thirty-four articles did not report any discount rate for future costs and health outcomes. Seven papers [82] [84] [88] [92] [110] [113] [130] applied a discount rate to costs only (i.e. either 3.0% or 3.5%). In eleven articles costs and effects were discounted at the same rate (i.e. 3.0% in nine cases, 5.0% in two). The last paper [126] adopted a 1.5% discount rate for outcomes and 4.0% for costs. Among the studies (44 out of 53) adopting a time horizon ≥ 1 year, discounting was not applied in more than half of the cases (25/44), while seven studies applied a discount factor to costs only; the remaining twelve articles discounted both costs and health outcomes at rates ranging between 3% and 5%.

All types of economic evaluations were represented other than cost-benefit analysis. Fourteen studies were cost-effectiveness analyses where outcomes were expressed in natural units, either intermediate (e.g. number of recurrences detected) or final ones (e.g. survival). Fifteen articles assessed the effects of an intervention in terms of

QALYs, either alone or in combination with physical outcomes (e.g. survival) and were classified as cost-utility analyses. Eight studies were cost-minimization analyses assuming the equivalence of outcomes between interventions and comparators; sixteen were cost-consequences analyses where costs and effects were not combined in a single metric and economic results were usually expressed in terms of cost per patient or total costs.

2.3.2 Study findings

Among the five studies [90] [93] [100] [105] [124] comparing a follow-up program for cancer survivors versus not performing one (i.e. 'do nothing' option), all showed the intervention strategy was a cost-effective option through the calculation of incremental cost per QALY (or life year) gained or recurrence detected.

Some studies (n=7) compared several diagnostic tools to be adopted in post-treatment setting and the results in terms of costs and health outcomes varied considerably. One study [126] evaluating PET combined with CT (PET-CT) showed that it was a cost-effective diagnostic instrument compared to CT alone or X-ray in lung cancer patients. Another study dealing with lung cancer [115] and comparing coincidence detection system imaging (CDET) with 18-fluorodeoxyglucose (FDG) with conventional technique imaging concluded that the two groups were similar in term of recurrences detected and survival, and FDG-18 was costlier. Three authors [94] [104] [117] reported that selected innovative tools (i.e. semi-automated microsatellite-analysis and NMP22 bladder check) were not cost-effective in detecting bladder cancer recurrences. One paper [80] comparing several diagnostic strategies for HIV-infected men with anal cancer found that a combination of high-resolution anoscopy (HRA) and cytology provided the greater benefit at an acceptable cost/QALY gained. Lastly, a modelling

study [92] revealed that the addition of an innovative diagnostic tool to the traditional cystoscopy improves detection rates with a minimal cost increase in bladder cancer.

Four papers aimed at women with cervical intraepithelial neoplasia (NIP) compared follow-up strategies involving diagnostic tools such as HPV testing, cytology, and colposcopy. Three of these studies [78] [88] [110] concluded that HPV testing was cost-effective compared to a conventional cytological approach; on the contrary, one study [114] concluded that HPV testing added limited improvement to survival at higher costs than cytology.

A broad group of papers (n=19) were categorized as economic evaluations of follow-up programs of different ‘intensity’. In seven cases [82] [87] [97] [101] [103] [118] [119], one or more diagnostic tests were added to routine surveillance (e.g. PET-CT imaging plus standard practice); of these, five [82] [87] [97] [103] [118] concluded that a less intensive follow-up program was clinically and economically justified for a variety of malignancies (i.e. breast, cervical and colorectal cancer, and melanoma). One study [101] showed that an intensive follow-up could yield a higher detection of recurrences before the appearance of clinical signs and a lower mortality rate; however, the ICER calculated as incremental cost per case detected was considered too high compared to the threshold recommended by the World Health Organization (i.e. three times Gross Domestic Product per capita). On the contrary, another study [119] stated that the addition of PET-CT scan in cervical cancer follow-up was cost-effective in terms of cost per recurrence prevented. A further group of seven studies [85] [86] [91] [96] [108] [113] [130] compared the same types of diagnostic exams but administered with different timing, either in terms of number of tests per year or of follow-up length (or both); among them, five studies stated that less frequent options were cost-effective [85] [86] [130] or even cost saving [91] [108] compared to a more intensive program. On the contrary, the studies by Macafee [113] and Erenay [96] showed that more intensive

strategies could be cost-effective compared to standard guidelines in colorectal cancer follow-up. In three articles [99] [125] [127] the definition of ‘intensity’ combined both concepts (i.e. increased frequency and additional tests); in two cases [99] [125], less intensive options were preferred since they provided comparable clinical outcomes (i.e. overall survival, number of recurrences detected) at significantly lower costs, whilst one study [127] showed that the incremental cost per curable recurrence detected by more frequent carcinoembryonic antigen (CEA) measurements was reasonably low. One study [121] showed that a risk-adapted follow-up, with the timing of clinical controls and radiological investigations modulated according to the risk of recurrence, was cost saving compared with a common strategy for all. Lastly, the study by Lu [111] performed a comparison of three follow-up strategies where the time in hospital was progressively shortened by a shift of care to the GP; once again, the simplified follow-up showed an acceptable cost-effectiveness profile.

The results from 17 studies dealing with different organizational aspects of post-treatment surveillance supported the current trend of moving towards less structured healthcare programs. However, clinical results were seldom expressed as relevant final outcomes (i.e. LYG or QALYs). Eight [102] [106] [107] [109] [116] [120] [123] [128] out of nine papers comparing hospital-based versus nurse-led follow-up revealed the latter option was less costly without compromising patients’ health or acceptability. A UK-based cost-minimization analysis [84] concluded that, given the equivalence of health outcomes, a nurse-led telephone follow-up compared to a traditional hospital-based one might reduce patient’s travel and productivity costs but did not lead to cost or salary savings in the National Health Service perspective. In a similar way, four studies [81] [83] [95] [98] showed that a general practitioner-led follow-up did not affect survival, quality of life, time to detection of recurrence, or patient’s satisfaction. Three studies [79] [112] [122] concluded that a mobile-app-based follow-up was cost-saving

compared to traditional in-person or telephone-based approaches. The study by Visser [129] showed no differences in efficacy outcomes and limited cost increase for group medical consultations compared to traditional individual visits in breast cancer follow-up.

Finally, the study by Coyle [89] assessed a survivorship care plan (SCP) for women after breast cancer treatment, including an educational session for patients and full follow-up guidelines for general practitioners. SCP was not cost-effective since the control group had better outcomes and lower costs than the SCP group.

2.3.3 Study quality assessment

The quality evaluation of the included studies based on the CHEERS checklist is summarized in Tables 2.3-2.4. Three items (i.e. preference measurement, model choice and model assumption) were applicable only to a limited number of studies; in detail, item 12 was related to cost-utility analyses using QALYs, while items 15-16 concerned modelling studies only. For this reason, the quality judgment was expressed in terms of percentage (instead of absolute number) of checklist criteria met.

Thirty-one papers were categorized as high-quality studies as more than 75% of criteria were fulfilled. For eighteen studies, the quality estimated was of average level (between 50% and 75% of items met), while the remaining four studies met less than 50% of the checklist criteria and were categorized as poor. The average proportion of items achieved was 76% (range: 33% - 100%).

The most commonly missing quality criteria (i.e. in more than half of the studies) were: not accounting for patients' heterogeneity in reporting results (Item 21; missing: 74%); outcomes and costs not discounted and/or justification not given for the adopted (or not

adopted) discount rate (Item 9; missing: 60%); a non-explicative title of the interventions compared and the economic evaluation performed (Item 1; missing: 57%).

On the other hand, the CHEERS items most often reported in the studies were the specification of the target population (Item 4; fulfilled: 98%) and the time horizon of the study (Item 8; fulfilled: 96%), followed by description of health outcomes and the methods and/or sources adopted to measure them (Items 10-11; fulfilled: 94-96%).

Despite the estimated quality being rather high (76%) across studies, several CHEERS points have been weakly considered by selected authors and need further discussion. First, discounting (Item 9) was disregarded by 32 studies (since two studies [101] [120] provided justification for not discounting, thus they were considered to fulfil the criterion); of these, twenty-one used a time horizon longer than 2 years, thus a discount factor was required. Moreover, even in studies adopting a short timeframe (i.e. ≤ 1 year) CHEERS recommends reporting a 0% rate for clarity. Among the few studies (n=21) which carried out some discount technique, most explicitly referred to published guidelines or health jurisdictions; all UK papers in this group [82] [84] [110] [113] adopted the most recent recommendations from NICE for discounting (i.e. 3.5%) but applied the rate to costs only.

A second issue is the specification of model choice and assumptions (Items 15-16). In fact, only four studies [82] [85] [87] [92] gave reasons for the specific type of model used and six modelling studies [79] [93] [99] [110] [113] [130] did not provide a graphical representation of the model structure; moreover, five articles [87] [103] [111] [113] [130] did not properly describe the assumptions underlying the decision-analytic model. Moreover, a table reporting cost (and utility, if required) parameters with probability distributions (Item 18) was not provided by twenty-two studies.

Thirdly, with reference to analytic methods (Item 17) a positive judgment ('yes') has been given if at least one of the following was reported in the article: methods for dealing with skewed, missing, or censored data; approaches to validate or adjust a model (e.g. half-cycle correction); methods for handling heterogeneity or uncertainty. Uncertainty was considered by the great majority (87%) of the studies, either through sensitivity analyses in modelling studies and statistical tests (e.g. t-test, chi-square) or 95% confidence intervals in RCTs; conversely, a limited number of papers fulfilled the other two methodological requirements. Data skewness was taken into account by seven studies [81] [84] [106] [109] [116] [120] [128] which adopted non-parametric bootstrapping techniques (e.g. Mann-Whitney test) for skewed cost data. One study [89] used standard methods for handling censored data; two recent studies [109] [129] specified how to deal with missing data (e.g. using multiple imputation). Among the modelling studies, only one [93] reported adjustments for lead-time bias.

Fourthly, the checklist section about costs (Items 13-14) was only partially fulfilled. In two papers adopting a societal perspective [88] [94] the source of travel, production and other patient costs was not reported. Among the papers which met this requirement, six studies [81] [84] [89] [106] [127] [129] empirically collected non-medical direct costs (i.e. transportation, co-payments, other patient/family expenses) through ad hoc surveys, cost diaries and local sources; in two studies [79] [115] transportation costs were calculated as a function of the distance between home and hospital. Productivity losses were estimated based on average national wages [79] [84] [100] [127] or using the friction cost method [129]. In one study [96], patient's costs were estimated from the literature.

Fifthly, with reference to Item 14 (currency, price date and conversion), twelve studies [83] [91] [94] [95] [97] [111] [112] [117] [118] [121] [122] [123] did not specify the year of reported costs, while three authors [97] [101] [103] failed to report the exchange

rate between currencies. The remaining 38 studies satisfied the criterion. In US-based studies, medical costs (in US dollars) were always adjusted using the medical component of the consumer price index (PCI) for the study year and mainly derived from Medicare reimbursement data; the only exception was a study [96] that estimated costs fully from the literature and updated them using the PCI. In the study from Norway [81] cost elements were converted from Norwegian kroner (NOK) into British pounds at the study year exchange rate. One study from Israel [78] reported healthcare costs in US dollars. Three studies [79] [97] [103] reported results both in local currencies (i.e. Canadian dollars and euros) and US dollars; all the other studies estimated costs in the same currency of the country where the analysis was performed.

Sixthly, twenty studies only reported results as ICERs, either as cost/QALY, cost/LYG or cost/additional treatable recurrence. Fifteen studies reported mean differences in effects and costs between (or among) the alternative interventions, without combining them in an ICER. These two groups of studies combined (n=35) were considered to fulfil the Item 19 criterion. Conversely, the remaining eighteen studies simply indicated mean values for the main categories of estimated costs and health outcomes for each follow-up strategy analysed; this approach is typical of cost-consequences analyses, where costs and consequences are not aggregated into a single measure.

Finally, intervention results in cancer care may vary according to patients' characteristics and disease severity; however, a limited group of papers (14 out of 53) handled population heterogeneity (Item 21) in the decision model. In these studies, results were stratified according to several factors, such as age, treatment, cancer stage, smoking status, presence/absence of metastasis, high/low recurrences risk, and symptomatic/asymptomatic disease.

Table 2.3 Quality assessment of the included studies (n=53) based on the CHEERS checklist (*Items:1-12*).

Ref.	Title	Structured abstract	Rationale/objectives	Target population	Setting	Study Perspective	Comparators	Time horizon	Discount rate	Health outcomes	Effectiveness measurement	Preferences measurement	Items sum	Quality
[78]	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	12/21	Average
[79]	Y	Y	Y	N	N	Y	N	Y	N	Y	N	NA	15/23	Average
[80]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	23/24	High
[81]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	20/21	High
[82]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	23/24	High
[83]	N	Y	Y	Y	N	N	Y	Y	N	Y	Y	NA	13/21	Average
[84]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	20/21	High
[85]	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	20/24	High
[86]	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	20/24	High
[87]	Y	N	N	Y	Y	Y	N	Y	N	Y	Y	Y	14/24	Average
[88]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	16/23	Average
[89]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	19/22	High
[90]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	23/24	High
[91]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	14/21	Average
[92]	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	NA	22/23	High
[93]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	22/24	High
[94]	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	NA	18/23	High
[95]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	16/21	High
[96]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	21/23	High
[97]	N	Y	Y	Y	N	N	N	Y	N	Y	Y	NA	11/21	Average
[98]	N	Y	N	Y	Y	N	Y	Y	N	Y	Y	NA	10/21	Poor
[99]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	21/24	High
[100]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	20/23	High
[101]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	19/21	High
[102]	N	Y	Y	Y	Y	Y	Y	Y	N	N	Y	NA	15/21	Average
[103]	N	Y	Y	Y	Y	Y	N	Y	N	Y	Y	N	13/24	Average
[104]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	14/21	Average
[105]	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	18/24	High

Table 2.3 (cont.) Quality assessment of the included studies (n=53) based on the CHEERS checklist (*Items:1-12*).

Ref.	Title	Structured abstract	Rationale/objectives	Target population	Setting	Study perspective	Comparators	Time horizon	Discount rate	Health outcomes	Effectiveness measurement	Preferences measurement	Items sum	Quality												
[106]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	20/22	High												
[107]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	17/21	High												
[108]	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	12/21	Average												
[109]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	17/21	High												
[110]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	NA	20/23	High												
[111]	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	NA	15/23	Average												
[112]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	16/21	High												
[113]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	18/23	High												
[114]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	24/24	High												
[115]	N	N	N	Y	N	Y	Y	Y	N	Y	Y	NA	13/21	Average												
[116]	N	Y	N	Y	Y	Y	Y	Y	N	Y	Y	NA	15/21	Average												
[117]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	14/21	Average												
[118]	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	NA	7/21	Poor												
[119]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	19/23	High												
[120]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	18/21	High												
[121]	N	Y	Y	Y	N	N	N	Y	N	Y	Y	NA	10/21	Poor												
[122]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	9/21	Poor												
[123]	N	Y	Y	Y	Y	Y	Y	N	N	Y	Y	NA	12/21	Average												
[124]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	15/21	Average												
[125]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA	22/23	High												
[126]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	22/24	High												
[127]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	18/21	High												
[128]	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	15/21	Average												
[129]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NA	17/21	High												
[130]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	21/24	High												
Y	23	43%	47	89%	48	90%	52	98%	48	90%	46	87%	48	90%	51	96%	21	40%	50	94%	51	96%	13	24%		
N	30	57%	6	11%	5	10%	1	2%	5	10%	7	13%	5	10%	2	4%	32	60%	3	6%	2	4%	2	4%		
NA	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	38	72%		

Table 2.4 Quality assessment of the included studies (n=53) based on the CHEERS checklist (*Items:13-24*).

Ref.	Resources/ costs	Currency/ price date	Model choice	Model Assumptions	Analytic methods	Study parameters	Δcosts/ Δoutcomes	Uncertainty	Heterogeneity (subgroups)	Discussion	Funding Source	Conflict interest	Items sum	Quality
[78]	Y	Y	NA	NA	Y	N	N	Y	N	N	N	N	12/21	Average
[79]	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	15/23	Average
[80]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	23/24	High
[81]	Y	Y	NA	NA	Y	Y	Y	Y	N	Y	Y	Y	20/21	High
[82]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	23/24	High
[83]	Y	N	NA	NA	Y	Y	N	Y	N	N	Y	Y	13/21	Average
[84]	Y	Y	NA	NA	Y	Y	Y	Y	N	Y	Y	Y	20/21	High
[85]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	20/24	High
[86]	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	20/24	High
[87]	N	Y	Y	N	Y	N	Y	Y	Y	N	N	N	14/24	Average
[88]	N	Y	Y	Y	Y	Y	N	Y	N	N	N	N	16/23	Average
[89]	Y	Y	NA	NA	Y	Y	Y	Y	N	N	Y	Y	19/22	High
[90]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	23/24	High
[91]	N	N	NA	NA	Y	N	N	Y	N	Y	Y	Y	14/21	Average
[92]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	22/23	High
[93]	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	22/24	High
[94]	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	18/23	High
[95]	Y	N	NA	NA	Y	N	Y	Y	N	Y	Y	Y	16/21	High
[96]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	21/23	High
[97]	Y	N	NA	NA	Y	Y	N	Y	N	N	N	Y	11/21	Average
[98]	N	N	NA	NA	Y	N	N	Y	N	N	Y	N	10/21	Poor
[99]	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y	21/24	High
[100]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	20/23	High
[101]	Y	N	NA	NA	Y	N	Y	Y	Y	Y	Y	Y	19/21	High
[102]	Y	Y	NA	NA	Y	N	Y	Y	Y	N	Y	N	15/21	Average
[103]	Y	N	Y	N	N	N	Y	N	N	N	Y	Y	13/24	Average
[104]	Y	Y	NA	NA	N	N	N	N	N	Y	Y	Y	14/21	Average
[105]	Y	Y	Y	Y	Y	N	Y	Y	N	Y	N	N	18/24	High

Table 2.4 (cont.) Quality assessment of the included studies (n=53) based on the CHEERS checklist (*Items:13-24*).

Ref.	Resources/ costs		Currency/ price date		Model choice		Model Assumptions		Analytic methods		Study Parameters		Δcosts/ Δoutcomes		Uncertainty		Heterogeneity (subgroups)		Discussion		Funding Source		Conflict interest		Items sum	Quality
[106]	Y		Y		NA		NA		Y		Y		Y		Y		N		Y		Y		Y		20/22	High
[107]	Y		Y		NA		NA		Y		Y		Y		Y		N		Y		N		N		17/21	High
[108]	Y		Y		NA		NA		N		Y		N		N		N		N		Y		N		12/21	Average
[109]	Y		Y		NA		NA		Y		N		Y		Y		N		Y		Y		Y		17/21	High
[110]	Y		Y		N		Y		Y		Y		Y		Y		N		Y		Y		Y		20/23	High
[111]	Y		N		Y		Y		Y		N		N		Y		N		N		N		N		15/23	Average
[112]	Y		N		NA		NA		Y		N		Y		Y		N		Y		Y		Y		16/21	High
[113]	Y		Y		N		N		Y		Y		Y		Y		Y		N		N		Y		18/23	High
[114]	Y		Y		Y		Y		Y		Y		Y		Y		Y		Y		Y		Y		24/24	High
[115]	Y		Y		NA		NA		Y		Y		N		Y		N		Y		N		Y		13/21	Average
[116]	Y		Y		NA		NA		Y		N		N		Y		N		Y		Y		Y		15/21	Average
[117]	Y		N		NA		NA		N		N		N		Y		N		Y		N		Y		14/21	Average
[118]	N		N		NA		NA		N		N		N		N		N		N		N		N		7/21	Poor
[119]	Y		Y		Y		Y		Y		Y		Y		Y		N		Y		N		Y		19/23	High
[120]	Y		Y		NA		NA		Y		Y		N		Y		N		Y		N		Y		18/21	High
[121]	N		N		NA		NA		Y		N		N		Y		Y		Y		N		N		10/21	Poor
[122]	N		N		NA		NA		N		N		N		N		N		N		N		N		9/21	Poor
[123]	Y		N		NA		NA		Y		N		N		Y		N		Y		N		N		12/21	Average
[124]	Y		Y		NA		NA		N		N		Y		N		Y		Y		N		Y		15/21	Average
[125]	Y		Y		Y		Y		Y		Y		Y		Y		N		Y		Y		Y		22/23	High
[126]	Y		Y		Y		Y		Y		Y		Y		Y		Y		Y		N		Y		22/24	High
[127]	Y		Y		NA		NA		Y		Y		Y		Y		N		N		Y		Y		18/21	High
[128]	Y		Y		NA		NA		Y		N		N		Y		N		Y		Y		Y		15/21	Average
[129]	Y		Y		NA		NA		Y		Y		Y		Y		N		N		Y		Y		17/21	High
[130]	Y		Y		N		N		Y		Y		Y		Y		N		Y		Y		Y		21/24	High
Y	43	81%	38	72%	18	34%	19	36%	46	87%	31	58%	35	66%	47	89%	14	26%	37	70%	33	62%	39	74%		
N	10	19%	15	28%	6	11%	5	9%	7	13%	22	42%	18	34%	6	11%	39	74%	16	30%	20	38%	14	26%		
NA	0	0%	0	0%	29	55%	29	55%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%		

2.4 Discussion

2.4.1 Summary of evidence

This is the first-time health and economic outcomes from post-treatment follow-up interventions have been compared across countries and for all types of cancer. There exist around twenty-five reviews published in the last few years dealing with cancer follow-up care, but they are all limited to clinical studies only, specific cancer populations or types of interventions (e.g. primary versus secondary care). A dated paper by Edelman [131] reviewed surveillance strategies and assessed follow-up costs for the most common malignancies based on the studies retrieved. A study by Hex [132] aimed at reviewing the cost-effectiveness of follow-up care in paediatric tumours and highlighted a trend towards risk-based personalized approaches for long-term childhood cancer survivors. The target population addressed by Hex was exactly complementary to that of the current study, as this search was focused on adult cancer patients only. Moreover, none of these reviews addressed the quality of the included studies.

This review provides insights into the clinical and economic value of a variety of post-treatment follow-up programs across many types of malignancies. Due to recent improvements in cancer therapies and survival rates, the number of patients requiring post-treatment services is rapidly increasing and posing a substantial burden on healthcare systems. The fifty-three studies included in the review represent the best economic evidence available around cancer follow-up. From study findings, a general tendency emerged towards less intensive options in terms of frequency of visits and/or length of program, risk-adapted follow-up according to age or tumour stage, and service delivery in primary care or through mobile-app technologies replacing traditional hospital-based investigations. In most studies, these simplified follow-up schemes were to be preferred to the more intensive ones according to their favourable cost-

effectiveness profile. However, study results, even for the same type of follow-up, were contradictory and estimates varied considerably by study setting and cancer type. In some cases, indeed, the addition of sophisticated diagnostic techniques, such as PET-CT and hexaminolevulinate hydrochloride-guided blue-light flexible cystoscopy (HAL BLFC) were recommended as cost-effective options. Moreover, most studies reported the equivalence (or a non-significant difference) of health outcomes between traditional and novel options for post-treatment surveillance, with a cost saving when less intensive or non-hospital-based programs were implemented; indeed, none of them led to significant improvements in health outcomes such as the number of recurrences detected or overall patient's survival.

The average quality score (76%) of the studies retrieved was good, with most of them (n=31; 58%) performing very well in reporting economic evaluations. Among high-quality studies, cost-utility analyses, Markov model- and UK/US-based studies were mainly identified; these studies, indeed, generally measure outcomes in terms of QALYs and cost/QALY is the ratio adopted by more recent studies to determine the cost-effectiveness of healthcare programs. Markov model is an appropriate instrument to conduct economic evaluations in chronic diseases where the occurrence of the events (e.g. cancer recurrence) is uncertain and these may happen more than once [133]. Compared to other types of modelling (e.g. discrete event simulation) equally valid for the purpose, the Markov model requires less clinical information but a validation of the underlying assumptions. Moreover, studies conducted in UK and US are more likely to adhere to recommendations from NICE or other HTA agencies. Finally, the average quality score (i.e. 81%) of the studies published since 2015 (n=14) is considerably higher than the mean score (i.e. 61%) across the articles dated between 2000 and 2005. Thus, it is reasonable to assume that more recent economic evaluations tend to better adhere to published recommendations in this field.

2.4.2 Critical issues

A number of issues characterizing long-term cancer survivorship should be carefully evaluated by the health economics literature in this field. First, the setting (e.g. GP-versus hospital-based) where follow-up care is conducted can greatly affect patient's quality of life and private costs (i.e. travel expenses and productivity losses), especially in peripheral areas with long distances to travel to hospitals. Secondly, little is known about potential damages (and related costs) of follow-up, in terms of patient's dissatisfaction, long-term toxicity, and false positive results; thus, economic evaluations including also these cost categories are encouraged. Thirdly, the topic of heterogeneity in cancer patients affected by the same malignancy is still unexplored in the literature. Post-treatment surveillance programs, indeed, may yield different survival gains according to age, cancer stage and comorbidities; thus, economic evaluations should routinely report differences in cost-effectiveness results for relevant subgroups of patients. A further weakness observed in the reviewed articles is related to the choice of health outcomes; most studies, indeed, evaluate follow-up interventions in terms of intermediate outcomes (e.g. number of recurrences detected). However, the 'value' of an early diagnosis of cancer relapse is closely linked to the availability and effectiveness of secondary treatments able to extend survival. Future economic studies are encouraged to adopt longer timeframes in order to catch the full health effects and cost paths arising from different surveillance options and potential curative treatments administered in case of recurrence. A longer time horizon is also necessary to capture long-term side effects related to intensive radiological examinations. As medical advances have improved post-treatment prognosis, the health issues experienced by cancer survivors tend to be more episodic and to occur over a longer timeframe. Thus, model-based economic evaluations are increasingly required to extend clinical trial results over patient's lifetime.

2.4.3 Strengths and limitations

This systematic review presents a number of limitations. First, the databases searched were limited to the most common ones (i.e. PubMed, Medline, and EMBASE). This may result in selection bias that was reduced by searching the references of each selected article manually. Secondly, as in all searches of the published literature, there may be publication bias and unpublished studies could affect the review findings; however, this is particularly common in studies funded by private companies of which there are few in this review as it deals with non-pharmaceutical interventions. Thirdly, most of the included studies were conducted in the developed world, thus likely representing the most healthy and affluent group of cancer survivors [66]. For these reasons, the extension of these results to other settings needs to be done cautiously.

Moreover, comparison of economic outcomes is complicated due to the high variability in time periods, currencies, and health systems involved. Although it is reasonable to assume that the relative price of some common diagnostic technologies (e.g. PET-CT scan) might be similar across many developed countries [74], other cost categories (e.g. nurse salaries, consultation fees) or reimbursement policies may differ a lot. Even in studies assessing standard health outcomes such as survival or QALYs, comparison is hard to perform due to a wide range of cancers addressed in the review and, even across studies related to the same malignancy, heterogeneous follow-up interventions and patient populations by age or cancer stage. Moreover, the concentration on economic evaluations as inclusion criterion may have excluded other important clinical studies in cancer follow-up research.

A further limitation is the use of the CHEERS checklist as a measure of quality in economic evaluations. This checklist gives an indication of how much the published studies adhere to reporting criteria but does not state their relative importance. For

example, reporting discounting may be more relevant than funding source and a simple addition of the criteria met may result in a misleading assessment of quality. However, as the CHEERS checklist does not provide any weights to be applied to quality criteria, summing the number of items achieved with a qualitative discussion in text was assumed an appropriate methodology to differentiate the quality of studies.

In general, the use of checklists to evaluate the quality of economic evaluation studies can be viewed as overly simplistic. For example, the standard checklists emphasize how well the study is reported rather than whether it can inform good policy decisions, since they evaluate the publication itself and not its implications for clinical practice. Moreover, not reporting an element in the article does not necessarily mean that the authors in the analysis have not addressed that aspect. In addition, most checklist criteria require that the reasons behind the choice of a given item (e.g. time horizon) are specified, but the authors often disregard this aspect and just indicate the parameter value; thus, the simple addition of the items reported by each study does not inform about the appropriateness of a method. Especially in model-based economic evaluations, different assumptions around the study parameters may significantly alter the cost-effectiveness results with important consequences for evidence-based medical and policy decisions. A well-reported study will not necessarily be fit for purpose but at least it may be easier to determine whether it is fit for purpose compared with a less well-reported study.

In spite of these limitations and compared to previous guidelines (e.g. Drummond [134]), the CHEERS checklist appears more comprehensive and suitable for model-based studies which are becoming increasingly important, partly due to the financial and logistic constraints on performing primary data collections (e.g. RCTs) but also because of the well-known limitations of trial-based evaluations for informing decision-making. The checklist, indeed, asks authors to specify more details about model assumptions and

analytic methods (i.e. skewed, missing, censored, and pooled data) that were not provided in older guidelines. Moreover, a greater emphasis is given to the need of characterizing heterogeneity in reporting study results, in line with the recent trend towards a more personalized medicine according to patient's characteristics.

2.5 Conclusions

Health economic analyses are increasingly used to inform policy-makers about the efficient allocation of limited healthcare resources. Economic evaluations in cancer care have been mainly applied to drug therapies, while less evidence is available for other types of interventions. In recent years, a debate arose among oncologists around how to design post-treatment programs for cancer survivors. This chapter summarizes the current body of knowledge regarding economic evaluations in cancer follow-up and may help clinicians and policy-makers interpret health economic results according to study quality and update post-treatment surveillance schemes based on a sound scientific evidence.

The quality of the studies retrieved from the literature is generally high. However, many of them report cost-effectiveness results as intermediate outcomes (e.g. cost per recurrences detected), or without combining costs and outcomes in a single measure (i.e. as cost-consequence analyses); these study characteristics pose an issue of comparability with economic evaluations in other disease areas and with commonly accepted thresholds, which are expressed as incremental cost per QALY gained. High-quality studies in the review often use a modelling approach, confirming that model is increasingly an unavoidable instrument in health economic analyses to project the costs and the effects of a healthcare intervention beyond the usual limited length of clinical trial. Although judging the quality of scientific work is inevitably controversial, the CHEERS checklist appears suitably up-to-date and comprehensive to facilitate this task.

This review indicates that less intensive, nurse-led and primary-care based follow-up schemes are frequently clinically equivalent and economically justified in oncology. However, no studies reported any efficacy gains in favour of these simplified programs and results vary considerably across studies. HPV testing appears a cost-effective alternative to traditional cytological approach for cervical cancer patients. There is also evidence of increasing interest in delivering post-treatment services using technology (e.g. mobile apps), although not yet sufficient to state whether remote cancer follow-up may replace traditional face-to-face surveillance.

Most of the reviewed literature focused on widely spread neoplasms such as breast, colorectal and cervical cancer. In the context of this thesis, the search identified the first economic evaluation study [112] comparing two follow-up programs in HNC. However, the study presents several weaknesses, including the use of patient-reported outcomes of limited clinical significance (e.g. patient's satisfaction and loss to follow-up), and a short time horizon (i.e. 6 months); thus, a model-based economic evaluation comparing follow-up strategies of different intensity over the patient's lifetime is still lacking in HNC, and the rationale for a valuable contribution from this thesis is made evident. Overall, high-quality scientific evidence on the cost-effectiveness of surveillance, and especially for less common malignancies, continues to be urgently required.

3 A SYSTEMATIC LITERATURE REVIEW OF HEALTH STATE UTILITY VALUES IN HEAD AND NECK CANCER

3.1 Introduction

Cost-utility models are increasingly used to establish whether the cost of a new treatment is justified in terms of health gains. This approach usually adopts the QALY as a measure of health effectiveness. As already described in Chapter I, the QALY corresponds to *the time spent in a series of quality-weighted health states*, where the weights represent the desirability of living in that state [36] [135]. The basic idea is that individuals move through health states over time and that each health state has a preference weight attached to it [36], also known as a HSUV. Thus, the HSUV can be interpreted as the strength of preference for a given health state on a cardinal scale anchored at 0 ('death') and 1 ('full health'), with some instruments also allowing for negative values representing states worse than death [136]. Therefore, QALYs are obtained by summing-up the products of the time spent in each health state and its corresponding preference-based value [137].

HSUVs can be estimated in a variety of ways including direct methods, multi-attribute utility instruments (MAUIs), mapping functions and expert opinion. The most common ways of eliciting HSUVs directly are gambling with respect to a hypothetical treatment that may result in perfect health or death (standard gamble, SG) or trading-off part of future life for a shorter time in perfect health (time trade-off, TTO) [138]. A further, simpler option is to use a VAS, also known as rating scale, which provides an immediate valuation of the current (or a hypothetical) health state on a graduated scale, usually ranging between 0 and 100. This technique is generally considered to be

methodologically inferior to choice-tasks such as SG and TTO, which incorporate some extra information about the individual risk attitude [137]; VAS scores, indeed, are elicited in a choice-less context, and thus do not require respondents to make trade-offs within their utility function [139]. Moreover, rating scales are well-known to present measurement biases such as context bias, spacing-out bias, and end-aversion bias [137] [140]. Context bias occurs when several VAS questions are presented at the same time and the ratings for some tasks are affected by the context of the questionnaire, while spacing-out bias refers to respondents spacing out their scores to fill the entire range presented. The end-aversion bias occurs instead when respondents avoid using the extreme ends of the scale, thus compromising the integrity of the measurement [141]. Additionally, there is now consensus that HRQoL is a multi-dimensional concept, which includes domains related to physical, mental, emotional, and social functioning that are difficult to measure on a single scale [142].

Direct measurement of health utility through SG or TTO can be complicated and time-consuming and lead to incomparable results across the studies due to arbitrary health state descriptions (also called ‘vignettes’) [143] [144]. Consequently, in recent years, HSUVs have been increasingly estimated indirectly using MAUIs. These tools are formed of a generic HRQoL questionnaire and an accompanying formula or set of weights (or “tariffs”) elicited from a sample of the general population for converting responses into HSUVs; thus, the utility measure can be considered as a preference-based evaluation of a given health state described by the dimensions of the tool [145] [146]. NICE and the European Network for HTA (EUnetHTA) recommend the EQ-5D (<https://euroqol.org>) [147] [148]. Accordingly, the TTO with a 10-year time horizon is the most frequently used approach among the direct techniques, because of greater comparability with the method used to develop the EQ-5D scoring algorithm [149]. The other generic MAUIs mostly adopted in the literature [145] are the Health Utility Index

(HUI mark 2, HUI2 or mark 3, HUI3) [150], the Short Form-6-dimension (SF-6D) questionnaire derived from the 36-item Short Form Survey (SF-36) (www.sheffield.ac.uk), the 15D (www.15d-instrument.net), the Quality of Wellbeing (QWB) index [151] and the Assessment of Quality of Life (AQoL) instruments [152].

In many situations, clinical studies neither administer preference-based MAUIs nor elicit HSUVs directly but collect instead disease-specific HRQoL data or other clinical measures that are not associated with a preference-based scoring system; thus, QALY calculation from these studies is not possible. Therefore, “mapping” or “cross-walking” has been developed to predict HSUVs from non-preference-based scores, provided that a statistical relationship can be established between the two instruments and, sometimes, allowing for the mediating effect of demographic and clinical characteristics [153]. In most cases, however, it is still preferable to collect HSUVs directly or use MAUIs, and mapping should be viewed as a “second-best” solution [154].

This chapter focuses on HSUVs in HNC. Patients with HNC often undergo several rounds of treatment during which they experience acute toxicity and other side effects, such as loss of verbal abilities, difficulties in swallowing, and considerable pain [155]. This HRQoL impairment may continue long after treatment through persistent functional deficits, physical disfigurement, psychological distress, and recurrent disease. There is an extensive HRQoL literature in HNC, although mainly comprised of disease-specific, non-preference-based data unsuitable for cost-utility comparisons. Due to the paucity of HSUVs for some health states in HNC, some previous cost-effectiveness analyses [156] [157] relied on values calculated for other cancers (such as breast or lung) to populate their models with utility parameters. A systematic review published in 2006 [158] identified eight studies providing utility values in HNC elicited through VAS, TTO or SG. The current work extends the collection of utility values related to this medical condition by systematically reviewing the studies published to date. This

review considers for inclusion studies of any design in which utility values in HNC were:

- directly elicited using standard techniques such as TTO or SG either in patient-based studies or in the general population;
- calculated indirectly from patient's responses to generic MAUIs (e.g. EQ-5D) through a set of tool- and country-specific preference weights;
- predicted from non-preference based HRQoL instruments using mapping algorithms.

The PRISMA statement [72] is not entirely applicable to systematic reviews of HSUVs [159], since the standard Population, Intervention, Comparator, and Outcome (PICO) elements do not provide a useful framework for identifying utility values for health states that are not necessarily attached to a given intervention [160]. Thus, this study followed the recommendations provided by Papaioannou et al. [160]. The ultimate objective is to generate a database of HSUVs that might be useful to populate future cost-utility models of interventions in HNC, including the one comparing alternative follow-up programs based on the HETeCo trial, which, indeed, stops collecting EQ-5D after any patient's relapse (Chapter V). In addition, the included studies were critically appraised by highlighting a few elements that should be considered when selecting utility parameters for modelling.

3.2 Methods

A systematic literature search was carried out of the PubMed, EMBASE and Cochrane Library databases for studies published from 2000 until the end of 2016 using a range of free-text terms in title/abstract (Figure 3.1). Since Medical Subject Headings (MeSH) terms provide little coverage of HSUVs [159] [160], a few relevant free-text terms were identified by referring to the published recommendations [160] and recent analogous

systematic reviews [159] [161] [162]. Tool- (e.g. EQ-5D) and method-specific (e.g. SG) terms were combined with vocabulary related to HNC including the most frequent cancer sites; in using free-text terms, it was considered that some instruments may be referred to or spelled in different ways. The VAS term was not explicitly included among the keywords, due to the above-mentioned limitations in using this tool for measuring utility. Other search strings were used to identify cost-effectiveness and cost-utility studies using HSUVs to calculate QALYs. A direct search of utility weights in the Tufts Cost-Effectiveness Analysis (CEA) Registry [163] and the University Sheffield School of Health and Related Research Health Utilities Database was done (ScHARRHUD) [164]. An additional search was carried out of the Health Economics Research Centre (HERC) database [165] [166] to retrieve mapping studies deriving utility values from non-preference-based instruments in HNC. The relevant databases were selected based on previous recommendations [160] and systematic reviews on the topic [167]. Web searches of grey literature were not performed to avoid obtaining contents which are frequently subject to changes and cannot be identified in a systematic manner.

All search results were extracted in an Excel spreadsheet and duplicates removed. Titles and abstracts were screened by two independent reviewers (i.e. the candidate and the supervisor) and records excluded if not meeting the inclusion criteria; full-text papers were retrieved in case of doubtful results. Articles estimating HNC utility values using established methods were included; studies using the VAS instrument were not considered for inclusion, unless alongside other valuation techniques. This choice is consistent with the suggestion that VAS should be used as an introductory task but not as a definitive method to elicit utility values alone [168]. The included studies had to be published as full-text with no time or language restrictions; conference abstracts, editorials, and reviews were not suitable for inclusion. Studies were excluded if they did

not report original utility values in HNC; however, the bibliography of studies referring to secondary sources for HSUVs was checked to avoid missing any relevant publications. The reviewers resolved any disagreements by discussion until consensus was reached.

Figure 3.1 Free-text terms for electronic database searching.

1. head neck
2. oropharyn*
3. hypopharyn*
4. laryn*
5. oral cavity
6. cancer or carcinoma
7. (1 or 2 or 3 or 4 or 5) and 6
8. health utilit*
9. EQ 5D or EQ5D or EuroQol or Euro Qol
10. SF 6D or SF6D or short form 6D
11. 15D
12. QWB
13. AQoL or assessment quality life
14. HUI or health utility index
15. standard gamble
16. time trade off or time tradeoff
17. cost utility
18. cost effectiveness
19. economic evaluation
20. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19
21. 7 and 20

The characteristics of the included studies were extracted by the candidate using a form developed following previous studies [159] [160] [161], and subsequently crosschecked by the supervisor. Information collected included: study country, study design, sample size, valuation technique, administration method, cancer subsite addressed, and clinical and demographic characteristics of respondents. For each HSUV, the number of respondents, the point estimate (i.e. mean or median) and its measure of variance (e.g. standard deviation) were recorded; the same information was collected for each study subgroup (or time point) whenever applicable.

Although there are no agreed reporting standards for HSUVs studies, the methodological quality of each included study was evaluated through a set of generic criteria as reported by the guidelines from Papaioannou et al [160]. Thereafter, one point was awarded to each of the following criteria: (1) sample size ≥ 100 ; (2) description of respondent selection and recruitment; (3) description of inclusion/exclusion criteria; (4) response rate $\geq 60\%$ [169]; (5) reporting of the amount and reasons of loss to follow-up (only for longitudinal studies); (6) reporting of missing data pattern and methods to deal with it; (7) appropriateness of measure (based on the authors' judgment). Lastly, the scores were summed for each article to yield an overall quality score, ranging from 0 to 7 where higher scores indicated higher quality [170]. Any other problems arising from the studies (criterion 8) were narratively discussed. Additionally, ISPOR recently published a set of recommendations for mapping studies [153] that were used to evaluate the quality of mapping studies retrieved by the systematic search.

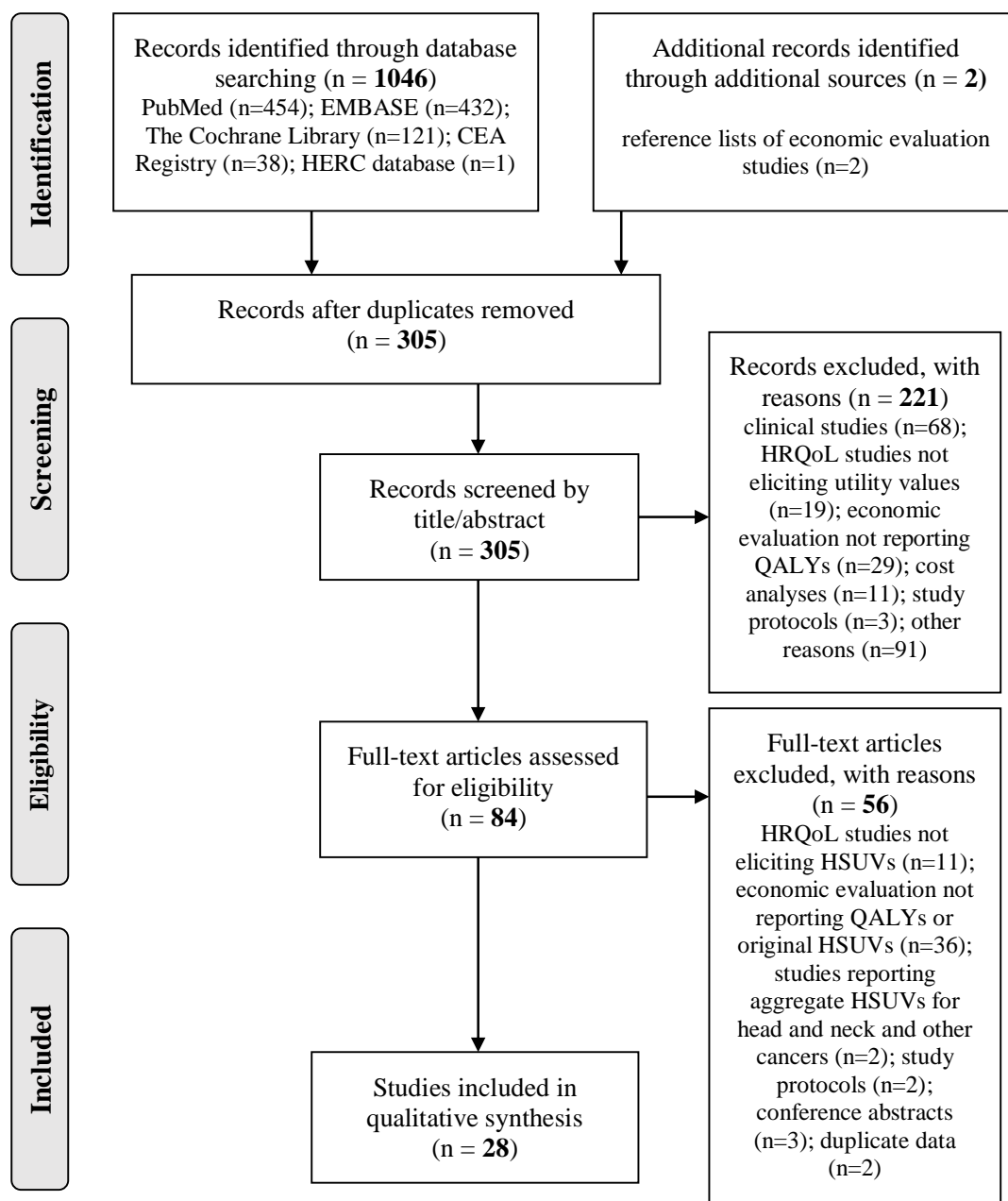
3.3 Results

3.3.1 Study selection

The PRISMA diagram [72] for this literature search is presented in Figure 3.2. In total, the search strategy identified 1048 articles: 1046 were retrieved by searching the online databases (PubMed; EMBASE; The Cochrane Library; CEA Registry; HERC database), and two by manually searching the bibliography of model-based economic evaluations retrieved from the online search. No articles were obtained from the SchARRHUD database. After removing 743 duplicates, 305 records were scanned for title/abstract and 221 were excluded in this first phase for a variety of reasons reported in the chart. Subsequently, 84 full-text articles were retrieved and a further 56 records

were excluded for not complying with the inclusion/exclusion criteria. Accordingly, 28 studies were definitively included in the review.

Figure 3.2 PRISMA flow chart.



3.3.2 Study characteristics

The 28 journal articles included in the review are categorized into three groups: studies using direct elicitation methods (n=10), studies administering MAUIs (n=13) and studies deriving HSUVs using mapping (n=3); two studies [171] [172] adopted both direct methods and MAUIs. The characteristics of the 25 studies using direct and indirect techniques (i.e. MAUIs) are listed in Table 3.1, while the three mapping studies are separately described in Table 3.2.

Studies using direct or indirect methods

Among the studies using direct elicitation techniques, SG alone was adopted in two cases [155] [173] and TTO alone in five [174] [175] [176] [177] [178]. In four studies [172] [179] [180] [181], more than one direct methodology (i.e. SG, TTO, VAS) was adopted to derive utility values. The study by Noel et al. [172] compared these direct techniques with MAUIs (i.e. EQ-5D, HUI3), while a further study [171] used both TTO and EQ-5D instruments.

In studies administering MAUIs, EQ-5D was the most common (n=11); five of these studies [171] [172] [182] [183] [184] did not report which scoring algorithm was used, two studies [185] [186] explicitly adopted the UK algorithm, another two [187] [188] adopted the US one, one study [189] used the Dutch tariff and another one [190] the Belgian one. Moreover, nine of the studies using EQ-5D [171] [182] [183] [185] [186] [187] [188] [189] [190] explicitly referred to the 3-level version (EQ-5D-3L) and one [172] to the newer 5-level one (EQ-5D-5L); one study [184] did not specify the instrument's version adopted. Additional generic, preference-based HRQoL tools retrieved by this search were 15D (n=2), HUI3 (n=2) and SF-6D (n=1); no studies used the QWB scale or the AQL-8D utility instrument.

Table 3.1 Studies estimating HSUVs in HNC using direct or indirect methods (n=25).

Author (year)	Country	Study design	Cancer subsite(s)	Valuation method	Mode of administration	Sample size	Response rate	Participants	(Mean) age; % male
Aro (2016) [191]	Finland	Longitudinal	All	15D	Self-completion (by post)	214	72%	Patients	63.0; 66%
Conway (2012) [173]	Australia	Cross-sectional	Oropharynx	SG	1-hour group session	99	84%	Healthy subjects	43.0; 54%
de Almeida (2014) [179]	US	Cross-sectional	Oropharynx	SG; VAS	Face-to-face interview	59	NA	Healthy subjects (n=50) and experts (n=9)	Healthy subjects: 34.8; 42%. Experts: 45.3; 89%
del Barco Morillo (2016) [182]	Spain	Longitudinal	All	EQ-5D-3L	NA	40	NA	Patients	61 (Median); 87%
Govers (2016) [189]	Netherlands	Cross-sectional	Oral cavity	EQ-5D-3L	Self-completion (by post)	181	62%	Patients	64.4; ≥50%
Hamilton (2016) [174]	UK	Cross-sectional	Larynx	TTO	Face-to-face interview	114	NA	Healthy subjects (n=51) and COPD patients (n=63)	67.3; 49%
Higgins (2011) [192]	Canada	Cross-sectional	Larynx	HUI3	Self-completion	30	NA	Patients	NA
Hollenbeak (2001) [175]	US	Cross-sectional	All	TTO	NA	8	80%	Patients	NA
Jalukar (1998) [176]	US	Cross-sectional	All	TTO	Self-completion (on site for patients; by email for healthcare professionals)	185	Patients: 78%; healthcare professionals: 42%; students: NA	Patients (n=49); healthcare professionals (n=50); students (n=86)	Patients: 57.2; 71%. Healthcare professionals: 40.1; 40%. Students: NA
Kent (2015) [193]	US	Cross-sectional	Oral cavity and pharynx	SF-6D/VR-6D	Mail or telephone	580	62%	Patients	67.7; 60%
Llewellyn-Thomas (1993) [180]	Canada	Longitudinal	Larynx	TTO/VAS	Interview	66	NA	Patients	NA; 86%

Table 3.1 (cont.) Studies estimating HSUVs in HNC using direct or indirect methods (n=25).

Author (year)	Country	Study design	Cancer subsite(s)	Valuation method	Mode of administration	Sample size	Response rate	Participants	(Mean) age; % male
Loimu (2015) [194]	Finland	Longitudinal	Pharynx, larynx, nasal cavity	15D	Self-completion: on site (first assessment); by post (afterwards)	64	76%	Patients	61.6; 75%
Marcellusi (2015) [171]	Italy	Cross-sectional	All	TTO; EQ-5D-3L	Computer-guided	79	NA	Patients	65.0; 78.5%
Noel (2015) [172]	Canada	Cross-sectional	All	SG; TTO; VAS; EQ-5D-5L; HUI3	Face-to-face interview	100	79%	Patients	61.0; 75%
Ouattassi (2016) [183]	Morocco	Cross-sectional	All	EQ-5D-3L	Self-completion	120	NA	Patients	57.0; 60%
Parrilla (2015) [184]	Italy	Longitudinal	Larynx	EQ-5D	Self-completion	30	NA	Patients	68.7; 93%
Pickard (2016) [187]	US	Cross-sectional	All	EQ-5D-3L	Self-completion	50	NA	Patients	56.0; NA
Pottel (2015) [190]	Belgium	Longitudinal	All	EQ-5D-3L	Self-completion or interview on site (first assessment); by post (afterwards)	81	81%	Patients	72.0; 86%
Ramaekers (2011) [186]	Netherlands	Cross-sectional	All	EQ-5D-3L	Self-completion	396	93%	Patients	63.2; 70%
Ringash (2000) [177]	Canada	Cross-sectional	Larynx	TTO	Face-to-face interview	120	49%	Patients	65; 83%
Rogers (2006) [185]	UK	Cross-sectional	Oral cavity and oropharynx	EQ-5D-3L	Self-completion (by post)	224	64%	Patients	65; 58%
Szabo (2012) [155]	Canada	Cross-sectional	Larynx, lip, oral cavity, oropharynx	SG	Interview using script and prop	101	95%	Healthy subjects	47; 48%

Table 3.1 (cont.) Studies estimating HSUVs in HNC using direct or indirect methods (n=25).

Author (year)	Country	Study design	Cancer subsite(s)	Valuation method	Mode of administration	Sample size	Response rate	Participants	(Mean) age; % male
Truong (2016) [188]	US	RCT	Oropharynx, hypopharynx, larynx	EQ-5D-3L	Self-completion	818	87%	Patients	Arm CIS: 56.1; 86%. Arm CET/CIS: 57.3; 89%
van der Donk (1995) [181]	Netherlands	Cross-sectional	Larynx	TTO/SG/VAS	Face-to-face interview	39	NA	Laryngeal cancer patients (n=10), FOM cancer patients (n=10), experts (n=9), healthy subjects (n=10)	Laryngeal cancer patients: 62; NA. FOM cancer patients: 56; NA. Experts: 43; NA. Healthy subjects: 36; NA.
Weiss (1994) [178]	US	Cross-sectional	Pharynx, larynx	TTO	NA	3	NA	Clinical experts	NA

CIS: radiation-cisplatin without cetuximab; CET/CIS: radiation-cisplatin with cetuximab; COPD: chronic obstructive pulmonary disease; EQ-5D-3L: EuroQol 5-dimension 3-Level; EQ-5D-5L: EuroQol 5-dimension 5-Level; FOM: floor-of-the-mouth; HNC: head and neck cancer; HSUV: health state utility value; HUI3: Health Utility Index Mark 3; NA: not available; SF-6D: Short Form-6-dimension; SG: standard gamble; TTO: time trade off; VAS: visual analogue scale; VR-6D: Veterans RAND-6-dimension.

The 25 articles reported on HNC utility-related studies conducted in several European (Belgium, n=1; Finland, n=2; Italy, n=2; Netherlands, n=3 Spain; n=1 United Kingdom, n=2) and non-European countries (Australia, n=1; Canada, n=5 Morocco, n=1; United States, n=7). The great majority of the HSUVs came from cross-sectional surveys (n=18); the remaining articles (n=7) adopted a longitudinal design, including five prospective cohort studies [180] [184] [190] [191] [194] and two clinical trials [182] [188].

Sample sizes varied widely from 3 [178] to 818 [188], with a mean of 152 respondents per study. The response rate was between 49% [177] and 95% [155]. In most of the studies (n=18), the participants were HNC patients at various stages of disease and treatment pathway; in two studies [155] [173] healthy individuals from the general population were surveyed through the SG techniques, while in one case [178] the utility assessment was based on a consultation with a panel of experienced physicians. The remaining four studies [174] [176] [179] [181] retrieved utility measures from multiple subjects (i.e. healthy people, clinical experts, HNC patients and patients with other medical conditions) and reported HSUVs from each group separately.

In studies recruiting HNC patients, most were male, and the mean age was always above 55. Conversely, responders were generally younger and with a higher proportion of females in studies surveying individuals from the general population or clinical experts. The range of cancer subsites addressed by each study was quite broad: ten studies [171] [172] [175] [176] [182] [183] [186] [187] [190] [191] generally investigated utility in HNC without specifying any cancer site, six [174] [177] [180] [181] [184] [192] were related to laryngeal cancer, two [173] [179] addressed cancer in the oropharynx, one [189] recruited patients affected by cancer in the oral cavity and the remaining six [155] [178] [185] [188] [194] [193] focused on selected multiple sites (e.g. oropharynx, hypopharynx, and larynx).

The most common way (n=12 [176] [183] [184] [185] [186] [187] [188] [189] [190] [191] [192] [194]) of collecting utility data was by self-completion of a written survey (administered on site or by post/e-mail), followed by face-to-face interviews (n=6 [172] [174] [177] [179] [180] [181]); four studies adopted different administration options including group session (n=1 [173]), telephone or mail interview (n=1 [193]), interview using a script/prop (n=1 [155]), and computer-guided data collection (n=1 [171]). The administration method was not specified in three cases [175] [178] [182]. When HSUVs were obtained from the patients, the survey (or the interview) was usually scheduled during a clinical appointment or a hospital admission; in longitudinal studies [190] [194], surveys after the first were frequently delivered by post to the patient's address.

Mapping studies

The three studies deriving HSUVs in HNC using a mapping technique are described in Table 3.2. Among them, the first one [195] developed an original mapping algorithm using responses from HNC patients and was retrieved from the HERC database. Ordinary Least Square (OLS) regression was applied to establish a statistical relationship between the University of Washington Quality of Life questionnaire version 4 (UW QOL v4) and the EQ-5D-3L using a dataset of 89 patients treated for HNC. Thereafter, the responses of an additional 48 patients enrolled in the study were used as a validation database. The second study [196] was a cost-utility analysis reporting a mapping formula without any details on the technique adopted; the rationale for this mapping was based on a previous article showing a comparable responsiveness of EQ-5D and EORTC QLQ-C30 in patients with liver metastases [197]; QLQ-C30 data were retrieved from a randomized trial in HNC [198]. The third study [199] was a model-based economic evaluation reporting HSUVs for several HNC-related health states by applying an existing OLS model developed from data collected on various patients [200] to HRQoL data retrieved from a study in nasopharyngeal cancer [201].

Table 3.2 Mapping studies predicting HSUVs in HNC (n=3).

Author (year)	Country	Mapping technique	From (tool 1)	To (tool 2)	Sample's description (algorithm)	Sample size (algorithm)	Study ref. (algorithm)	Sample's description (tool 1)	Sample size (tool 1)	Study ref. (tool 1)
Chan (2014) [195]	Canada	OLS	UW QOL v4	EQ-5D-3L	Patients treated for HNC	89 (estimation); 48 (validation)				
Parthan (2009) [196]	UK	NS	EORTC QLQ-C30	EQ-5D-3L	NS	NS		Patients with locally advanced inoperable HNC	358	Vermorken (2007) [198]
Yong (2012) [199]	Canada	Application of a published algorithm using OLS	SF-36	HUI2	Various patients	6921	Nichol (2001) [200]	Patients with early stage nasopharyngeal cancer	51	Pow (2006) [201]

EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire; EQ-5D-3L: EuroQol 5-dimension 3-Level; HNC: head and neck cancer; HSUV: health state utility value; HUI3: Health Utility Index Mark 3; NS: not specified; OLS: ordinary least square regression; SF-36: 36-Item Short Form Health Survey; UW QOL v4: University of Washington Quality of Life questionnaire, version 4.

3.3.3 Study quality assessment

The quality assessment of the 25 studies using direct or indirect methods was based upon eight criteria, of which seven were given a score (Table 3.3). In all studies, the instrument adopted to estimate HSUVs was considered appropriate in relation with the participants enrolled. Additionally, most studies (84%) reported a description of the participants recruitment process, whilst only 56% of them clearly stated the inclusion/exclusion criteria. Information on missing data and techniques to deal with them were reported by a limited number of studies (24%). In half of the studies, the sample size was rather small (<100) and response rate was either low (<60%) or not reported. In reviewing these studies, it was evident that a few additional issues (criterion 8 [160]) should be considered when selecting sources to populate health economic models with utility parameters. First, some of the included studies are quite dated (published before 2000), thus describing health states that might not be realistic nowadays because of emerging treatment modalities, improvements in treatment-related morbidity and organ preservation techniques. Second, there might be potential sources of bias in reporting HRQoL results in clinical studies investigating one or more interventions, although the number of comparative trials retrieved by this search was very limited. Third, in some studies [171] [183] [193] [195], and especially those analysing HRQoL in multiple cancers including head and neck [171] [193], patient's characteristics (e.g. cancer stage/site, treatment phase) are poorly reported, thus making it difficult to match the study's HSUVs with the health states described in a cost-effectiveness model. Lastly, the great majority of studies are cross-sectional surveys, representing the quickest and cheapest method for gathering HRQoL data; however, longitudinal data collections are often more valuable since they facilitate capture of changes in utility values as cancer progresses through different phases.

With reference to mapping studies, only one [195] reported details on the developed algorithm, thus preventing a comparative evaluation of studies. This study presented a four-variable model to predict EQ-5D-3L utilities using OLS regression; coefficient values and error terms were clearly reported, and box-plot distributions of actual and predicted utilities provided. However, the authors did not justify the model choice in relation to the observed EQ-5D distribution, nor any additional tests or judgments made. The goodness-of-fit was presented as R^2 , mean absolute error (MAE) and root mean squared error (RMSE), which are considered of limited value in the mapping field. No demographic or clinical variables were included as covariates, which was recognized as a study limitation by the same authors. Moreover, when the sample size is small (as it was in this study), the most recent guidelines do not recommend splitting it for empirical validation [153].

3.3.4 Overview of HSUVs

A total of 346 original HSUVs were retrieved from 27 studies included in the review (Table A3.1), since one study [176] reported results only graphically in the article. The studies [172] [179] [181] providing the highest number of HSUVs (i.e. over 40) either adopted multiple techniques or interviewed several groups of respondents that yielded different values for each health state. In other cases [180] [184] [188] [190] [191] [194] [199], different HSUVs have been collected by the same participants over the study time points. HSUVs were reported as means in the great majority of studies (n=25), of which four [155] [173] [186] [188] also reported the median; the remaining two studies [182] [190] calculated a median value only. Among the measures of variance, standard deviation was the most frequently adopted (n=12), followed by the min-max range (n=7), and the interquartile range (n=5); several studies reported more than one measure type. In some cases [174] [178] [181] [192] [199], no measures of variability were reported, thus limiting the usefulness of the health state utility data.

Table 3.3 Quality assessment criteria in studies using direct or indirect methods (n=25).

Criteria*	[191]	[173]	[179]	[182]	[189]	[174]	[192]	[175]	[176]	[193]	[180]	[194]	[171]	[172]	[183]	[184]	[187]	[190]	[186]	[177]	[185]	[155]	[188]	[181]	[178]	Tot
Sample size	1	0	0	0	1	1	0	0	1	1	0	0	0	1	1	0	0	0	1	1	1	1	1	0	0	12
Selection and recruitment	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	21
Inclusion/exclusion criteria	0	1	1	1	0	1	0	0	1	0	1	0	1	1	0	1	0	1	1	1	0	1	0	1	0	14
Response rate**	1	1	0	0	1	0	0	1	0	1	0	1	0	1	0	0	0	1	1	0	1	1	1	0	0	12
Loss to follow-up***	1	C	C	0	C	C	C	C	C	C	0	1	C	C	C	1	C	1	C	C	C	C	1	C	C	5
Missing data	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0	0	0	1	1	0	6
Appropriateness of measure	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	25
Total score	5/7	4/6	3/6	3/7	4/6	4/6	1/6	2/6	4/6	5/6	3/7	5/7	3/6	5/6	3/6	4/7	2/6	6/7	6/6	4/6	4/6	5/6	5/7	4/6	1/6	

*The criteria come from Papaioannou et al. 2013 [160] ** Response rate is set equal to 0 if <60% or not reported in the study; ***C: cross-sectional study (not applicable criterion).

3.4 Discussion

This study reviews systematically published studies reporting HSUVs in HNC. Compared to a previous review [158], many more studies have been identified, most of which use the EQ-5D and were published from 2011 onwards. Overall, this review shows that HNC patients suffer from substantial HRQoL impairment over the different disease phases. However, there is a lack of research into the HRQoL in the recurrent and/or metastatic health states, with only one study [182] reporting a median EQ-5D utility value (i.e. 0.7) from the patients, which is less useful for the purposes of economic evaluation that focuses on mean costs and effects. Another study [179] elicits values for a range of recurrent disease states from healthy subjects and clinical experts using SG and VAS and obtains extremely heterogeneous results across the types of participants and methods. The same paucity of HSUVs was observed for treatment-related complications, which are addressed by three studies [155] [179] [180] only, possibly because of the infrequency of some of these events that restricts the data from patients in that health state.

Differences in utilities were found across studies even in the pre-treatment state. The choice of baseline utility is particularly relevant because it affects the incremental gain achievable by different therapeutic options [147], thus potentially biasing the estimated cost-effectiveness. The two Finnish studies [191] [194] using the 15D yielded higher utility values in patients shortly after diagnosis than those using the EQ-5D [188] [189]. This phenomenon has previously been observed in studies addressing other medical conditions [167] [202] [203]. There are many possible explanations for these discrepancies: different number of dimensions; the EQ-5D has generally been valued using TTO rather than VAS [204]; the preference weights have come from different populations (a Finnish value set is usually adopted for 15D) [202]; the EQ-5D, unlike

the 15D, can take negative values [204] [205]. The participants' characteristics might have also affected study results. For example, a study [190] addressing HRQoL in patients aged ≥ 65 years with HNC consistently provided lower HSUVs than other studies in either the pre-treatment, treatment, and follow-up phases, probably because of comorbidities and functional impairments usually affecting elderly people independently from cancer. Moreover, the use of different scoring algorithms may have contributed to variation in HSUVs in studies administering the EQ-5D.

Heterogeneity in utility values was particularly evident in the studies applying more than one technique to evaluate the same health state. Among them, in a study reporting HSUVs for different treatments, treatment-related complications, and remission/recurrence states in oropharyngeal cancer [179], the values obtained using a VAS scale were consistently lower than for the SG. In the study by Marcellusi et al. [171], patients in follow-up after treatment for HNC reported lower utility values when performing the TTO task than when responding to the EQ-5D questionnaire. Another study [172] compared five different (direct and indirect) methods to retrieve HSUVs from patients experiencing a similar health state (i.e. three months after completion of treatments and no evidence of recurrent disease). Unlike Marcellusi et al. [171], the method yielding the highest utility value in the overall sample (n=100) was TTO (0.94), followed by SG (0.91), EQ-5D (0.82), VAS (0.76) and HUI3 (0.75). That VAS scores are consistently lower than SG scores is well-known in the literature; in 2001, Torrance et al [168], after reviewing several studies, concluded that the relationship between the two instruments can be represented by a concave curve passing through 0 and 1. Moreover, the indirect methods involving MAUIs have been shown to yield systematically lower utility values than the direct ones in a wide range of diseases [206] for a variety of reasons. First, in MAUIs participants are not asked to consider their health status relative to death and thus, there is no disincentive in reporting more severe

health problems [207]. Second, respondents are forced to describe their complex medical conditions through a limited number of attributes, thus ignoring any positive feelings that would boost utility values. Third, it is likely that the general population used to obtain tariffs for MAUIs make a different trade-off between a given health state and death because they tend to be younger and healthier. Finally, the vignettes described in direct valuation tasks are usually more detailed than the MAUI health states [206]. In studies comparing alternative MAUIs, EQ-5D has been shown to provide higher utilities values compared to HUI2 and HUI3, which in turn yield higher values than SF-6D. As for the differences between EQ-5D and 15D, potential explanations are likely to be found in descriptive systems, preference measurement, source of community preferences, and scoring methods [208].

In addition, studies can be classified by the type of responders who valued the health states, either patients or healthy subjects. In the literature, some argue that patients are best placed to value the relevant health states, while others advocate valuation by healthy people who will not directly benefit from a new treatment but, in tax-based systems, will bear its cost. The latter claim that this will provide an unbiased estimate of the hypothetical health states [140] [209] and more consistency across appraisal of very different interventions. The review by Komatsuzaki et al. [158] showed that patients usually reported lower utilities than physicians and healthy people for health states associated to HNC. In the current review, only a few studies recruited participants from the general population, thus limiting the number of utilities comparisons across different types of responders. One study [181] confirms the conclusions reached by the previous review [158], whilst others [176] [179] found healthy subjects consistently providing lower utility estimates compared to patients and healthcare professionals.

This study facilitates the identification of HSUVs for use in future HNC economic evaluations, including the one presented in this thesis (Chapter V). The number of

retrieved studies was quite large, with almost 350 distinct HSUVs collected from them. Most of the utility values were collected during the treatment phase or shortly after the completion of treatment, whilst limited evidence is available for the health-related utility assessment in HNC recurrent and end-of-life states. Due to the variety of health state definitions and valuation techniques across the studies, it was not possible to perform a quantitative synthesis of the results [136]. Moreover, unlike cost-effectiveness studies where structured guidelines exist to support authors and reviewers in assessing their quality [73], recommendations for valuation studies specifically aimed at measuring HSUVs are more fragmented or method-specific [144]. In this review, the assessment of study quality was based on a set of generic recommendations elaborated by a previous study [160] and arbitrarily modified to allow a quantitative scoring of the studies adopting direct and indirect techniques to estimate HSUVs; for mapping studies, the analysis relied instead on recent ISPOR guidelines [153].

Although there is no universally accepted theoretical basis for choosing direct or indirect methods [206], the use of the EQ-5D is favoured by several agencies including NICE, the Canadian Agency for Drugs and Technologies in Health and the French National Authority for Health [136]. In a recent position statement [210], NICE recommends the use of EQ-5D-3L for base-case analyses, or mapping EQ-5D-5L responses onto the 3L valuation set, to derive HSUVs, since further research is needed to explore the impact of adopting the 5L valuation set on technology appraisal. In model-based cost-effectiveness studies, where there is a choice of HSUVs, those using the value set of the jurisdiction for which a decision is being made are usually preferred. Moreover, HSUVs should be collected from studies enrolling patients with demographic and clinical characteristics that mostly resemble those of potential recipients of the intervention under investigation in the model. Until now, studies relying on direct techniques represent the only available source to retrieve HSUVs for

recurrent disease, palliative states, or treatment-related complications in HNC. Although considered as qualitatively inferior to MAUIs [136], these methods can provide values for cost-effectiveness analyses where the ‘vignettes’ presented in the choice task fit with the health states addressed in the model. Finally, in the absence of preference-based data, mapping from disease-specific instruments to generic MAUIs may represent a valuable alternative [209]; however, the only algorithm published to date in HNC [195] does not map from one of the HRQoL tools most frequently adopted in cancer studies, such as the EORTC QLQ-C30 [211] and the FACT-G [212]. Greater availability of mapping functions would facilitate the comparison of treatments using HRQoL data from many randomized controlled trials that only collected disease-specific health status information. Overall, the use of different techniques for utility elicitation might have substantial implications in cost-utility analyses; for example, it has been shown [206] that MAUIs, compared to direct valuation, tend to favour non-lifesaving treatments over interventions preventing or delaying death. Thus, regulatory bodies should avoid a mixture of methods in their decision processes to avoid a biased allocation of healthcare resources. Moreover, health economic modelers are always recommended to extensively test the uncertainty around the utility parameters in sensitivity analyses [206].

3.5 Conclusions

This study improves understanding of preference-based HRQoL measurement in HNC by systematically reviewing and critically evaluating studies that estimated HSUVs in this cancer setting. Utility values are an essential parameter but also a major source of uncertainty in model-based economic evaluations, where it is common to select them from a single study based on clinical considerations [136] [162]. Further studies on the health-related utility assessment from HNC patients using MAUIs in recurrent and

terminal states are encouraged; this is a major issue in models evaluating follow-up programs, which inevitably includes such severe health states. Additional research on mapping algorithms to convert disease-specific HRQoL results onto preference-based HSUVs would be of value in this cancer population; this thesis, indeed, provides the first set of algorithms converting EORTC QLQ-C30/-H&N35 scores into EQ-5D utility values using HNC data (Chapter IV). Overall, the methods used to identify utility values within a growing body of HRQoL literature should be increasingly systematic and justified in future studies.

4 MAPPING THE CANCER-SPECIFIC EORTC QLQ-C30 AND QLQ-H&N35 TO THE GENERIC EQ-5D-5L IN HEAD AND NECK CANCER

4.1 Introduction

Overall survival and progression free-survival are commonly accepted primary endpoints in cancer studies; however, due to the aggressiveness of treatments, HRQoL is often measured as well through a variety of tools [213]. As extensively discussed in Chapter III, cost-utility analyses are increasingly required to demonstrate the clinical and economic value of expensive cancer therapies. The primary outcome of these analyses is the QALY that, for its calculation, requires the assessment of HSUVs, defined as the preference weights associated to each health state experienced by the patient over time. Some generic HRQoL instruments, also known as MAUIs, such as the EQ-5D questionnaire, provide utility values for the health states described by the tool's dimensions and levels based on the preferences elicited from the general population. However, these tools are not routinely included in clinical studies, because clinicians tend to prefer disease-specific questionnaires, which are more sensitive to changes in symptomatology and cover a wider range of health issues related to a condition [62] [214]. Among them, the EORTC QLQ-C30 (QLQ-C30 thereafter) is the most commonly used HRQoL tool in cancer [214] and has been widely applied in several oncological studies across Europe [215]. In their original form, QLQ-C30 scores cannot be used directly in economic evaluation studies, as they are not measures of utility elicited from the general population [216]. EORTC-8D was recently developed using the QLQ-C30 for use as a cancer-specific preference-based measure [64]. However, HSUVs generated from generic MAUIs (and, among them, preferably the

EQ-5D) are usually recommended for comparability across different therapeutic areas [217].

Mapping (or cross walking) is a useful tool to calculate patient-level HSUVs in studies that do not adopt any generic preference-based MAUIs [218]. The mapping's task is to establish a statistical relationship between a 'source' measure and, usually, a generic 'target' one. In the absence of generic preference-based data, NICE endorses mapping from other HRQoL measures collected in the relevant trials to the EQ-5D. Over the last few years, several mapping functions for deriving EQ-5D utilities from QLQ-C30 scores have been published for a variety of cancers. However, the existing studies do not cover the full spectrum of malignancies and it is not clear whether any function might reasonably be extended to other types of cancers [219]. Moreover, most algorithms mapped to the 3-level version of EQ-5D (EQ-5D-3L) [165], which is beginning to be replaced by the 5-level one (EQ-5D-5L). Finally, linear regression is the most widely applied technique for mapping, although it has been shown to be systematically biased in modelling EQ-5D data since it poorly fits with their typical distribution characteristics (i.e. boundedness, skewness, multimodality) [216] [220].

This thesis is focussed on HNC, which is a major public health concern in the developed world. As reported in Chapter III, treatments for HNC are usually aggressive and significantly affect patient's HRQoL through functional impairments, physical disfigurement, and psychological distress. Surgical procedures often cause facial alterations or change an individual's appearance, with subsequent worsening in social interaction and eating. Some of these consequences (such as pain) may improve over time, whilst others (e.g. dysphagia, speaking/chewing difficulties, and dryness of mouth) may last for several months after treatment ends [221]. Thus, HRQoL is increasingly viewed as an important outcome in clinical studies of HNC, alongside other traditional parameters such as survival, recurrence, and drug toxicity [222] [223].

A variety of validated instruments are available to measure HRQoL in HNC; among them, the most popular are the QLQ-C30 supplemented with a 35-item HNC cancer module (QLQ-H&N35; H&N35 thereafter), the University of Washington Quality of Life (UWQOL) questionnaire and the FACT-G/FACT-H&N questionnaire [221]. However, as shown in Chapter III, the literature on HSUVs and, particularly, their measurement using EQ-5D are still limited in HNC, especially for advanced cancer stages. Moreover, until now, there are no published original algorithms to map EQ-5D utility values from QLQ-C30 and H&N35.

This chapter contributes to filling knowledge “gaps” in mapping in three different ways. First, using data from the ongoing HETeCo trial, it maps from the QLQ-C30 and H&N35 scales/items to EQ-5D-5L in HNC, thus extending the mapping literature in cancer in general and providing the first mapping in HNC using the EORTC questionnaires. Second, it provides a set of mapping functions for the available EQ-5D-5L tariffs, which correspond to the preference weights elicited from the general population in different countries, so that future users can select the most appropriate model for their data. Third, it tests (and compares) alternative mapping techniques as suggested by the most recent literature on the topic. Overall, this work aims at facilitating the assessment of HSUVs in HNC in the absence of original EQ-5D data, and, consequently, the performance of cost-utility comparisons of novel treatments in this cancer population.

4.2 Methods

In carrying out the analyses, the recently published guide to good practices from ISPOR [153] is mainly followed. The MAPS statement [224] [225], a 23-item checklist of recommendations that authors should consider when reporting mapping studies, is also completed, and attached to the appendix (Table A4.1). Furthermore, the analyses refer

to the recommendations from Longworth et al. [215] and other recent studies obtaining EQ-5D utilities from QLQ-C30 [63] [226].

4.2.1 Instruments

The EQ-5D is a commonly used generic instrument for measuring HRQoL in a variety of medical conditions. The instrument consists of five dimensions (i.e. mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) plus a 100-point visual analogue scale (EQ-VAS) for rating overall health status. In its original version (EQ-5D-3L), each dimension has three levels: no problems, moderate problems, and extreme problems; two intermediate levels (i.e. slight problems and severe problems) have been added between levels 1 and 2 and levels 2 and 3, respectively (<http://www.euroqol.org>) in the EQ-5D-5L, which is adopted in the HETeCo trial. EQ-5D responses can be converted into a summary utility score using one of the existing sets of country preference weights (or “tariffs”). A number of tariff sets (i.e. England [227], Netherlands [228], Canada [229], Uruguay [230], Japan [231], South Korea [232], and China [233]), as reported on the EuroQol website (last updated: 18th April 2017) are used to calculate the EQ-5D-5L overall utility in this chapter. Where a full dataset of HSUVs is not provided, the preferred model as specified by the authors is used to calculate EQ-5D-5L values.

The QLQ-C30 is a cross-culturally validated 30-item questionnaire to assess HRQoL in cancer patients. Version 3.0 is currently the standard recommended for use in clinical studies. A scoring system transforms raw responses into 15 overall indexes including five functional scales (physical, role, emotional, cognitive, and social), three symptoms scales (fatigue, nausea or vomiting, and pain), a global health status, and six single items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea, and financial difficulties). Possible scores range from 0 to 100, with higher mean scores on the

functional scales and global status representing better health whilst higher mean scores on the symptom scales/items indicate worse symptomatology [211].

The H&N35 is a supplement of the QLQ-C30 instrument that is recommended for use in HNC patients. The module comprises 35 questions that can be combined into 18 summary scores including seven multi-scale items (i.e. pain, swallowing, senses, speech, social eating, social contact, and sexuality) and eleven single items. For all scores (0-100), higher numbers indicate more health problems [211].

4.2.2 Sample and data collection

Data for this study are obtained from the ongoing multicentre HETeCo trial [33], already described in Chapter I. Patients' demographics and clinical information, together with informed consent, are collected at enrolment. HRQoL tools (i.e. EQ-5D-5L, QLQ-C30 and H&N35) are administered at the enrolment and at every other visit in the first two years and then at each visit; follow-up visits occur every 2-6 months according to cancer site and study time. This chapter uses HRQoL data from a sample of patients enrolled in the trial until March 2018; data from all time points are pooled to increase the sample size and therefore the statistical precision of the estimates. Missing values are assumed to occur completely at random; for QLQ-C30 and H&N35 functional/symptom scales, the approach reported in the scoring manual is followed to calculate the missing scores, whenever at least half the items are completed [211]. Single QLQ-C30 and H&N35 items and EQ-5D-5L responses are imputed instead through the Last Observation Carried Forward (LOCF) method given the longitudinal nature of the data; according to this rule, the missing item is not imputed when occurring at the first visit. Questionnaires without a corresponding observation in the alternative instrument are dropped from the analysis [63].

4.2.3 Data analysis

As preliminary analyses, the degree of overlap between the source measures (QLQ-C30 and H&N35) and the target one (EQ-5D-5L) is assessed using Spearman's rank correlations to justify a mapping exercise [224]. The same technique is applied to QLQ-C30 and H&N35 summary scores separately to state whether any independent variables are highly correlated (i.e. correlation coefficient $> |0.7|$) and thus not recommended for inclusion within the same regression model [215]. Moreover, a plot of EQ-5D-5L value distributions is provided according to different country value sets to help inform the identification of appropriate regression techniques for mapping; models' choice also follows the existing literature [153] [215].

Three mapping techniques are applied to model the EQ-5D-5L utility values: (1) linear mixed-effects regression (*mixed* command in STATA); (2) random-effects Tobit model (*xttobit* command in STATA) (3) adjusted limited dependent variable mixture model (ALDVMM, *aldvmm* command in STATA).

- (1) The linear model with a random effect is an extension of the OLS model that allows for multiple observations per patient (i.e. the 'cluster'); patient's responses at different time points, indeed, are likely to be correlated. This model is also termed a mixed-effects model because the parameters are a mix of fixed and random variables; the between-cluster variability is modelled with a random effect, i.e. as a random intercept term at the patient level [226]. The model is estimated through maximum likelihood (ML) estimator; robust standard errors are used to protect against non-normality [234].
- (2) In linear models, predictions are free to range from negative to positive infinity and therefore may assume values outside the existing range of the EQ-5D-5L utilities (e.g. -0.281 and 1 for the English tariff). Moreover, they tend to

overpredict utilities in poor health states and, conversely, under-predict utilities in patients in relatively good health [220]. The Tobit model, also known as a censored regression model, estimates linear relationships among variables by accounting for the bounded nature of EQ-5D. In this study, the lower limit (health state 55555) for the dependent variable (EQ-5D-5L utility) varies according to the country set adopted, while the upper limit (health state 11111) is the same except for the Canadian value set [229], where the maximum value is fixed at 0.95 (instead of 1). The predicted values above or below the limits take the value of the thresholds themselves, so that they remain in the existing instrument range [63] [215] [218]. Again, a random-effects model (*xttobit*) is adopted to account for non-independent observations, using patient id as the panel variable.

- (3) The EQ-5D data typically show a few additional characteristics beyond the left and right boundedness. There is generally a “mass” of observations at the maximum value of 1 corresponding to perfect health, and a “gap” between this bulk of observations at 1 and the next feasible value (e.g. 0.951 for health state 11211 using English tariffs). However, due to the increased number of levels in the new version, this decrement is smaller for EQ-5D-5L compared to EQ-5D-3L, where the next value after 1 is 0.883 [227]. Moreover, the EQ-5D data tend to present negative skewness and multimodal distribution, which may invalidate the normality assumption of linear models. The Tobit model, which has been developed as an alternative to the linear model for dealing with bounding only, may not adequately address all these features. The ALDVM model was recently developed as a flexible alternative, and with better performance compared to traditional regression techniques in modelling EQ-5D data with non-linear or unknown distributions. Further details about the model can be found elsewhere

[220] [234] [235] [236]. In brief, ALDVMM is an adjusted Tobit model that allows several latent classes to be considered simultaneously within a data distribution, each expressing a different relationship between the EQ-5D dependent variable and the set of independent variables; thus, the same variable can be highly significant in certain components and not in others, so that it may be erroneously excluded in standard models [215]. ALDVMM is considered semiparametric, in the middle between a fully parametric model with one component only and a non-parametric one where the number of classes coincides with the sample size [234]. In this study, robust standard errors with patient id as a cluster variable are used to reflect the correlations between per patient observations [220].

In all developed models, the EQ-5D-5L summary score is the dependent variable, while the explanatory variables are the QLQ-C30 and H&N35 scale/item results. Modelling from QLQ-C30 increases the usability of the algorithms in cancers other than HNC, while functions with H&N35 variables are likely to be more sensitive and responsive to the symptoms associated with HNC [237]. Because of the small sample size, models including both QLQ-C30 and H&N35 scores are not estimated, according to the rule of 10 observations per variable [215]. The signs of the QLQ-C30 functional scales and global health status are expected to be positive, while those of the QLQ-C30 and H&N35 symptom scales/items are expected to be negative. Age and gender (female=1; male=0) are included as covariates in all models. In the interest of developing a parsimonious mapping, backward elimination with a significance level of 0.05 is used to select variables entering the final models, except for ALDVMM where some variables may be significant in one component and not in the other(s) [215] [236]; in one-component ALDVM models (that are like Tobit but accounting for “gaps” in EQ-5D distributions [235]), backward selection is applied as usual. However, since this

selection process increases the model's internal validity at the expenses of generalizability [215], both full and reduced models are reported. All the developed models are additive, thus implying linear independence between explanatory variables [226] [237]. Linear model coefficients can be easily applied to external datasets to generate EQ-5D utilities [63] [226], although predicting from ALDVMM is more complex since the calculation involves the probability each value falls within one of the different model components [226].

Model goodness-of-fit is measured using penalized likelihood criteria including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), where the smaller the value, the better the model fit; the BIC is also used as indicator to select the best number of components within each ALVDMM [235]. For each model, the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are calculated to estimate the magnitude of difference between the observed and predicted values [234]. In details, the MAE is the average of the absolute differences, while the RMSE is the root of the average of the squared differences [62]. Moreover, summary statistics of predictions are reported for each model [236]. Scatterplots of observed versus predicted EQ-5D-5L are also displayed to investigate how goodness-of-fit varies across the data distribution [62] [63]. Predicted EQ-5D-5L utilities are obtained using the command *predict*. All statistical analyses are performed using STATA version 14.0 (StataCorp LP, College Station, TX).

4.3 Results

4.3.1 Descriptive statistics

Baseline demographic and clinical variables of the patients (n=97) enrolled in the study after being treated for their primary HNC are presented in Table 4.1. Mean age is around 63 and most patients (80.4%) are men. HNC is mostly localized in the oral

cavity (45.4%), followed by the oropharynx (27.8%), larynx (21.6%) and hypopharynx (5.2%). The great majority of the patients (61.9%) are ex-smokers, while 18.5% are still smoking at the enrolment time and 19.6% never smoked.

Table 4.1 Sample's characteristics.

	N (%)
No. of patients	97 (100.00)
<i>Demographics</i>	
Male	78 (80.41)
Age (mean \pm SD; range)	63.07 \pm 10.58; 33-90
<i>Cancer site</i>	
Oral cavity	44 (45.36)
Oropharynx	27 (27.84)
Larynx	21 (21.65)
Hypopharynx	5 (5.15)
<i>Smoking status</i>	
Ex-smoker	60 (61.86)
Never smoked	19 (19.59)
Current smoker	18 (18.56)

SD: standard deviation.

Descriptive statistics of EQ-5D-5L and QLQ-C30 and H&N35 scores, pooling all available information across study visits, are summarized in Table 4.2. The number of observations per patient range between 1 and 7. A total of 84 unique EQ-5D-5L profiles are reported in the database, the most frequent one being 11111 (17.5%) followed by 11121 (13.5%), and 11122 (5.2%). Level 1 (no problems) is reported most frequently in all dimensions except for pain/discomfort, where the modal response is level 2 (slight problems). The dimension reporting the highest number of responses at level 1 is self-care (84.3%), while level 5 (extreme problems) responses are 1.3% in usual activity, pain/discomfort, and anxiety/depression, and 0% in mobility and self-care. The Uruguayan preference weights yield the highest average utility (0.905 \pm 0.11), followed by the English (0.839 \pm 0.16), Canadian (0.817 \pm 0.15), Korean (0.812 \pm 0.13), Chinese (0.803 \pm 0.21), Japanese (0.789 \pm 0.16), and Dutch (0.786 \pm 0.19).

The mean QLQ-C30 global health score is equal to 66.73 (\pm 18.60) and varies across the entire interval from 0 to 100. Among functional scales, cognitive functioning is the one

presenting the highest score (87.36 ± 18.10), while emotional is the scale yielding the lowest value (78.54 ± 21.33). The worst problem experienced by patients is fatigue (24.08 ± 21.62) followed by insomnia (23.63 ± 27.58), financial problems (18.99 ± 30.07) and constipation (18.44 ± 26.73). Among H&N35 scores, the three symptoms reporting higher scores (that indicate more problems) are dry mouth (42.14 ± 31.43), sticky saliva (39.02 ± 31.79), and senses problems (26.29 ± 24.23).

Table 4.2 Summary statistics of EQ-5D-5L and QLQ-C30/-H&N35 scores.

<i>EQ-5D-5L (responses)</i>	N	%					
Mobility							
Level 1	146	63.8					
Level 2	49	21.4					
Level 3	30	13.1					
Level 4	4	1.7					
Level 5	0	0.0					
Self-care							
Level 1	193	84.3					
Level 2	19	8.3					
Level 3	13	5.7					
Level 4	4	1.7					
Level 5	0	0.0					
Usual activities							
Level 1	129	56.3					
Level 2	58	25.4					
Level 3	27	11.8					
Level 4	12	5.2					
Level 5	3	1.3					
Pain							
Level 1	69	30.1					
Level 2	100	43.7					
Level 3	53	23.1					
Level 4	4	1.8					
Level 5	3	1.3					
Anxiety/depression							
Level 1	120	52.4					
Level 2	61	26.7					
Level 3	44	19.2					
Level 4	1	0.4					
Level 5	3	1.3					
<i>EQ-5D-5L (utility)</i>	N	Mean	SD	Min	Max	Below 0	Ceiling at 1
England	229	0.839	0.16	0.042	1.000	0 (0%)	40 (17.5%)
Netherlands	229	0.786	0.19	-0.180	1.000	2 (0.9%)	40 (17.5%)
Canada	229	0.817	0.15	0.068	0.949	0 (0%)	0 (0%)
Uruguay	229	0.905	0.11	0.414	1.000	0 (0%)	40 (17.5%)
Korea	229	0.812	0.13	0.303	1.000	0 (0%)	40 (17.5%)
Japan	229	0.789	0.16	0.251	1.000	0 (0%)	40 (17.5%)
China	229	0.803	0.21	-0.061	1.000	2 (0.9%)	40 (17.5%)
<i>EQ-5D-5L (VAS)</i>	232	70.00	18.05	0	100		

EQ-5D-5L: EuroQol Five-Dimension Five-Level; SD: standard deviation; VAS: visual analogue scale.

Table 4.2 (cont.) Summary statistics of EQ-5D-5L and QLQ-C30/-H&N35 scores.

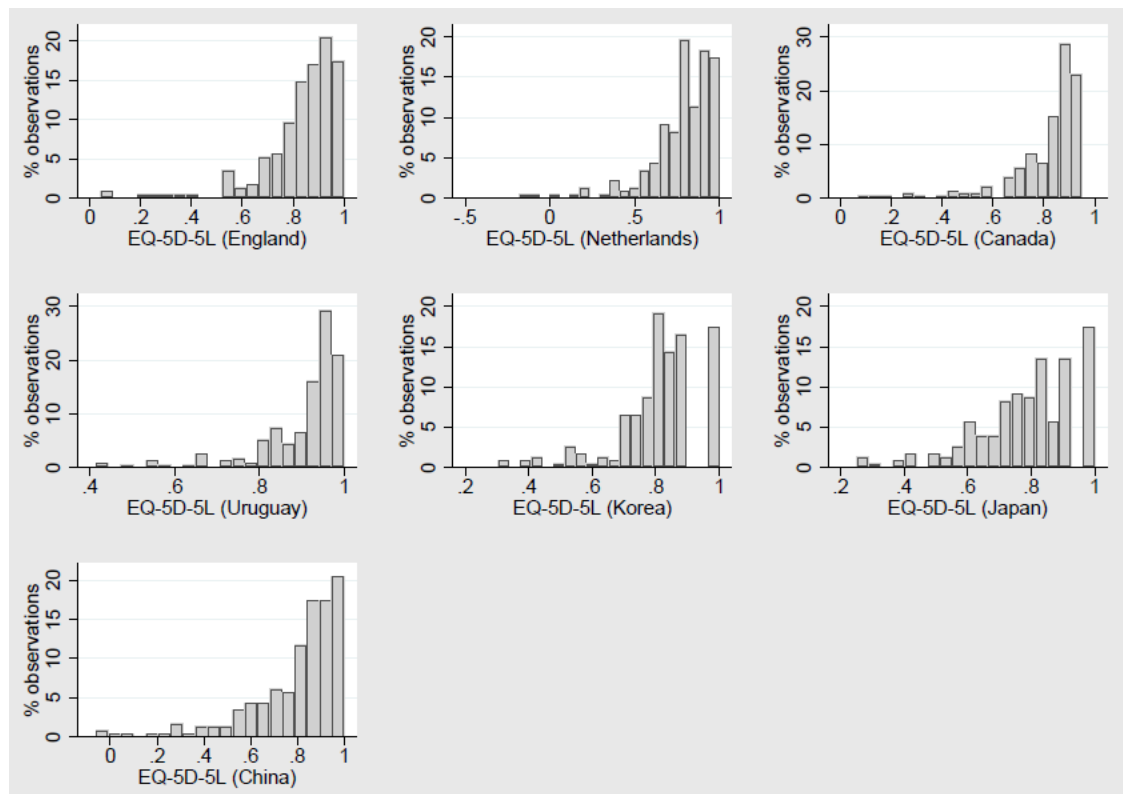
<i>QLQ-C30</i>	N	Mean	SD	Min	Max
Global health status (GH)	243	66.73	18.60	0	100
Physical functioning (PF)	245	82.51	17.60	13.33	100
Role functioning (RF)	245	83.06	22.43	0	100
Emotional functioning (EF)	244	78.54	21.33	0	100
Cognitive functioning (CF)	244	87.36	18.10	16.67	100
Social functioning (SF)	244	82.79	22.92	0	100
Fatigue (FA)	245	24.08	21.62	0	100
Nausea and vomiting (NV)	245	3.67	8.93	0	66.67
Pain (PA)	245	15.78	19.80	0	100
Dyspnoea (DY)	245	15.78	22.49	0	100
Insomnia (SL)	244	23.63	27.58	0	100
Appetite loss (AP)	245	14.69	24.92	0	100
Constipation (CO)	244	18.44	26.73	0	100
Diarrhoea (DI)	244	3.82	11.47	0	66.67
Financial problems (FI)	244	18.99	30.07	0	100
<i>QLQ-H&N35</i>					
Pain (HNPA)	246	17.89	19.12	0	100
Swallowing (HNSW)	246	18.49	19.76	0	91.67
Senses problems (HNSE)	246	26.29	24.23	0	100
Speech problems (HNSP)	246	23.01	22.41	0	100
Trouble with social eating (HNSO)	246	21.78	21.58	0	100
Trouble with social contact (HNSC)	246	13.03	17.80	0	93.33
Less sexuality (HNSX)	243	23.11	29.72	0	100
Teeth (HNTE)	245	22.45	33.46	0	100
Opening mouth (HNOM)	246	25.34	30.18	0	100
Dry mouth (HNDR)	246	42.14	31.43	0	100
Sticky saliva (HNSS)	246	39.02	31.79	0	100
Coughing (HNCO)	246	13.55	20.13	0	100
Felt ill (HNFI)	246	4.20	12.61	0	100
Pain killers (HNPK)	246	24.80	43.27	0	100
Nutritional supplements (HNNU)	246	23.58	42.53	0	100
Feeding tube (HNFE)	246	4.47	20.71	0	100
Weight loss (HNWL)	246	15.45	36.21	0	100
Weight gain (HNWG)	246	23.58	42.53	0	100

QLQ-C30: 30-item Core Quality of Life Questionnaire; QLQ-H&N35: 35-item Head and Neck Cancer Quality of Life Questionnaire.

After removing 20 QLQ-C30 questionnaires and another 20 H&N35 without a correspondence in the EQ-5D-5L measure, 225 EQ-5D-5L/QLQ-C30 and 226 EQ-5D-5L/QLQ-H&N35 questionnaire pairs are finally available for the mapping exercise. Moreover, 11 missing EQ-5D-5L responses are imputed based on LOCF; in QLQ-C30 questionnaires, 13 functional/symptom scales are imputed based on EORTC manual instructions and other 22 scores using LOCF; in H&N35, 11 symptom scales are imputed using the manual, while 40 scales/items are imputed based on LOCF.

A plot of the EQ-5D-5L utility distribution for each value set is reported in Figure 4.1. Some common features can be observed, such as the presence of large spikes (especially at the health states closest to the perfect health), negative skewness, and multimodality; in Korean and Japanese values, a clear “gap” is evident between the mass of observations at 1 and the next feasible values.

Figure 4.1 Observed EQ-5D-5L utility values.



4.3.2 Mapping functions

Table 4.3 reports Spearman’s rank-order correlation coefficients among EQ-5D-5L utilities and QLQ-C30/H&N35 scores. As expected, a positive correlation is found between EQ-5D-5L and QLQ-C30 global health status/functional scores, whilst a negative correlation exists with QLQ-C30 and H&N35 symptom scales and single items (except for weight gain that can be interpreted, indeed, as a healthy effect in cancer patients and is not statistically significant). All correlation coefficients (apart from weight gain) show a statistically significant value ($p < 0.05$); thus, the amount of overlap

between the sources (QLQ-C30 and H&N35) and the target measure (EQ-5D-5L) is considered high enough to perform the mapping. Tables 4.4 display instead Spearman correlation coefficients for the models' explanatory variables only (QLQ-C30 and H&N35 scores); none of coefficients is above |0.7| and, consequently, no variables are excluded from the model at this stage.

Table 4.3 Spearman's rank correlation coefficients between EQ-5D-5L utilities and QLQ-C30/-H&N35 summary scores.

	EQ-5D-5L						
	England	Netherlands	Canada	Uruguay	Korea	Japan	China
GH	0.5557*	0.5497*	0.5591*	0.5713*	0.5623*	0.5787*	0.5844*
PF	0.6511*	0.6344*	0.6530*	0.6305*	0.6769*	0.6760*	0.6737*
RF	0.6126*	0.6158*	0.6321*	0.6373*	0.6336*	0.6319*	0.6421*
EF	0.6503*	0.6585*	0.6192*	0.5719*	0.5922*	0.6444*	0.6090*
CF	0.4322*	0.4320*	0.4126*	0.4314*	0.4127*	0.4380*	0.4160*
SF	0.5663*	0.5662*	0.5524*	0.5771*	0.5484*	0.5748*	0.5617*
FA	-0.5538*	-0.5534*	-0.5433*	-0.5436*	-0.5353*	-0.5506*	-0.5346*
NV	-0.3081*	-0.3003*	-0.2822*	-0.2864*	-0.2855*	-0.3078*	-0.2844*
PA	-0.5804*	-0.5766*	-0.6102*	-0.5784*	-0.5698*	-0.5860*	-0.5994*
DY	-0.4445*	-0.4267*	-0.4296*	-0.3966*	-0.4493*	-0.4396*	-0.4189*
SL	-0.5256*	-0.5282*	-0.5078*	-0.4895*	-0.4805*	-0.5185*	-0.5054*
AP	-0.3934*	-0.3953*	-0.3489*	-0.3347*	-0.3422*	-0.3910*	-0.3491*
CO	-0.1970*	-0.2038*	-0.1819*	-0.2233*	-0.1854*	-0.2151*	-0.1966*
DI	-0.2173*	-0.2195*	-0.2099*	-0.1842*	-0.2101*	-0.2096*	-0.2055*
FI	-0.4449*	-0.4447*	-0.4530*	-0.4627*	-0.4370*	-0.4396*	-0.4485*
HNPA	-0.4790*	-0.4714*	-0.4884*	-0.4579*	-0.4590*	-0.4640*	-0.4720*
HNSW	-0.3720*	-0.3883*	-0.3861*	-0.3782*	-0.3546*	-0.3702*	-0.3917*
HNSE	-0.1942*	-0.1988*	-0.1736*	-0.2020*	-0.1683*	-0.2005*	-0.1812*
HNSP	-0.4502*	-0.4661*	-0.4351*	-0.4375*	-0.4161*	-0.4469*	-0.4289*
HNSO	-0.3176*	-0.3251*	-0.3265*	-0.3638*	-0.3154*	-0.3253*	-0.3322*
HNSC	-0.5166*	-0.5328*	-0.4969*	-0.5246*	-0.4893*	-0.5136*	-0.4958*
HNSX	-0.1665*	-0.1743*	-0.1626*	-0.1739*	-0.1716*	-0.1781*	-0.1723*
HNTE	-0.2680*	-0.2605*	-0.2622*	-0.2577*	-0.2421*	-0.2570*	-0.2608*
HNOM	-0.2703*	-0.2639*	-0.2703*	-0.3052*	-0.2760*	-0.2855*	-0.2816*
HNDR	-0.2051*	-0.1987*	-0.2107*	-0.1744*	-0.2082*	-0.2069*	-0.2106*
HNSS	-0.2658*	-0.2732*	-0.2646*	-0.2350*	-0.2387*	-0.2564*	-0.2544*
HNCO	-0.2281*	-0.2225*	-0.2252*	-0.2179*	-0.2254*	-0.2215*	-0.2226*
HNFI	-0.3579*	-0.3543*	-0.3641*	-0.3675*	-0.3512*	-0.3645*	-0.3644*
HNPK	-0.3146*	-0.3277*	-0.3449*	-0.3220*	-0.3287*	-0.3303*	-0.3575*
HNNU	-0.1948*	-0.1982*	-0.1621*	-0.1917*	-0.1725*	-0.1934*	-0.1740*
HNFE	-0.1545*	-0.1630*	-0.1506*	-0.1588*	-0.1486*	-0.1446*	-0.1423*
HNWL	-0.1475*	-0.1538*	-0.1501*	-0.1414*	-0.1375*	-0.1547*	-0.1491*
HNWG	0.0727	0.0702	0.0817	0.1031	0.0860	0.1054	0.1029

* p<0.05. GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems; HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain.

Tables 4.4 Spearman's rank correlation coefficients between QLQ-C30/-H&N35 summary scores.

Table 4.4 (A) Spearman's rank correlation coefficients between *QLQ-C30* summary scores.

	GH	PF	RF	EF	CF	SF	FA	NV	PA	DY	SL	AP	CO	DI	FI
GH	1														
PF	0.5615	1													
RF	0.5651	0.6070	1												
EF	0.4133	0.4671	0.5494	1											
CF	0.4214	0.4770	0.4281	0.4891	1										
SF	0.4280	0.5065	0.5881	0.5324	0.3832	1									
FA	-0.5079	-0.6637	-0.6251	-0.5981	-0.5771	-0.4792	1								
NV	-0.1399	-0.2429	-0.2986	-0.3055	-0.2576	-0.2645	0.3739	1							
PA	-0.4325	-0.5234	-0.5306	-0.5432	-0.3496	-0.4703	0.5290	0.2297	1						
DY	-0.3715	-0.5424	-0.5233	-0.4558	-0.3944	-0.3167	0.5071	0.1999	0.3490	1					
SL	-0.3287	-0.4762	-0.4598	-0.5790	-0.3033	-0.5006	0.4927	0.2392	0.4518	0.3917	1				
AP	-0.3394	-0.4324	-0.3378	-0.4500	-0.3260	-0.3706	0.4742	0.3359	0.3560	0.3086	0.4002	1			
CO	-0.2094	-0.2324	-0.2363	-0.2651	-0.3948	-0.1082	0.2734	0.2530	0.1425	0.1387	0.1708	0.1743	1		
DI	-0.1587	-0.1681	-0.1635	-0.2276	-0.1420	-0.1140	0.1745	0.0990	0.1068	0.1627	0.2145	0.0883	0.0086	1	
FI	-0.3159	-0.3065	-0.4560	-0.4074	-0.2367	-0.5172	0.3614	0.1285	0.4080	0.2642	0.4767	0.2175	-0.0431	-0.0415	1

Table 4.4 (B) Spearman's rank correlation coefficients between *QLQ-H&N35* summary scores.

	HNSW	HNSX	HNSP	HNSO	HNSC	HNSX	HNTE	HNOM	HNDR	HNSS	HNSO	HNFI	HNP	HNSO	HNFE	HNWL	HNWG	
HNSW	1																	
HNSX	0.4634	1																
HNSP	0.2276	0.2416	1															
HNSO	0.3165	0.4112	0.2835	1														
HNSC	0.4241	0.5768	0.3816	0.3616	1													
HNSX	0.2898	0.3157	0.2644	0.5838	0.5539	1												
HNTE	0.2152	0.3040	0.2718	0.3102	0.4032	0.3614	1											
HNOM	0.4340	0.1634	0.1272	0.2469	0.2159	0.2537	0.1677	1										
HNDR	0.3204	0.4025	0.1647	0.1561	0.4544	0.2745	0.2818	0.1802	1									
HNSS	0.4496	0.3808	0.3049	0.1528	0.2677	0.0977	0.2785	0.1539	0.3323	1								
HNSO	0.4175	0.3895	0.2444	0.3154	0.3367	0.2462	0.2055	0.2273	0.2662	0.5157	1							
HNFI	0.1215	0.1974	0.0411	0.2997	0.1543	0.2453	0.0769	0.1435	-0.0207	0.0312	0.1142	1						
HNP	0.2915	0.3209	0.0707	0.2130	0.1889	0.1847	0.0356	0.0969	0.0595	0.1443	0.0899	0.1771	1					
HNSO	0.3300	0.2833	0.0138	0.1411	0.1897	0.1766	0.0486	0.1463	0.0424	0.1472	0.1712	0.1142	0.3499	1				
HNFE	0.1663	0.2119	0.0397	0.0408	0.1871	0.1910	-0.0449	-0.0090	0.2208	0.2192	0.2013	0.0893	0.2097	0.1553	1			
HNWL	0.1396	0.2028	0.0321	0.2141	0.2217	0.2050	0.2831	0.2235	0.1254	0.0830	0.1358	0.0752	-0.0016	0.1477	0.0698	1		
HNWG	0.2649	0.2804	0.1005	0.2670	0.2502	0.2000	0.1351	0.1036	0.1670	0.1787	0.2295	0.1590	0.0709	0.0522	0.0234	0.2430	1	
	-0.0873	-0.0376	-0.1515	-0.0142	-0.1032	0.0194	0.0616	-0.0413	-0.0439	-0.0123	-0.0131	0.0627	-0.0751	0.0342	0.0206	0.0208	-0.1809	1

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems; HNSW: pain; HNSX: swallowing; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNNS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNP: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain.

Tables A4.2-A4.3 in the appendix show the full regression results for EQ-5D-5L utilities predicted through QLQ-C30 and H&N35 scores using linear mixed-effects and random-effects Tobit models. The two regression techniques are broadly similar in terms of significant variables, although Tobit has a much poorer goodness-of-fit in terms of AIC/BIC. Their accuracy prediction in terms of MAE/RMSE is instead almost the same. Therefore, in tables 4.5-4.6, only results from the best performing linear-mixed models are reported; with all tariff sets, reduced models perform better than full models. In models using QLQ-C30 scores (Table 4.5), physical functioning is always positive and highly statistically significant ($p < 0.001$), and emotional functioning is significant in some models only. Among the symptoms scales, the financial difficulties score has a significant negative impact on EQ-5D-5L utility, as do nausea and vomiting, pain, and diarrhoea, although with slight differences across the value sets adopted; constipation shows a counterintuitive positive coefficient. The global health score is positive and statistically significant in all models. In models using H&N35 scores (Table 4.6), pain, trouble with social contact, felt ill and, occasionally, opening mouth, weight loss, weight gain and female gender have a significant impact on EQ-5D-5L utility. Age does not show any significant association and is thus removed from all models. Figures 4.2-4.3 show the plots of predicted versus observed values for all country tariff sets and QLQ-C30/-H&N35 models, respectively. The diagonal line shows the line of perfect correlation; the vertical distance between the points and the line represents the error between observed and predicted utility values. In both scatter plot sets, low EQ-5D-5L values tend to be overestimated, while models underestimate utility in healthier states, and consistently fail to predict the maximum value of 1 (three predictions only are equal to 1, compared to 40 observed values); thus, the min-max range of predicted values is smaller compared to the observed ones. No predicted values fall outside the theoretical EQ-5D-5L range. For all the tariff sets adopted, the

algorithms using QLQ-C30 scales/items fit better than those regressing H&N35. Moreover, in both QLQ-C30 and H&N35 models, a better goodness-of-fit is observed for the Canadian, Uruguayan, and Korean value sets.

Additionally, ALDVM models (Tables A4.4-A4.5) are separately run for QLQ-C30 and H&N35 scores, by referring to the characteristics of each EQ-5D-5L value set distribution and using the BIC as a criterion for identifying the best number of components [235]; due to the small sample, models with more than three components are not tested [234]. A synthesis of the best performing models is reported in tables 4.7-4.8. In terms of goodness-of-fit, ALDVM models with one component only perform worse than Tobit models and, accordingly, are not considered for model comparison. Conversely, the 2- and 3- component models are those generally presenting the lowest AIC/BIC across all model types, although BIC reflects a penalty for model complexity given the small size of the database [215]. Predictive errors (MAE/RMSE) resulting from ALDVMM are comparable with those obtained from the other two regression techniques. The number and type of significant covariates vary considerably in these models, which yield a complex pattern of relationships between the EQ-5D-5L utility and QLQ-C30/-H&N35 scales/items. A better goodness-of-fit is obtained by using Canadian and Dutch value sets in QLQ-C30 models, and English, Canadian, and Uruguayan ones in H&N35 models, for which data allow the application of multi-component models. As observed for linear-mixed models, also the ALDVMM tends to overpredict EQ-5D-5L for the poorest states and underpredict the highest values; however, the dots appear closer to the diagonal line, especially on the left side of the utility scale. As above, models using QLQ-C30 fit better than H&N35 ones (Figures 4.4-4.5).

Table 4.5 Synthesis of best performing linear mixed-effects models using QLQ-C30.

	England			Netherlands			Canada			Uruguay			South Korea			Japan			China		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	p	Coef.	SE	p
Intercept	0.4050	0.0751	0.000	0.2859	0.0929	0.002	0.5628	0.0555	0.000	0.7288	0.0399	0.000	0.5110	0.0466	0.000	0.3203	0.0538	0.000	0.3730	0.0753	0.000
GH	0.0011	0.0004	0.012	0.0014	0.0005	0.005	0.0009	0.0003	0.002	0.0007	0.0002	0.003	0.0012	0.0004	0.001	0.0014	0.0004	0.000	0.0016	0.0004	0.000
PF	0.0034	0.0008	0.000	0.0037	0.0009	0.000	0.0028	0.0007	0.000	0.0018	0.0005	0.000	0.0030	0.0006	0.000	0.0035	0.0006	0.000	0.0045	0.0009	0.000
RF																					
EF	0.0012	0.0005	0.022	0.0016	0.0007	0.022										0.0013	0.0004	0.002			
CF																					
SF																					
FA																					
NV	-0.0027	0.0012	0.022	-0.0030	0.0013	0.028							-0.0018	0.0009	0.041	-0.0021	0.0010	0.038			
PA							-0.0011	0.0004	0.003	-0.0008	0.0003	0.013							-0.0016	0.0005	0.004
DY																					
SL																					
AP																					
CO	0.0007	0.0002	0.006	0.0008	0.0003	0.010	0.0006	0.0002	0.005	0.0005	0.0002	0.005							0.0007	0.0003	0.012
DI							-0.0020	0.0009	0.026	-0.0013	0.0006	0.030							-0.0019	0.0008	0.012
FI	-0.0008	0.0003	0.002	-0.0011	0.0003	0.001	-0.0009	0.0002	0.000	-0.0007	0.0002	0.000	-0.0009	0.0002	0.000	-0.0008	0.0002	0.001	-0.0012	0.0003	0.000
Female																					
Age																					
<i>Goodness-of-fit statistics</i>																					
AIC	-413.60			-323.78			-473.48			-590.39			-459.93			-431.40			-336.13		
BIC	-382.90			-293.07			-442.77			-559.69			-436.05			-404.11			-305.43		
MAE	0.0698			0.0883			0.0653			0.0498			0.0671			0.0745			0.0923		
RMSE	0.1012			0.1256			0.0997			0.0738			0.0895			0.0949			0.1296		
Mean	0.843			0.790			0.824			0.910			0.813			0.792			0.813		
SD	0.104			0.123			0.093			0.065			0.083			0.113			0.137		
Min	0.401			0.287			0.447			0.654			0.480			0.350			0.279		
Max	1.000			0.970			0.955			1.006			0.928			0.951			0.994		
Below 0	0; 0%			0; 0%			0; 0%			0; 0%			0; 0%			0; 0%			0; 0%		
Ceiling 1	1; 0.4%			0; 0%			0; 0%			2; 0.9%			0; 0%			0; 0%			0; 0%		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SD: standard deviation; SE: standard error.

Table 4.6 Synthesis of best performing linear mixed-effects models using QLQ-H&N35.

	England			Netherlands			Canada			Uruguay			South Korea			Japan			China		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	p	Coef.	SE	P	Coef.	SE	P	Coef.	SE	p	Coef.	SE	p
Intercept	0.9200	0.0118	0.000	0.8818	0.0141	0.000	0.8908	0.0097	0.000	0.9575	0.0081	0.000	0.8721	0.0118	0.000	0.8735	0.0160	0.000	0.9106	0.0182	0.000
HNPA	-0.0020	0.0005	0.000	-0.0024	0.0006	0.000	-0.0017	0.0004	0.000	-0.0011	0.0003	0.001	-0.0017	0.0004	0.000	-0.0018	0.0005	0.000	-0.0023	0.0007	0.000
HNSW																					
HNSE																					
HNSP																					
HNSO																					
HNSC	-0.0026	0.0006	0.000	-0.0031	0.0007	0.000	-0.0023	0.0006	0.000	-0.0018	0.0004	0.000	-0.0022	0.0005	0.000	-0.0025	0.0005	0.000	-0.0034	0.0008	0.000
HNSX																					
HNTE																					
HNOM																-0.0009	0.0003	0.001	-0.0009	0.0004	0.018
HNDR																					
HNSS																					
HNCO																					
HNFI	-0.0014	0.0005	0.004	-0.0018	0.0006	0.003	-0.0013	0.0004	0.002	-0.0010	0.0003	0.003	-0.0013	0.0004	0.002	-0.0012	0.0005	0.007	-0.0016	0.0006	0.006
HNPk																					
HNNU																					
HNFE																					
HNWL																0.0004	0.0002	0.026	0.0005	0.0002	0.026
HNWG																0.0003	0.0001	0.023	0.0004	0.0002	0.028
Female													0.0437	0.0209	0.036						
Age																					
<i>Goodness-of-fit statistics</i>																					
AIC	-377.11			-292.37			-435.47			-565.08			-431.10			-377.78			-294.74		
BIC	-356.59			-271.85			-414.95			-544.55			-407.16			-346.99			-263.96		
MAE	0.0847			0.1031			0.0776			0.0560			0.0754			0.0973			0.1150		
RMSE	0.1252			0.1520			0.1201			0.0848			0.1048			0.1253			0.1611		
Mean	0.845			0.792			0.825			0.910			0.817			0.795			0.814		
SD	0.074			0.089			0.066			0.048			0.063			0.079			0.100		
Min	0.578			0.463			0.580			0.732			0.572			0.537			0.486		
Max	0.920			0.882			0.891			0.957			0.916			0.909			0.960		
Below 0	0; 0%			0; 0%			0; 0%			0; 0%			0; 0%			0; 0%			0; 0%		
Ceiling 1	0; 0%			0; 0%			0; 0%			0; 0%			0; 0%			0; 0%			0; 0%		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SD: standard deviation; SE: standard error.

Figure 4.2 Scatterplots of observed versus predicted EQ-5D-5L for linear-mixed models using QLQ-C30.

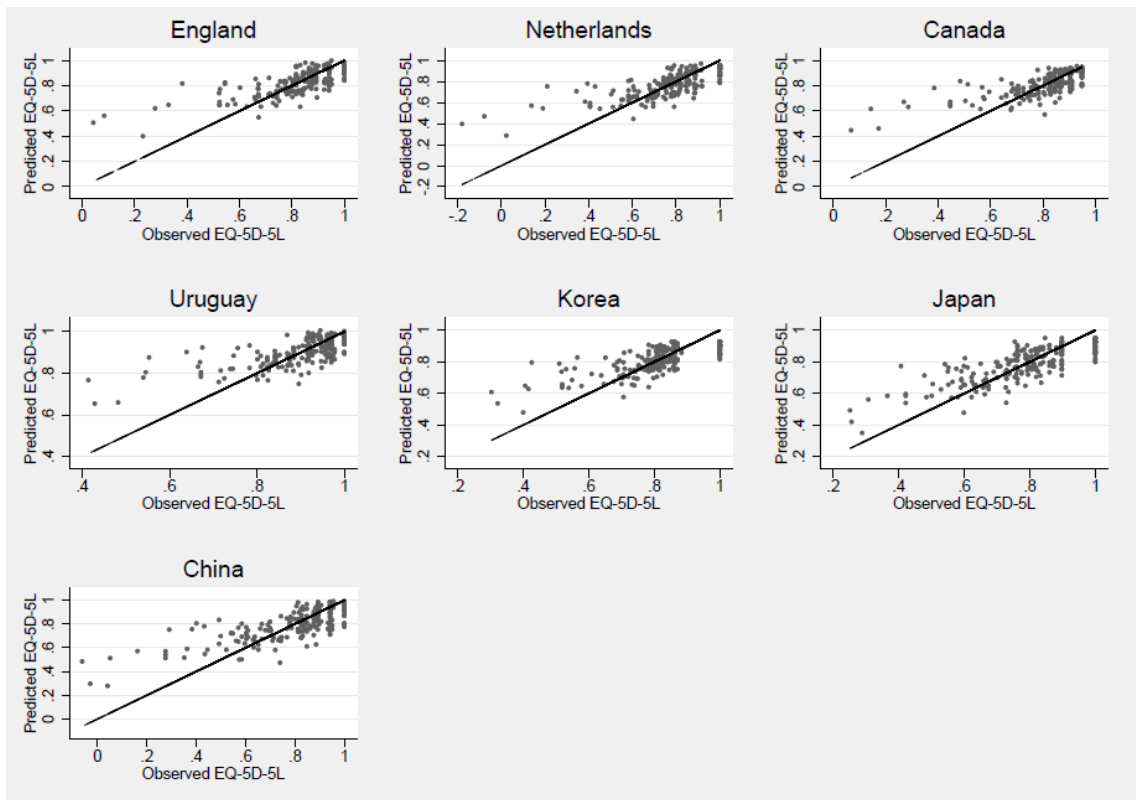


Figure 4.3 Scatterplots of observed versus predicted EQ-5D-5L for linear-mixed models using QLQ-H&N35.

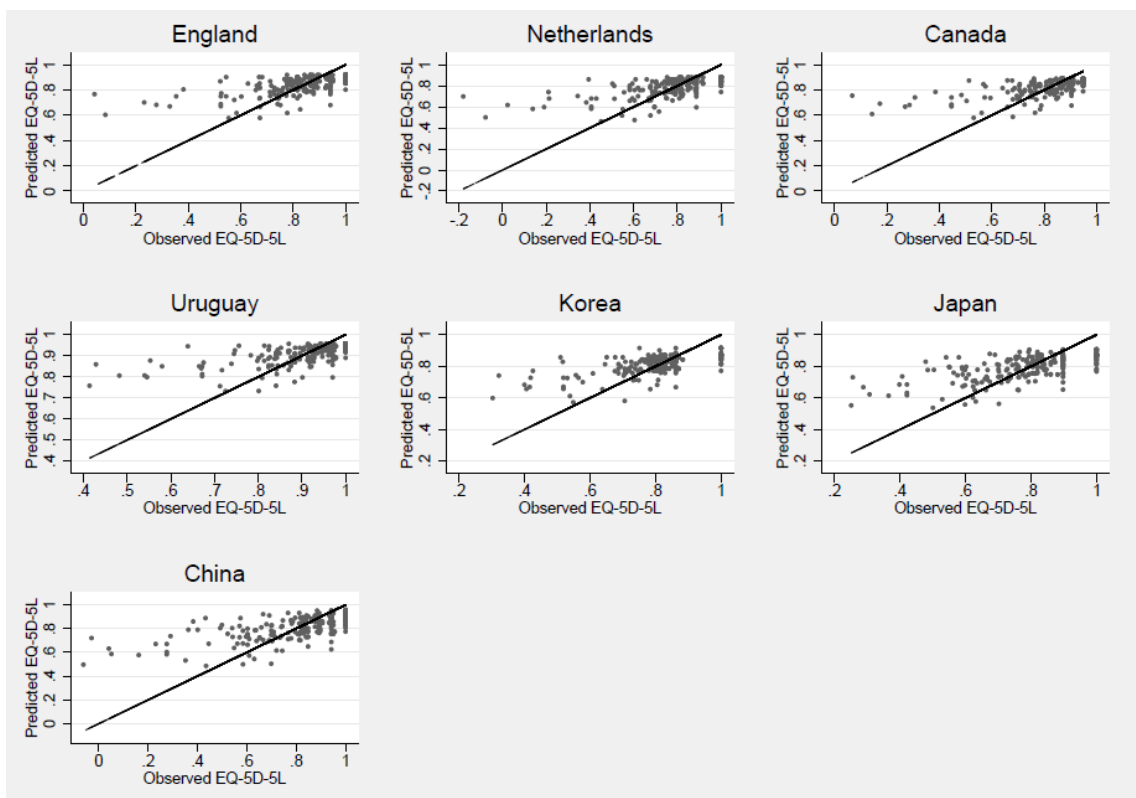


Table 4.7 Synthesis of best performing ALDVM models using QLQ-C30.

	England						Netherlands								
	Component 1			Component 2			Component 1			Component 2			Component 3		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	p	Coef.	SE	P
Intercept	0.8685	0.0039	0.000	0.3562	0.1514	0.019	0.9311	0.0023	0.000	0.8319	0.0553	0.000	-0.5478	0.3278	0.095
GH	0.0010	<0.0001	0.000	0.0010	0.0007	0.144	0.0003	<0.0001	0.000	0.0009	0.0007	0.168	0.0034	0.0015	0.019
PF	-0.0001	<0.0001	0.000	0.0042	0.0009	0.000	0.0025	<0.0001	0.000	-0.0008	0.0007	0.238	0.0087	0.0014	0.000
RF	-0.0005	<0.0001	0.000	0.0012	0.0007	0.082	0.0002	<0.0001	0.000	-0.0002	0.0001	0.244	0.0022	0.0010	0.029
EF	0.0007	<0.0001	0.000	0.0016	0.0005	0.003	-0.0011	<0.0001	0.000	0.0013	0.0003	0.000	0.0054	0.0016	0.001
CF	0.0006	<0.0001	0.000	-0.0007	0.0007	0.290	-0.0008	<0.0001	0.000	0.0005	0.0004	0.148	-0.0010	0.0015	0.514
SF	-0.0003	<0.0001	0.000	0.0003	0.0005	0.551	0.0008	<0.0001	0.000	-0.0008	0.0002	0.000	-0.0014	0.0013	0.279
FA	-0.0013	<0.0001	0.000	0.0005	0.0007	0.541	-0.0049	<0.0001	0.000	0.0002	0.0004	0.594	0.0036	0.0015	0.018
NV	0.0005	<0.0001	0.000	-0.0037	0.0018	0.034	0.0017	<0.0001	0.000	-0.0009	0.0003	0.001	-0.0117	0.0022	0.000
PA	-0.0008	<0.0001	0.000	-0.0013	0.0006	0.044	-0.0017	<0.0001	0.000	-0.0003	0.0003	0.219	-0.0012	0.0009	0.200
DY	-0.0011	<0.0001	0.000	0.0003	0.0005	0.552	-0.0034	<0.0001	0.000	-0.0007	0.0003	0.011	0.0034	0.0013	0.009
SL	0.0004	<0.0001	0.000	0.0003	0.0004	0.487	0.0006	<0.0001	0.000	-0.0013	0.0003	0.000	0.0016	0.0010	0.104
AP	-0.0002	<0.0001	0.000	0.0002	0.0005	0.701	-0.0004	<0.0001	0.000	0.0006	0.0002	0.015	0.0008	0.0011	0.469
CO	0.0013	<0.0001	0.000	-0.0004	0.0003	0.271	0.0001	<0.0001	0.000	0.0002	0.0002	0.325	-0.0003	0.0008	0.669
DI	0.0012	<0.0001	0.000	-0.0022	0.0009	0.018	-0.0040	<0.0001	0.000	-0.0002	0.0004	0.553	-0.0026	0.0018	0.158
FI	-0.0011	<0.0001	0.000	-0.0009	0.0004	0.017	0.0005	<0.0001	0.000	-0.0003	0.0003	0.363	-0.0028	0.0009	0.003
Female	0.0398	0.0007	0.000	0.0477	0.0217	0.028	0.0313	0.0004	0.000	-0.0090	0.0166	0.585	0.0634	0.0501	0.205
Age	-0.0012	<0.0001	0.000	-0.0013	0.0010	0.211	-0.0013	<0.0001	0.000	-0.0005	0.0008	0.563	-0.0014	0.0021	0.504
Probability (constant)	-1.6222	0.2222	0.000				-1.4306	0.3064	0.000	-0.2280	0.3213	0.478			
/lns	-7.0701	0.1472	0.000	-2.2525	0.0872	0.000	-9.9034	0.4471	0.000	-3.7173	0.2172	0.000	-2.0715	0.1109	0.000
Sigma	0.0008	0.0001		0.1051	0.0092		<0.0001	<0.0001		0.0243	0.0053		0.1260	0.0140	
Probability	0.1649	0.0306		0.8351	0.0306		0.1175	0.0256		0.3911	0.0689		0.4913	0.0738	
<i>Goodness-of-fit statistics</i>															
AIC	-341.64						-383.64								
BIC	-208.59						-192.59								
MAE	0.0666						0.0872								
RMSE	0.0943						0.1196								
Mean	0.837						0.794								
SD	0.128						0.143								
Min	0.260						0.161								
Max	0.995						0.963								
Below 0	0; 0%						0; 0%								
Ceiling 1	0; 0%						0; 0%								

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SD: standard deviation; SE: standard error.

Table 4.7 (cont.) Synthesis of best performing ALDVM models using QLQ-C30.

	Canada						China								
	Component 1			Component 2			Component 1			Component 2			Component 3		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	p	Coef.	SE	P
Intercept	0.7767	0.0412	0.000	0.3011	0.1806	0.095	0.7636	0.0480	0.000	-0.2102	0.4200	0.617	1.0463	0.0032	0.000
GH	0.0007	0.0003	0.043	0.0001	0.0005	0.789	0.0018	0.0004	0.000	0.0013	0.0010	0.201	<0.0001	<0.0001	0.037
PF	0.0010	0.0002	0.000	0.0035	0.0011	0.001	0.0014	0.0005	0.008	0.0079	0.0021	0.000	<0.0001	<0.0001	0.657
RF	-0.0001	0.0001	0.252	0.0029	0.0008	0.000	0.0000	0.0001	0.517	0.0018	0.0009	0.039	0.0001	<0.0001	0.000
EF	0.0014	0.0002	0.000	0.0005	0.0006	0.422	0.0009	0.0004	0.026	0.0022	0.0009	0.018	0.0001	<0.0001	0.000
CF	-0.0003	0.0001	0.018	0.0006	0.0006	0.355	<0.0001	0.0001	0.939	0.0009	0.0010	0.396	-0.0002	<0.0001	0.000
SF	-0.0005	0.0002	0.024	-0.0006	0.0005	0.200	-0.0009	0.0001	0.000	-0.0007	0.0006	0.219	0.0003	<0.0001	0.000
FA	0.0013	0.0002	0.000	0.0001	0.0008	0.926	0.0010	0.0002	0.000	<0.0001	0.0014	0.977	-0.0005	<0.0001	0.000
NV	-0.0012	0.0002	0.000	-0.0041	0.0014	0.003	-0.0030	0.0003	0.000	-0.0069	0.0027	0.010	-0.0052	<0.0001	0.000
PA	-0.0012	0.0001	0.000	-0.0017	0.0008	0.028	-0.0015	0.0004	0.000	-0.0011	0.0014	0.401	-0.0060	<0.0001	0.000
DY	0.0007	0.0002	0.000	0.0004	0.0006	0.478	0.0019	0.0003	0.000	0.0011	0.0012	0.358	-0.0026	<0.0001	0.000
SL	-0.0008	0.0002	0.000	0.0002	0.0003	0.507	-0.0021	0.0003	0.000	0.0008	0.0006	0.213	-0.0001	<0.0001	0.000
AP	0.0005	0.0001	0.000	0.0009	0.0005	0.071	0.0011	0.0001	0.000	0.0017	0.0008	0.035	-0.0008	<0.0001	0.000
CO	-0.0004	0.0001	0.000	0.0005	0.0003	0.115	-0.0008	0.0001	0.000	0.0005	0.0005	0.360	-0.0007	<0.0001	0.000
DI	-0.0012	0.0002	0.000	-0.0016	0.0010	0.099	-0.0017	0.0002	0.000	-0.0004	0.0011	0.732	-0.0043	<0.0001	0.000
FI	-0.0003	0.0002	0.212	-0.0006	0.0004	0.174	-0.0001	0.0001	0.642	-0.0015	0.0005	0.001	-0.0002	<0.0001	0.000
Female	-0.0188	0.0054	0.001	0.0453	0.0210	0.031	0.0086	0.0213	0.687	0.0369	0.0301	0.220	-0.1055	0.0005	0.000
Age	-0.0006	0.0005	0.215	-0.0010	0.0010	0.330	-0.0010	0.0006	0.120	-0.0014	0.0017	0.435	-0.0002	<0.0001	0.000
Probability (constant)	-0.4538	0.3152	0.150				0.4192	0.3344	0.210	1.4316	0.2279	0.000			
/lns	-4.3653	0.4044	0.000	-2.5282	0.1004	0.000	-4.4991	0.4309	0.000	-2.2115	0.0974	0.000	-7.5695	0.1339	0.000
Sigma	0.0127	0.0051		0.0798	0.0080		0.0111	0.0048		0.1095	0.0107		0.0005	0.0001	
Probability	0.3884	0.0749		0.6115	0.0749		0.2268	0.0561		0.6241	0.0575		0.1491	0.0273	
<i>Goodness-of-fit statistics</i>															
AIC	-547.84						-313.69								
BIC	-414.78						-112.40								
MAE	0.0624						0.0856								
RMSE	0.0921						0.1187								
Mean	0.827						0.800								
SD	0.107						0.169								
Min	0.380						0.094								
Max	0.973						0.989								
Below 0	0; 0%						0; 0%								
Ceiling 1	0; 0%						0; 0%								

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SD: standard deviation; SE: standard error.

Table 4.8 Synthesis of best performing ALDVM models using QLQ-H&N35.

	England									Canada								
	Component 1			Component 2			Component 3			Component 1			Component 2			Component 3		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	p	Coef.	SE	P
Intercept	1.0587	0.0548	0.000	0.5542	0.0013	0.000	0.9462	0.0019	0.000	0.8644	0.0200	0.000	0.9904	0.0171	0.000	0.5233	0.0915	0.000
HNPA	-0.0038	0.0007	0.000	-0.0009	<0.0001	0.000	-0.0001	<0.0001	0.000	-0.0003	0.0003	0.282	-0.0024	0.0003	0.000	-0.0065	0.0009	0.000
HNSW	-0.0004	0.0006	0.488	-0.0038	<0.0001	0.000	0.0001	<0.0001	0.000	-0.0006	0.0006	0.352	<0.0001	0.0002	0.797	0.0006	0.0005	0.205
HNSE	0.0002	0.0004	0.570	0.0024	<0.0001	0.000	-0.0001	<0.0001	0.000	-0.0009	0.0002	0.000	<0.0001	0.0001	0.783	0.0029	0.0004	0.000
HNSP	-0.0002	0.0005	0.667	0.0017	<0.0001	0.000	0.0006	<0.0001	0.000	0.0018	0.0003	0.000	-0.0005	0.0002	0.023	0.0009	0.0010	0.359
HNSO	0.0003	0.0006	0.616	0.0072	<0.0001	0.000	0.0010	<0.0001	0.000	0.0012	0.0002	0.000	-0.0003	0.0003	0.291	0.0046	0.0011	0.000
HNSC	-0.0035	0.0009	0.000	-0.0024	<0.0001	0.000	-0.0023	<0.0001	0.000	-0.0064	0.0004	0.000	0.0009	0.0003	0.007	-0.0028	0.0003	0.000
HNSX	0.0009	0.0003	0.002	-0.0017	<0.0001	0.000	-0.0003	<0.0001	0.000	0.0003	0.0001	0.021	0.0004	0.0001	0.002	-0.0013	0.0005	0.013
HNTE	<0.0001	0.0003	0.986	-0.0018	<0.0001	0.000	0.0007	<0.0001	0.000	-0.0005	0.0002	0.003	<0.0001	0.0001	0.861	0.0009	0.0004	0.021
HNOM	-0.0005	0.0003	0.121	0.0010	<0.0001	0.000	-0.0008	<0.0001	0.000	-0.0004	0.0002	0.047	-0.0003	0.0001	0.018	0.0010	0.0005	0.040
HNDR	-0.0002	0.0003	0.564	0.0033	<0.0001	0.000	-0.0003	<0.0001	0.000	-0.0001	0.0003	0.844	-0.0001	0.0001	0.386	0.0017	0.0009	0.067
HNSS	<0.0001	0.0003	0.902	-0.0016	<0.0001	0.000	-0.0006	<0.0001	0.000	0.0003	0.0002	0.106	-0.0005	0.0002	0.004	0.0005	0.0003	0.089
HNCO	-0.0008	0.0004	0.079	-0.0068	<0.0001	0.000	-0.0010	<0.0001	0.000	0.0002	0.0002	0.321	-0.0009	0.0002	0.000	-0.0039	0.0005	0.000
HNFI	-0.0012	0.0007	0.083	-0.0049	<0.0001	0.000	-0.0020	<0.0001	0.000	-0.0119	0.0006	0.000	0.0006	0.0004	0.124	-0.0002	0.0003	0.579
HNPk	-0.0005	0.0002	0.019	-0.0011	<0.0001	0.000	0.0005	<0.0001	0.000	-0.0005	0.0001	0.001	-0.0002	0.0001	0.003	-0.0006	0.0002	0.004
HNNU	-0.0002	0.0002	0.242	-0.0037	<0.0001	0.000	0.0003	<0.0001	0.000	<0.0001	0.0001	0.828	<0.0001	0.0001	0.592	-0.0008	0.0002	0.000
HNFE	-0.0003	0.0003	0.449	-0.0019	<0.0001	0.000	-0.0002	<0.0001	0.000	0.0009	0.0003	0.000	-0.0012	0.0001	0.000	0.0026	0.0003	0.000
HNWL	0.0005	0.0003	0.054	0.0027	<0.0001	0.000	<0.0001	<0.0001	0.000	-0.0004	0.0001	0.000	0.0004	0.0001	0.000	0.0007	0.0005	0.159
HNWG	0.0004	0.0001	0.006	0.0001	<0.0001	0.000	-0.0006	<0.0001	0.000	-0.0002	0.0001	0.083	0.0001	0.0001	0.050	0.0013	0.0003	0.000
Female	0.0407	0.0228	0.075	0.2872	0.0004	0.000	-0.0394	0.0002	0.000	0.0489	0.0113	0.000	-0.0060	0.0074	0.413	0.1885	0.0374	0.000
Age	-0.0012	0.0008	0.142	0.0020	<0.0001	0.000	-0.0005	<0.0001	0.000	0.0008	0.0003	0.010	-0.0008	0.0003	0.007	0.0002	0.0007	0.727
Probability (constant)	1.8421	0.2146	0.000	-0.0618	0.3030	0.838				0.7863	0.2810	0.005	0.9643	0.2808	0.001			
/lns	-2.4564	0.0695	0.000	-8.2578	0.1251	0.000	-10.0190	0.3320	0.000	-3.5098	0.1262	0.000	-3.8609	0.0964	0.000	-3.5452	0.1148	0.000
Sigma	0.0857	0.0060		0.0003	<0.0001		<0.0001	<0.0001		0.0299	0.0038		0.0210	0.0020		0.0289	0.0033	
Probability	0.7648	0.0295		0.1139	0.0233		0.1212	0.0228		0.3773	0.0515		0.4508	0.0540		0.1719	0.0361	
<i>Goodness-of-fit statistics</i>																		
AIC	-566.40									-535.52								
BIC	-365.64									-304.14								
MAE	0.0834									0.0758								
RMSE	0.1133									0.1097								
Mean	0.831									0.814								
SD	0.109									0.095								
Min	0.368									0.389								
Max	0.977									0.959								
Below 0	0; 0%									0; 0%								
Ceiling 1	0; 0%									0; 0%								

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SD: standard deviation; SE: standard error.

Table 4.8 (cont.) Synthesis of best performing ALDVM models using QLQ-H&N35.

	Uruguay									China					
	Component 1			Component 2			Component 3			Component 1			Component 2		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	p	Coef.	SE	P
Intercept	1.0471	0.0282	0.000	0.8676	0.1005	0.000	0.9980	<0.0001	0.000	0.9966	0.0024	0.000	1.0846	0.1082	0.000
HNPA	-0.0010	0.0003	0.000	-0.0031	0.0012	0.010	-0.0012	<0.0001	0.000	-0.0021	<0.0001	0.000	-0.0035	0.0011	0.002
HNSW	0.0001	0.0002	0.718	0.0036	0.0008	0.000	0.0005	<0.0001	0.000	0.0019	<0.0001	0.000	0.0003	0.0012	0.778
HNSE	-0.0003	0.0002	0.095	0.0053	0.0006	0.000	-0.0004	<0.0001	0.000	-0.0003	<0.0001	0.000	0.0010	0.0006	0.128
HNSP	0.0002	0.0002	0.390	-0.0067	0.0012	0.000	-0.0005	<0.0001	0.000	0.0013	<0.0001	0.000	-0.0009	0.0009	0.320
HNSO	0.0005	0.0003	0.067	0.0014	0.0006	0.018	-0.0002	<0.0001	0.000	<0.0001	<0.0001	0.925	0.0025	0.0011	0.023
HNSC	-0.0008	0.0005	0.102	0.0015	0.0008	0.069	-0.0012	<0.0001	0.000	-0.0012	<0.0001	0.000	-0.0075	0.0016	0.000
HNSX	<0.0001	0.0001	0.898	-0.0015	0.0004	0.000	-0.0001	<0.0001	0.000	-0.0008	<0.0001	0.000	-0.0003	0.0007	0.610
HNTE	0.0001	0.0001	0.142	-0.0019	0.0006	0.002	-0.0001	<0.0001	0.000	0.0001	<0.0001	0.000	-0.0001	0.0004	0.798
HNOM	-0.0008	0.0001	0.000	0.0007	0.0006	0.207	0.0001	<0.0001	0.000	-0.0020	<0.0001	0.000	-0.0010	0.0005	0.040
HNDR	0.0001	0.0001	0.668	0.0028	0.0008	0.000	-0.0001	<0.0001	0.000	0.0003	<0.0001	0.000	-0.0007	0.0004	0.085
HNSS	-0.0002	0.0002	0.200	-0.0036	0.0009	0.000	-0.0002	<0.0001	0.000	-0.0020	<0.0001	0.000	0.0006	0.0005	0.213
HNCO	-0.0003	0.0002	0.153	-0.0006	0.0006	0.354	-0.0002	<0.0001	0.000	-0.0006	<0.0001	0.000	-0.0013	0.0008	0.091
HNFI	-0.0009	0.0002	0.000	0.0010	0.0008	0.231	-0.0011	<0.0001	0.000	-0.0178	<0.0001	0.000	-0.0015	0.0009	0.107
HNPk	-0.0002	0.0001	0.076	-0.0033	0.0005	0.000	-0.0004	<0.0001	0.000	0.0001	<0.0001	0.000	-0.0010	0.0003	0.003
HNNU	-0.0001	0.0001	0.403	-0.0009	0.0003	0.003	-0.0009	<0.0001	0.000	-0.0001	<0.0001	0.000	-0.0005	0.0004	0.169
HNFE	-0.0004	0.0002	0.127	0.0056	0.0013	0.000	-0.0011	<0.0001	0.000	-0.0019	<0.0001	0.000	0.0006	0.0006	0.280
HNWL	<0.0001	0.0001	0.886	-0.0035	0.0008	0.000	0.0001	<0.0001	0.000	-0.0006	<0.0001	0.000	0.0013	0.0004	0.001
HNWG	0.0001	0.0001	0.063	-0.0012	0.0004	0.006	0.0003	<0.0001	0.000	-0.0004	<0.0001	0.000	0.0012	0.0003	0.000
Female	-0.0045	0.0106	0.671	0.1739	0.0324	0.000	0.0398	0.0001	0.000	0.0291	0.0008	0.000	0.0799	0.0356	0.025
Age	-0.0008	0.0004	0.043	0.0015	0.0021	0.468	-0.0001	<0.0001	0.000	0.0007	<0.0001	0.000	-0.0022	0.0016	0.185
Probability (constant)	1.7280	0.2565	0.000	0.5257	0.3090	0.089				-1.3532	0.1775	0.000			
/lns	-3.3359	0.0801	0.000	-3.0249	0.1533	0.000	-9.3988	0.2156	0.000	-6.5865	0.1233	0.000	-1.8746	0.0788	0.000
Sigma	0.0356	0.0028		0.0486	0.0074		0.0001	<0.0001		0.0014	0.0002		0.1534	0.0121	
Probability	0.6765	0.0592		0.2033	0.0517		0.1202	0.0242		0.2053	0.0290		0.7946	0.0290	
<i>Goodness-of-fit statistics</i>															
AIC	-504.67									-203.65					
BIC	-283.49									-50.53					
MAE	0.0689									0.1022					
RMSE	0.1167									0.1375					
Mean	0.892									0.796					
SD	0.073									0.162					
Min	0.640									0.151					
Max	0.993									0.985					
Below 0	0; 0%									0; 0%					
Ceiling 1	0; 0%									0; 0%					

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SD: standard deviation; SE: standard error.

Figure 4.4 Scatterplots of observed versus predicted EQ-5D-5L for ALDVM models using QLQ-C30.

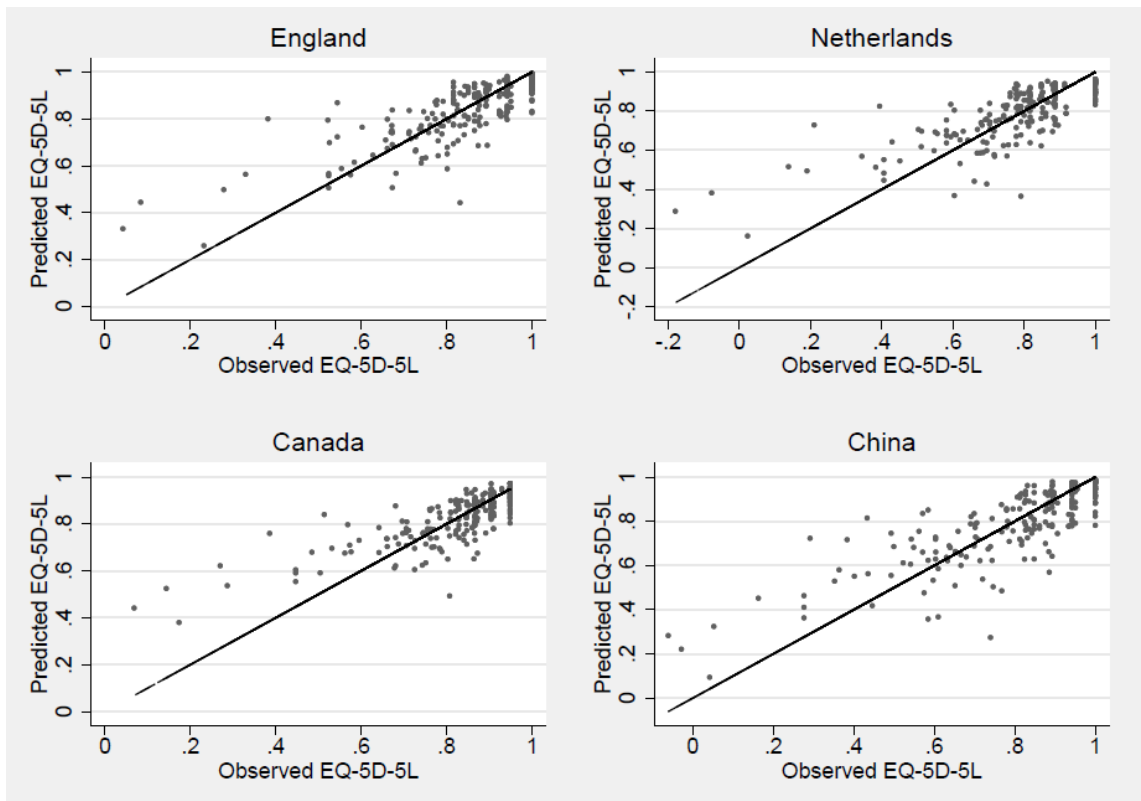
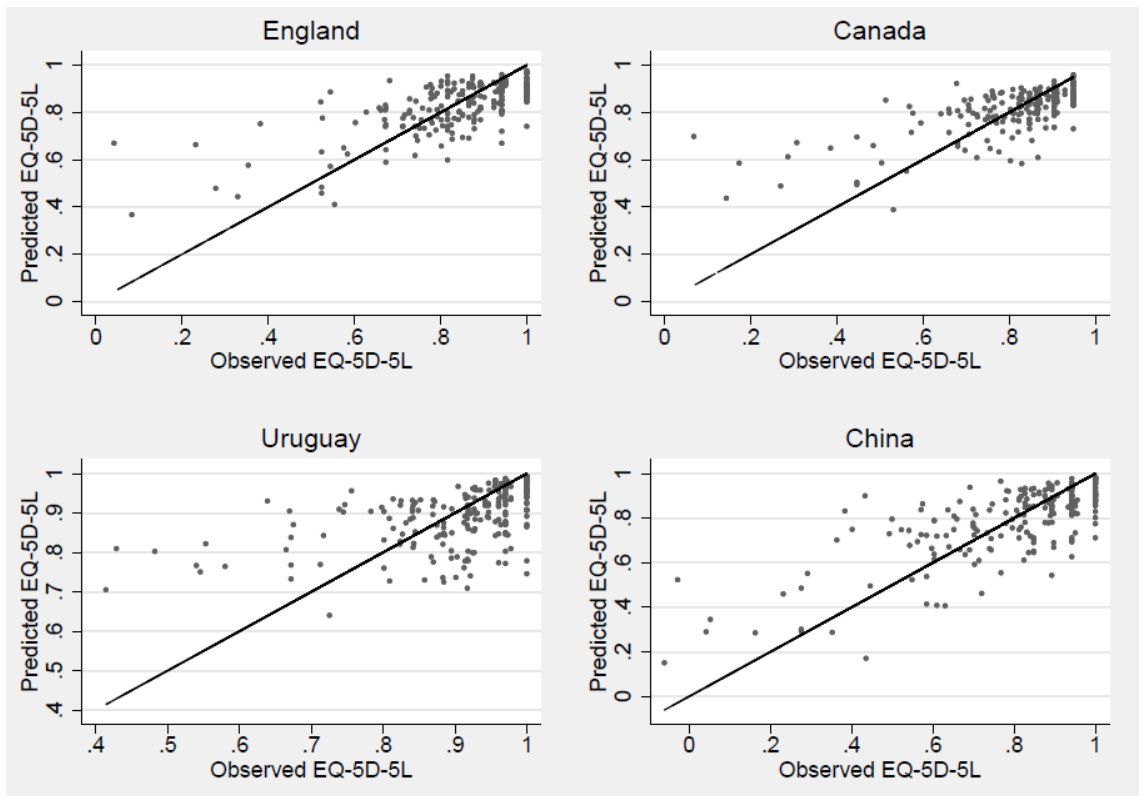


Figure 4.5 Scatterplots of observed versus predicted EQ-5D-5L for ALDVM models using QLQ-H&N35.



4.4 Discussion

This study mapped QLQ-C30 and H&N35 scales/items to EQ-5D-5L utilities in HNC patients, using data collected within the ongoing HETeCo trial. This is the first mapping study linking EQ-5D (either -3L or -5L) to QLQ-C30 (and H&N35) in HNC. The HERC database (version 6.0, last updated in January 2017) [165] [238] [239] currently includes 144 mapping studies that predict EQ-5D utilities from any disease-specific tool. Among them, 22 are related to cancer and 15 map from QLQ-C30. Moreover, PubMed was also searched using as keywords “mapping”, “EORTC”, “EQ-5D” and “cancer” to avoid missing any recent publication. In total, 15 articles mapping QLQ-C30 onto EQ-5D in cancer were identified. Among them, only the most recent [226] predicts utility scores for the 5-level version, and none addresses HNC. However, as described in Chapter III, one study [195] maps the EQ-5D-3L from the UWQOL Questionnaire (UW QOL v4) collected from a cohort of HNC patients shortly after treatment using OLS regression with forward stepwise selection ($p < 0.05$).

In this chapter, three different techniques which are being favoured by the most recent mapping literature [63] [220] [226] have been tested. Overall, the linear mixed-effects and the multi-component ALDVM models outperform the random-effects Tobit models in terms of AIC/BIC goodness-of-fit measures, and therefore are considered for more detailed analyses; moreover, as observed in a previous study on breast cancer [237], mapping the general questionnaire (QLQ-C30) provides a better fit than using the cancer-specific module (H&N35) irrespective of the model’s specification.

In linear-mixed models, a few recurrent significant QLQ-C30 scores are observed: physical functioning, which is composed by five items largely corresponding to the first three EQ-5D dimensions (mobility, self-care, and usual activity), emotional functioning which can relate to the anxiety/depression dimension of EQ-5D, and pain, resembling

the pain/discomfort dimension. Global health status, which is a 2-item scale synthesizing the self-perceived patient's health and HRQoL, can be interpreted as a general metrics resembling EQ-5D utility. Other symptom scales/items such as nausea and vomiting or financial difficulties that show a significant impact of EQ-5D-5L utility are harder to interpret. Similarly, in models using H&N35, pain clearly corresponds with pain/discomfort, while trouble with social contacts and felt ill can be interpreted as parts of the anxiety/depression domain. It is not surprising instead that the coefficients of cognitive functioning, fatigue, and insomnia are never statistically significant across the models, since the use of EQ-5D in cancer has been recently questioned for not including dimensions related to 'cognition', 'sleep', and especially 'vitality', which has been found to be significant in determining HRQoL [38].

This study presents a few limitations. First, the sample size is small (97 patients, 225/226 paired observations in models using QLQ-C30/-H&N35), thus affecting the precision of the estimated coefficients, especially for the more complex mixture models, which were limited to three components as a maximum. In the recent systematic review by Dakin [238], only 15% of the algorithms mapping to EQ-5D had less than 200 observations, and an equivalent proportion between 200 and 500. Thus, a re-evaluation of these models using a larger sample at the end of the trial is scheduled. Additionally, the limited number of observations prevented splitting the database by QLQ-C30/-H&N35 levels, to verify how the goodness-of-fit varies according to disease severity [234] [237]. Second, some caution should be adopted in using these algorithms, since it is well known that RCTs tend to recruit healthier patients than those usually observed in routine practice [234]. Moreover, data used for this mapping are collected from patients who have successfully completed their curative treatment(s) for primary cancer and, thus, are likely to be healthier than the overall HNC patients' population; their HRQoL measures, indeed, are overall quite high. Thus, using these mapping coefficients might

lead to biased QALY estimates in cost-utility models assessing novel interventions for patients with other characteristics or at different treatment stages. Moreover, all models tend to over/under predict EQ-5D at the extremes of the distribution, thus potentially affecting QALY results for seriously ill and end-of-life patients and, conversely, less severe ones. Third, unexpected significant positive coefficients were obtained for constipation, weight loss and weight gain, which are symptom scales calculated from QLQ-C30 and H&N35 responses. However, none of the studies retrieved from the literature have perfectly intuitive signs for the full sets of independent variables. Such counterintuitive signs for coefficients could be attributable to unknown relationships between QLQ-C30/-H&N35 scales/items within the regression models [237]. Moreover, these specific symptoms might also be interpreted in a positive sense for oncological patients, who often suffer from lack of appetite, diarrhoea, and cachexia [240]. Fourth, additional EQ-5D-5L country value sets are being added to the literature, as documented by frequent updates of the EuroQol website; thus, the list of seven tariff sets explored in this study cannot be considered exhaustive. Similarly, the selected regression techniques do not cover the whole spectrum of models proposed by the most recent literature. For example, Khan et al. suggested a non-linear beta binomial regression as a promising alternative that has shown an improved fit compared to traditional censored regression (e.g. Tobit) in modelling EQ-5D-5L data. This technique, indeed, can easily model skewed and multimodal data bound on a 0 to 1 interval (although any range is suitable by applying an appropriate data transformation into 0-1) [216] [226]. Moreover, the small number of available EQ-5D-5L profiles in the study sample prevented the application of more advanced techniques such as indirect methods for mapping. Among them, response mapping fits separate ordinal (or multinomial) logistic regression models to each of the five EQ-5D dimensions; thereafter, the expected utility value is calculated analytically using coefficients from

any preference-based algorithm [215] [220], thus overcoming the issue of country-specific mapping functions that are frequently available in the literature. However, the criterion of at least 10 observations for the smallest category usually applies to response mapping [215] and is not satisfied by the current database, which contains far less than 10 observations for levels 4 and 5 in each EQ-5D dimension (except for the usual activities dimension, level 4).

4.5 Conclusions

The developed functions are useful to obtain utility values in HNC clinical studies not collecting preference-based HRQoL data. Of the 1,815 HNC studies identified on ClinicalTrials.gov (accessed in November 2017), which is a database of privately and publicly funded clinical studies conducted worldwide, only 16 reported EQ-5D as an outcome measure, while QLQ-C30 and H&N35 (together or as stand-alone questionnaires) were adopted by 53 clinical trials; thus, studies using EORTC questionnaires only could obtain HSUVs by applying the algorithms presented in this chapter in order to perform trial-based cost-utility analyses. Moreover, since QLQ-C30 data were modelled alone, this study can inform also cost-utility analyses in other cancer types; however, the generalizability of the developed functions to a different cancer dataset should be carefully evaluated since, even for the same cancer, oncological patients vary widely in age, gender, and disease severity [219]. Additionally, models are developed using alternative value sets for selected countries, thus future researchers can select the preferred one according to the geographical area where the study is conducted. According to the most recent guidelines [153] [220], re-estimation of mapping results in an alternative dataset or other forms of evaluation (e.g. cross-sample validation) are not routinely required, especially for studies using small databases. However, the assessment of the algorithms' external validity in a larger

sample of HNC data may be of interest. Overall, the ALDVMM appears as a promising technique to model EQ-5D-5L data for mapping purposes, although the limited size of the database prevented the application of multi-component models for the full set of country tariffs. Further research is encouraged on this topic, as new country algorithms are being developed to value EQ-5D-5L and made progressively available on the EuroQol website. An Italian value set is particularly warranted for this study, since data are collected from HNC patients in Italy.

5 THE USE OF INTENSIVE RADIOLOGICAL ASSESSMENTS IN ROUTINE SURVEILLANCE AFTER TREATMENT FOR HEAD AND NECK CANCER: AN ECONOMIC EVALUATION

5.1 Introduction

HNC is a major public issue worldwide that causes significant morbidity and mortality despite clinical advances in diagnosis and treatment [241]. As anticipated in Chapter I, in Europe alone, around 143,000 people are diagnosed and more than 68,000 die each year because of the disease [17]. The incidence in Italy is about 16 cases per 100,000, half of whom are aged between 50 and 70; the male-female ratio is around 6:1 [3] [6]. Despite the routine introduction of combined-modality treatment, the 5-year overall survival rate is 40% to 60%, varying by age, cancer site and disease stage [1] [6] [61]. When diagnosed at a locally advanced stage, HNC tends to recur, with up to 50% of patients developing loco-regional or metastatic recurrences in the first few years after treatment depending on site [10] [25] [61] [242] [243] [244]; additionally, a constant rate of 2-3% per year of second primaries is observed [21] [22].

A few patients with loco-regional recurrences or second primaries can be salvaged by a potentially curative treatment with a possibility of long-term survival but a significant chance of treatment-related toxicity and intra-operative mortality [16] [17] [61]. Surgery is traditionally considered the treatment of choice for patients with a resectable loco-regional recurrence or a second primary and sufficiently good health status. In the last decade, high-dose re-irradiation has also shown reasonable improvements in loco-regional control and overall survival although at the expense of high, potentially life-

threatening toxicities (e.g. carotid rupture, fistula, bleeding) [25] [245]. Unfortunately, most recurrent patients are only suitable for palliative treatment that, in addition to best supportive care, usually includes systemic treatment with a combination of chemotherapeutics [26] [246]. The prognosis for patients with recurrent or metastatic disease not eligible for surgery or re-irradiation is very poor, with a median overall survival of around 10 months under the standard scheme of platinum-based chemotherapy plus cetuximab (i.e. an epidermal growth factor receptor – directed monoclonal antibody) [242] [247].

As extensively argued in Chapter I, a follow-up program is essential in the first few years after primary treatment to identify potentially curable relapses. However, there is no consensus in the medical community around the optimum timing of visits and number of radiological assessments to be carried out in this phase. Published recommendations are mostly based on retrospective studies and expert opinions, and clinical practice varies across countries, hospitals, and individual specialists [16] [32]. In particular, the added value of an intensive radiological assessment over a scheme based on self-reported symptoms (e.g. increased local pain, difficulty in swallowing, development of a new lump in the neck or other local symptoms) has not yet been confirmed in any prospective randomized study.

The research question for this study originated from the HETeCo trial [33], although most of the data are retrieved from other sources and combined in a model-based framework. This chapter, indeed, presents an exploratory model assessing the cost-effectiveness of the two strategies under evaluation in the ongoing RCT, where an intensive follow-up with frequent radiological investigations (including MRI, CT, and PET scans) is compared with minimal, symptom-based surveillance. A secondary objective is to identify the extent to which further collection of data from the trial is valuable for reducing the uncertainties that inevitably affect the model. The analyses

have been reported according to the CHEERS requirements for economic evaluations [73] (Table A5.1).

5.2 Methods

A decision-analytic Markov model is developed to assess the long-term health and economic consequences of the two strategies compared in the HETeCo trial. The model is populated using data from a variety of sources, including a retrospective study [248] conducted at the NCI (Milan, Italy) to replicate the pattern of cancer recurrences detected in the model; this study also acted as hypothesis generator for the trial. Additional parameters are obtained from the medical literature and expert opinion. The cost analysis is conducted from the perspective of a major Italian region (i.e. Lombardy).

5.2.1 Clinical studies

Full details regarding the HETeCo trial are reported in the protocol [33] and in Chapter I. Briefly, patients with a diagnosis of clinical or pathological stage III-IV squamous HNC in the oral cavity, oropharynx, larynx, or hypopharynx and without evidence of disease six months (± 1) after having received radiotherapy with curative intent (alone or with systemic therapy or in postoperative setting) are randomly allocated to one of two follow-up programs. The primary objective of the trial is to compare the two-year health outcomes and costs arising from two different surveillance schemes after primary treatment for HNC. Secondary objectives include the assessment of the number of potentially salvageable recurrences or second primaries, the cause-specific survival, and the overall survival in the two study groups.

The trial plans to randomize 330 patients (i.e. 165 per arm) to two alternative surveillance schemes. The non-intensive follow-up (arm A), designed according to the

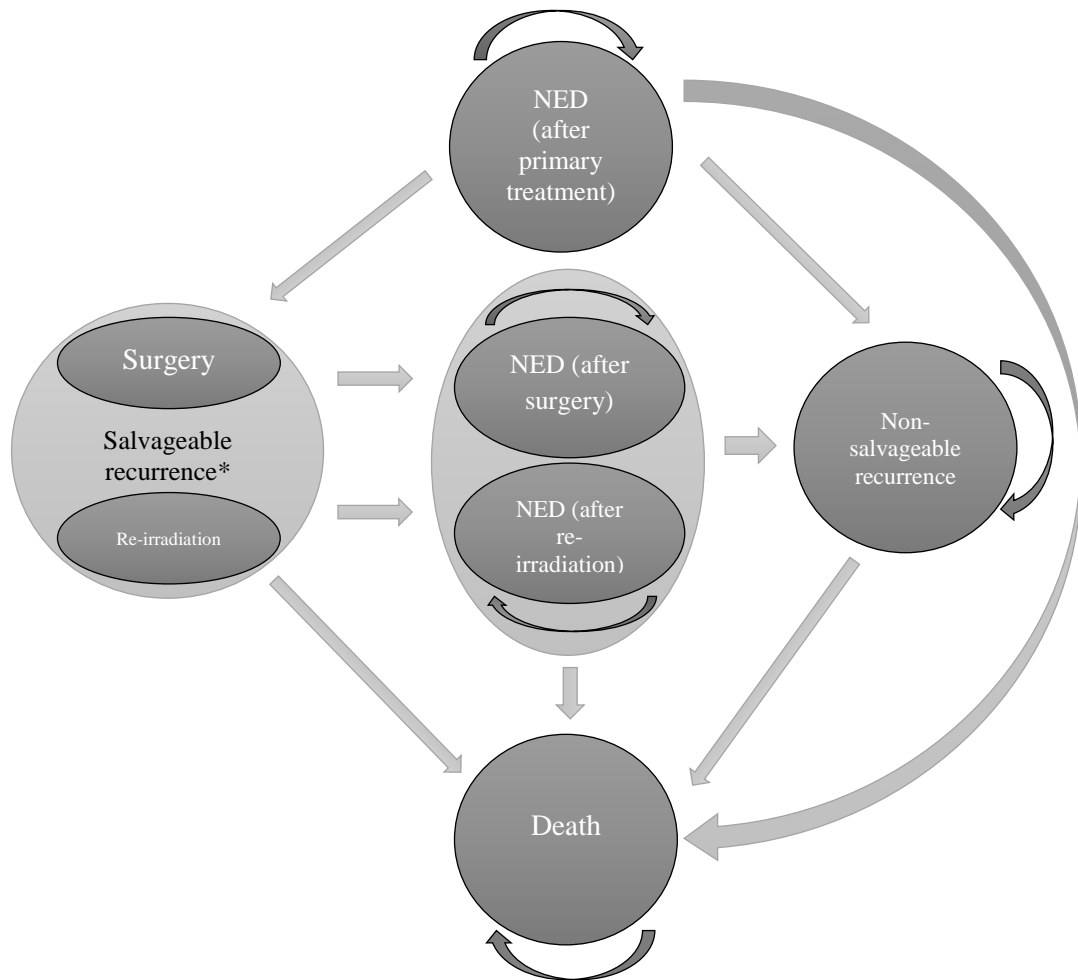
NCCN guidelines [34], comprises a few outpatient visits depending on cancer subsite. At each visit, patients receive both physical and fibre optic endoscopic examinations; laboratory tests including complete blood count and renal, hepatic, and thyroid function are performed once a year. Radiological assessment through MRI or CT is performed within six months of completion of treatment and thereafter only at the occurrence of new signs or symptoms. Patients are instructed how to recognize signs or symptoms of disease recurrence and contacted by phone between visits to monitor any health changes that might be related to cancer recurrence.

The alternative strategy (arm B) is a more intensive follow-up where outpatient visits are performed similarly to arm A, including fibre optic endoscopic examinations and laboratory tests. Imaging (MRI or CT) tests are scheduled for all patients twice a year in the first two years and annually in the third and fourth years. The choice between MRI and CT is made in accordance with institutional policies, but MRI is preferred for all subsites except for laryngeal cancer. PET scans are performed annually in the first three years in patients aged ≥ 50 years and with a smoking history of ≥ 20 pack/years.

Before starting the trial, a retrospective study was conducted at the NCI where the medical charts of 326 patients affected by stage III or IV HNC without evidence of disease 6 months after being treated with chemo-radiotherapy between 1998 to 2010 were reviewed [248]. According to the hospital's guidelines, all patients were enrolled in a 5-year follow-up program of outpatient visits and radiological examinations (i.e. MRI, CT, and PET), with frequency decreasing over time. In total, 113 patients (35%) were diagnosed with a recurrence or a second tumour. Specifically, 38 out of 113 (34%) cases presented loco-regional recurrences, 44 (39%) distant metastases and 31 (27%) second primary tumours; most recurrent patients (84%) were diagnosed during the first 3 years of follow-up.

5.2.2 Model structure

Figure 5.1 Markov state-transition diagram.



Legend. A hypothetical cohort of 1,000 62-year-old patients (i.e. mean age of the trial population) enter the model in the no evidence of disease (NED) state after the end of primary treatment. The individuals move between the health states based on a set of monthly transition probabilities. The symbol * indicates a temporary state, i.e. patients are not allowed to stay more than 1 cycle-month.

A Markov state-transition model (Figure 5.1) with mutually exclusive health states is developed to predict the lifelong costs and effects from the two follow-up strategies under investigation in the HETeCo trial. The choice of model was informed by a systematic literature review of economic evaluation studies in cancer follow-up (Chapter II) and previously published health economic models in HNC [157] [249]. Most of the reviewed studies, indeed, developed a Markov model to conduct cost-effectiveness analyses, whereas only few chose a discrete event simulation (DES) approach, which is not considered for this study due to the current unavailability of

individual-level patient data from the HETeCo trial to calibrate the model [85] [86]. Additionally, extensive discussions were conducted at the NCI with clinical experts including a medical oncologist, a radiotherapist, and a surgeon. The choice of classifying the model's health states based on the type of secondary treatment administered instead of the type of cancer relapse diagnosed (i.e. loco-regional recurrence, distant metastasis or second primary) was agreed with the clinicians and justified by more meaningful health and economic consequences being associated with alternative treatment options. The clinical opinion was particularly relevant to identify two alternative paths within the model for surgically treated and re-irradiated recurrent patients, respectively. All the clinical experts involved in this study approved the final model structure.

The model simulates the experience of a hypothetical cohort of 1,000 patients after being treated for primary stage III-IV HNC; mean age and gender ratio are representative of the patients enrolled in the trial (until May 2016). All patients enter the model free of disease six months after curative treatment for primary cancer and move through the different health states according to a set of transition probabilities. Recurrent patients are divided based on the intent of the treatment received (i.e. potentially curative or palliative); patients treated with curative intent are assigned to 'surgery' or 're-irradiation' states to capture the different costs and outcomes arising from the two therapeutic options. Patients without progression remain in the 'no evidence of disease' health states; the final, absorbing state is 'death'. The cycle length of the model is one month with a lifetime horizon. Utility values ranging from 0 (death) to 1 (perfect health) and costs are applied to the time spent in each health state. The model is run until the whole cohort (i.e. >99%) dies to estimate differences in life expectancy (i.e. LYG), quality-adjusted life expectancy (i.e. QALYs) and long-term costs associated with the two follow-up schemes.

5.2.3 *State-transition matrix*

Table 5.1 reports the clinical parameters derived from the published literature, unpublished data, and expert opinion; Table 5.2 presents the state-transition matrix with monthly transition probabilities between states, which are mainly derived from a combination of clinical parameters. Using the information from the retrospective study [248], the overall risk of relapse (i.e. local-regional recurrences, metastases and second primaries) is estimated at 29% in the first 3 years and 6% in the last two years of follow-up; 80% of the curable relapses undergo surgery and the remaining 20% are treated with re-irradiation. This recurrence risk is broadly similar to that reported by a recent study [244], which estimated the proportion of disease events (i.e. recurrence or death) to be 38% at 36 months. The proportion of potentially salvageable recurrences during follow-up in the study arm A (25%) is derived from the literature [250], while the percentage in group B (50%) is a clinical assumption of the HETeCo trial, which is intended to test whether a more intensive radiological assessment could detect a higher rate of salvageable relapses. In both groups, a 3% annual risk of second primary tumours is assumed from the 6th year onwards [21] [22], 47% of which are considered as curable [248]. As a simplification, all second primaries are hypothesized to affect the head and neck region in the model, disregarding other cancer localizations such as lung and oesophagus, and surgery is considered the only potentially curative treatment for second primaries. The ‘potentially curative treatment’ state is assumed temporary, meaning that each patient can only remain in it for one cycle. Indeed, the average length of hospitalization after salvage surgery has been reported as 30 days [251]. Up to 5% of the patients undergoing each salvage treatment (i.e. surgery or re-irradiation) are expected to die for causes related to re-treatment [252] [253]. The risk of relapsing after secondary treatment is estimated at 3% monthly, based on published studies [251] [253] [254]. These studies report almost the same value for salvage surgery and re-irradiation;

moreover, this parameter is consistent with the 0.009 weekly (i.e. 0.034 monthly) adopted by a previous cost-effectiveness model in HNC in Italy [157]. Any recurrence (or second primary) beyond the first is assumed to be treated with palliative intent only [192]. Patients receiving palliative chemotherapy are assumed to have a median survival of 10 months, corresponding to a 1-year overall survival of around 43% (equivalent to a 6.6% monthly mortality) [247] [255]; this value is comparable with the data reported by a review study [25] translated into a monthly probability (i.e. 6.8%). In each health state patients also experience a general risk of dying for reasons other than HNC; mortality rates for 5-year age groups divided by gender are obtained from official statistics for Lombardy [256]; a weighted monthly risk of dying is calculated by using the male-female ratio existing among the first 60 patients enrolled in the trial. Annual probability values reported in the literature are transformed into monthly probabilities using the appropriate formula linking probabilities and rates: $p = 1 - \exp(-r \cdot t)$ where p is probability, r is rate and t is the time expressed in months or years [257].

Table 5.1 Clinical parameters.

	Value	Distribution			Source
Male proportion	0.87	Beta	$\alpha=52$	$\beta=8$	HETeCo trial*
% of laryngeal cancer (requiring CT instead of MRI)	0.22	Beta	$\alpha=13$	$\beta=47$	HETeCo trial*
% of patients receiving PET (arm B)	0.57	Beta	$\alpha=34$	$\beta=26$	HETeCo trial*
Recurrence risk over follow-up (5 years)	0.35	Beta	$\alpha=113$	$\beta=213$	[248]
% recurrences in the first 3 years	0.84	Beta	$\alpha=15.2$	$\beta=2.9$	[248]
Ratio surgery/re-irradiation	80:20	Beta	$\alpha=19.2$	$\beta=4.8$	Clinical opinion
Potentially curable relapses (arm A)	0.25	Beta	$\alpha=74.8$	$\beta=224.3$	[250]
Potentially curable relapses (arm B)	0.50	Beta	$\alpha=49.5$	$\beta=49.5$	Clinical opinion
Annual risk of second primaries (6 th year onwards)	0.03	Beta	$\alpha=97.0$	$\beta=3135.4$	[21] [248]
Potentially curable second primaries (6 th year onwards)	0.47	Beta	$\alpha=15$	$\beta=16$	[248]
% of patients receiving home-based palliative care	0.135	Beta	$\alpha=20,985$	$\beta=134,461$	[258]
% of patients admitted to the hospital in the last month of life	0.78	Beta	$\alpha=8,947$	$\beta=2,523$	[259]

* Trial preliminary data (n=60 patients). HETeCo: Health and Economic Outcomes of Two Different Follow Up Strategies in Effectively Cured Advanced Head and Neck Cancer; MRI: magnetic resonance imaging; CT: computed tomography; PET: positron emission tomography; RCT: randomized controlled trial.

Table 5.2 State-transition matrix.

<i>Monthly transition probabilities (from/to)</i>		Value	Distribution			Source
NED (after primary treatment) Years 1st-3rd	NED (after primary treatment)	1-others	**			
	Salvageable recurrence (surgery)	A: 0.0019	**			[248]; Clinical opinion
		B: 0.0038				
	Salvageable recurrence (re-irradiation)	A: 0.0005	**			[248]; Clinical opinion
		B: 0.0010				
Non-salvageable recurrence (chemotherapy/palliative care)	A: 0.0072	**			[248]; Clinical opinion	
	B: 0.0048					
Death (for other causes) *	0.0032	Fixed			[256]	
NED (after primary treatment) Years 4th -5th	NED (after primary treatment)	1-others	**			[248]; Clinical opinion
	Salvageable recurrence (surgery)	A: 0.0005	**			[248]; Clinical opinion
		B: 0.0010				
	Salvageable recurrence (re-irradiation)	A: 0.0001	**			[248]; Clinical opinion
		B: 0.0002				
Non-salvageable recurrence (chemotherapy/palliative care)	A: 0.0018	**			[248]; Clinical opinion	
	B: 0.0012					
Death (for other causes) *	0.0032	Fixed			[256]	

Table 5.2 (cont.) State-transition matrix.

<i>Monthly transition probabilities (from/to)</i>		Value	Distribution			Source
NED (after primary treatment) From the 6th year onwards	NED (after primary treatment)	1-others	**			
	Salvageable recurrence (surgery)	0.0012	**			[248]
	Non-salvageable recurrence (chemotherapy/palliative care)	0.0013	**			[248]
	Death (for other causes)	0.0032	Fixed			[256]
Salvageable recurrence (surgery)	NED (after salvage surgery)	1-others	**			
	Treatment-related death	0.0380	Beta	$\alpha=10$	$\beta=251$	[252]
	Death (for other causes) *	0.0032				[256]
Salvageable recurrence (re-irradiation)	NED (after re-irradiation)	1-others	**			
	Treatment-related death	0.0490	Beta	$\alpha=2$	$\beta=39$	[253]
	Death (for other causes)	0.0032	Fixed			[256]
NED (after salvage surgery)	NED (after salvage surgery)	1-others	**			
	Non-salvageable recurrence (chemotherapy/palliative care)	0.0300		$\alpha=26$	$\beta=13$	[251]
	Death (for other causes) *	0.0032	Fixed			[256]
NED (after re-irradiation)	NED (after re-irradiation)	1-others	**			
	Non-salvageable recurrence (chemotherapy/palliative care)	0.0300		$\alpha=0.03$	$\beta=0.90$	[253]
	Death (for other causes) *	0.0032	Fixed			
Non-salvageable recurrence (chemotherapy/palliative care)	Non-salvageable recurrence (chemotherapy/palliative care)	1-others	**			
	Death from HNC	0.0660	Beta	$\alpha=93.3$	$\beta=1,320.8$	[247] [255]
	Death (for other causes) *	0.0032	Fixed			[256]

* The death rate is reported only for the age group 60-64 years corresponding to the age when the model starts. **These parameters are obtained from a combination of other values; thus, distributions are assigned to original values only. HETeCo: Health and Economic Outcomes of Two Different Follow Up Strategies in Effectively Cured Advanced Head and Neck Cancer; HNC: head and neck cancer; NED: no evidence of disease.

5.2.4 Health state utility values

The utility parameters are summarized in Table 5.3 and, since the HETeCo trial does not collect HRQoL measures after the patient is diagnosed with recurrence, mainly identified by a systematic literature review of HSUVs in HNC (Chapter III). An average utility value for the ‘no evidence of disease’ state (i.e. study recruitment time) is calculated from the EQ-5D-5L trial data using the English value set [227], in the absence of an Italian value set; the same value (i.e. 0.85) is confirmed by a cross-sectional study [186] recruiting a comparable population (i.e. HNC patients with a follow-up of at least 6 months after curative radiotherapy). Utility values for re-treatment health states are retrieved from a Canadian study [155] using the standard gamble technique to elicit preferences from the public; the value for the “non-salvageable recurrence” state also corresponds to that adopted (i.e. 0.33) by another HNC model study [157]. The utility parameter for the ‘no evidence of disease’ state after salvage surgery (i.e. 0.62) comes from a study [172] of HNC patients with no evidence of disease three months after completion of treatment and using several measurement methods, including EQ-5D; the same value is applied to patients treated with re-irradiation.

Table 5.3 Utility parameters.

Health state	Value	Distribution			Source
NED (after primary treatment)	0.85	Beta	$\alpha=1337.1$	$\beta=236.0$	HETeCo trial *; [186]
Salvageable recurrence (surgery)	0.57	Beta	$\alpha=348.7$	$\beta=263.1$	[155]
Salvageable recurrence (re-irradiation)	0.57	Beta	$\alpha=348.7$	$\beta=263.1$	[155]
NED (after salvage surgery)	0.62	Beta	$\alpha=22.2$	$\beta=13.6$	[172]
NED (after re-irradiation)	0.62	Beta	$\alpha=22.2$	$\beta=13.6$	Assumption
Non-salvageable recurrence (chemotherapy/palliative care)	0.34	Beta	$\alpha=190.4$	$\beta=369.6$	[155]
Death	0.00	Fixed			Assumption

* Trial preliminary data (n=60 patients). NED: no evidence of disease; HETeCo: Health and Economic Outcomes of Two Different Follow Up Strategies in Effectively Cured Advanced Head and Neck cancer.

5.2.5 Cost data

The cost analysis is conducted using a top-down approach from the perspective of the Lombardy regional healthcare system and thus includes only direct medical costs. Data on costs (not publicly available) are from Diagnosis-Related Groups (DRGs) and other regional tariffs (year 2016) for hospital admissions, specialist visits, radiological exams, laboratory tests and outpatient treatment regimens (Table 5.4). The cost of each follow-up program (A or B) performed in the ‘no evidence of disease’ state is calculated for 5 years (i.e. standard length of HNC follow-up in Italy and abroad), according to the description provided in the trial protocol in terms of healthcare resources consumption. An average monthly cost of around €12 is estimated for follow-up A, while two different monthly costs (€92 up to the 3rd year and €15 from the 4th year onwards) are estimated for follow-up B, recognizing that most radiological assessments are performed during the first three years. Using data from the first 60 patients enrolled in the trial, the cost of CT and MRI scans is weighted by the proportion of laryngeal and non-laryngeal cancers, respectively, while the PET cost is considered only for heavy smokers (i.e. ≥ 20 pack/years) aged more than 50. No surveillance costs are assumed from the 6th year onwards (Table 5.5). Patients surviving the ‘potentially curative treatment’ states are assumed to be monitored within a program of physical investigations resembling the less intensive scheme (arm A) with the primary objective of managing any re-treatment side effects. Since patients enter the ‘no evidence of disease’ state three months after salvage treatment at different times in the model, the monthly cost of this “secondary” follow-up is weighted by the proportion of people actually receiving the intervention at each cycle. The cost of salvage surgery is valued according to DRG 49 (i.e. major head and neck surgery). Re-irradiation is assumed to be prescribed as a cycle of intensity modulated radiation therapy (IMRT) sessions over a 5-week period (approximated at one month in the model) based on current practice in

Lombardy. Each re-irradiation cycle (≥ 5 sessions) is reimbursed at €10,000 and includes any specialist visit or examination that might be performed during the treatment. The standard platinum-based chemotherapy supplemented with 5-fluorouracil (5FU) and cetuximab is used for recurrences treated with palliative intent [247]. Monthly administration of cisplatin+5FU and weekly administration of cetuximab until one month before death is assumed in the model based on expert opinion. The average monthly cost of chemotherapy for the ‘palliatively treated recurrence’ state takes account of the higher drug dosage during the first administration of cetuximab. In Lombardy, the reimbursement of high cost chemotherapeutics, such as cetuximab, comprises the drug acquisition cost on behalf of the hospital plus a fixed tariff (MAC01) covering a preliminary specialist visit, laboratory exams and the drug administration. Additionally, an average cost of dying from HNC is assigned to each patient entering the ‘palliatively treated recurrence’ state calculated from published cost data for Italy according to the estimated consumption of formal end-of-life care (i.e. home-based assistance and hospital care). In detail, the proportion of terminally ill cancer patients receiving palliative care at home is estimated at 13.5% [258] at a cost of €2,100 [260]. Moreover, 78% of cancer patients [259] are assumed to be hospitalized in the last month of life at a cost of €4,000 [261]. The cost of alternative services (e.g. hospice and nursing home) is disregarded since it applies to less than 10% in Italy [258]. Monthly cost values for all model health states are reported in Table 5.6.

Table 5.4 Unit costs (€, 2016).

	Value	Code	Source
<i>Treatments</i>			
Major head and neck surgery	5,444	DRG 49	Regional tariff
Intensity modulated radiation therapy, including computed tomography (more than 5 sessions)	10,000	92.29.L	Regional tariff
Administration of high-cost chemotherapeutics (including specialist visits and laboratory tests)	44.00	MAC01	Regional tariff
Cetuximab (100 mg) *	153.56		File F
Cisplatin (50 mg) *	5.09		Local charge
5-Fluorouracil (500 mg) *	1.72		Local charge
Home-based palliative care (100 days)	2,100		[260]
Hospital admissions one month before death	4,000**		[261]
<i>Imaging</i>			
Computed tomography of head and neck with contrast	159.93	87.03.1/8	Regional tariff
Magnetic resonance imaging of head and neck with contrast	238.87	88.91.7	Regional tariff
Positron emission tomography (brain)	1,081.86	92.11.7	Regional tariff
<i>Specialist visits</i>			
First visit	22.50	89.7B.6	Regional tariff
Control visit	17.90	89.01.F	Regional tariff
<i>Laboratory tests</i>			
Blood cell count	4.05	90.62.2	Regional tariff
Urea	1.70	90.44.1	Regional tariff
Creatinine blood test	1.70	90.16.3	Regional tariff
Sodium	1.70	90.40.4	Regional tariff
Potassium	1.70	90.37.4	Regional tariff
Calcium	1.70	90.11.4	Regional tariff
Aspartate aminotransferase	1.70	90.09.2	Regional tariff
Alanine aminotransferase	1.70	90.04.5	Regional tariff
Bilirubin	1.70	90.10.4	Regional tariff
Alkaline phosphatase	1.70	90.23.5	Regional tariff
Thyrotropin	8.40	90.42.1	Regional tariff
Free thyroxin	9.50	90.42.3	Regional tariff
<i>Other</i>			
Phone call (cost/minute)	0.10		Telecom Italia

* Drug costs are value-added tax (VAT) excluded. ** This value is approximated from a graph. DRG: diagnosis-related group; MAC: macro ambulatory activity of high complexity.

Table 5.5 5-year follow-up costs in the HETeCo trial arms (A and B).

	Unit cost (€)	Follow-up (arm A)		Follow-up (arm B)				
		Quantity (5 years)	Total cost (€, 5 year)	Quantity (3 years)	Quantity (2 years)	Total cost (€, 3 years)	Total cost (€, 2 years)	Total cost (€, 5 years)
Specialist visit (first)	22.5	1	22.50	1	0	22.50	0.00	22.50
Specialist visit (control)	17.9	14	250.60	10	4	179.00	71.60	250.60
Blood cell count	4.05	6	24.30	4	2	16.20	8.10	24.30
Urea	1.7	6	10.20	4	2	6.80	3.40	10.20
Creatinine blood test	1.7	6	10.20	4	2	6.80	3.40	10.20
Sodium	1.7	6	10.20	4	2	6.80	3.40	10.20
Potassium	1.7	6	10.20	4	2	6.80	3.40	10.20
Calcium	1.7	6	10.20	4	2	6.80	3.40	10.20
Aspartate aminotransferase	1.7	6	10.20	4	2	6.80	3.40	10.20
Alanine aminotransferase	1.7	6	10.20	4	2	6.80	3.40	10.20
Alkaline phosphatase	1.85	6	11.10	4	2	7.40	3.70	11.10
Bilirubin	1.7	6	10.20	4	2	6.80	3.40	10.20
Thyrotropin	8.4	6	50.40	4	2	33.60	16.80	50.40
Free thyroxin	9.5	6	57.00	4	2	38.00	19.00	57.00
CT (% <i>larynx</i> *)	159.93	1 (0.22)	35.18	5 (0.22)	1 (0.22)	175.92	35.18	211.11
MRI (% <i>all other subsites</i> *)	238.87	1 (0.78)	186.32	5 (0.78)	1 (0.78)	931.59	186.32	1117.91
Phone call**	0.10/min	10	10.00	0	0	0	0	
PET (% ≥ 50 years and ≥ 20 pack/years*)	1081.86	0	0	3 (0.57)	0	1849.98	0	1849.98
Total			729.00			3308.60	367.90	3676.50
Total (monthly)			12.15			91.91	15.33	61.27

*Trial preliminary data (n=60 patients). ** Each call is assumed to last 10 minutes. MRI: magnetic resonance imaging; CT: computed tomography; PET: positron emission tomography.

Table 5.6 Health state description and monthly costs.

Health state	Description	Cost (€)
NED (after primary treatment): follow-up A	5-year follow-up program based on NCCN guidelines. Frequency of outpatient visits depending on cancer subsite. Laboratory tests performed once a year. Loco-regional imaging (MRI/CT*) performed within six months after treatment end and then recommended only at the occurrence of new signs or symptoms. Inter-visit phone calls to monitor patient's symptomatology.	12.1 No costs from the 6 th year onwards
NED (after primary treatment): follow-up B	Alternative 5-year follow-up program. Frequency of outpatient visits depending on cancer subsite. Laboratory tests performed once a year. Loco-regional imaging (MRI/CT*) requested two times/year in the first two years and once/year in the third and fourth years. PET scan performed yearly in the first three years only in high-risk patients (≥ 50 years and ≥ 20 pack/years).	Up to the 3 rd year: 91.9 From the 4 th year: 15.3 No costs from the 6 th year onwards
Salvageable recurrence (surgery)	Major head and neck surgical intervention with excision of tumour and surrounding tissues. Hospital stay of around one month.	5,444
Salvageable recurrence (re-irradiation)	Intensity-modulated radiation therapy with curative intent five days a week for 5-6 weeks.	10,000
NED (after salvage surgery/re-irradiation)	Same program described in NED (after primary treatment), follow-up A	12.1 No costs from the 6 th year onwards
Non-salvageable recurrence (chemotherapy/palliative care)	Combined treatment of platinum-based chemotherapy plus 5-fluorouracil (once a month) and cetuximab (once a week) administered up to one month before death. End-of-life care consisting in home-based palliative care and hospital admissions during the last month of life.	Chemotherapy (1 st month): 3,720.0 Chemotherapy: (2 nd month onwards): 3,263.9 (Average) monthly cost: 3,063.3** End-of-life care***: 3,403.5

*MRI is preferred for all HNC subsites except for larynx. **The average monthly cost for chemotherapy (including administration) is calculated based on an average stay of 14 months in the “non-salvageable recurrence” state. A gamma distribution ($\alpha=100$; $\beta=30.6$) is assigned to the obtained value (€3,063.3). ***End-of-life care cost is calculated as $13.5\% * 2,100 + 78\% * 4,000 = €3,403.5$ over a 3-month period and assigned to all patients entering the “non-salvageable recurrence” state. MRI: magnetic resonance imaging; CT: computed tomography; PET: positron emission tomography; NCCN: National Comprehensive Cancer Network; NED: no evidence of disease.

5.2.6 Cost-effectiveness analysis

Health outcomes (i.e. LYG and QALYs gained) and total costs are combined into an ICER = $\frac{Cost_B - Cost_A}{LYS_B - LYS_A}$ and ICUR = $\frac{Cost_B - Cost_A}{QALYS_B - QALYS_A}$ to represent the incremental cost of achieving one unit of health outcome when an intensive follow-up strategy (arm B) replaces a less intensive one (arm A). The ICUR obtained is compared with the range of €25,000-€40,000 [262] recommended by the Italian Health Economics Association to assess the cost-effectiveness of healthcare interventions. All costs and outcomes are discounted at 3% (converted to 0.247% monthly) following the same guidelines and expressed in Euro (€) 2016.

5.2.7 One-way sensitivity analysis

One-way sensitivity analyses explore the robustness of the base-case results by varying some key model parameters one at a time. First, the cost of salvage surgery is varied between €5,280.68 and €5,607.32 according to the hospital type as reported by official tariffs for Lombardy. Second, the annual discount factor for costs and outcomes is increased up to 3.5%, as recommended by NICE [263]. Third, the risk of overall relapse during follow-up is increased up to 50%, as reported by other studies [10]. Fourth, treatment-related mortality is set at zero reflecting a study [251] which reported no perioperative deaths in patients undergoing salvage surgery, although the sample size is very small (i.e. 41 patients) compared to the study used in the base case scenario (n=261) [252]. Similarly, the risk of dying because of re-irradiation-related toxicities is set equal to the lowest (i.e. 0%) and the highest (i.e. 11%) values from studies on IMRT and systematically reported in a review [26]. An intensive program of radiological assessments after a salvage treatment is not justifiable by any survival gains in case of a second recurrence; therefore, the base-case analysis applies the less intensive program cost to the ‘no evidence of disease’ state after surgery or re-irradiation. The impact of

this assumption is explored by assuming that patients have the same follow-up intervention (A or B) received in the ‘no evidence of disease’ state after primary treatment.

5.2.8 Two-way sensitivity analysis

A two-way sensitivity analysis is performed to assess the simultaneous effect of varying the proportion of salvageable recurrences in arm A and in arm B, with all else unchanged in the model. Cost-effectiveness is determined with respect to the Italian thresholds of €25,000 and €40,000 per QALY [262]. This analysis is important because of the uncertainty surrounding this parameter based on published literature and clinical opinion and not yet confirmed by the ongoing RCT.

5.2.9 Probabilistic sensitivity analysis

A probabilistic sensitivity analysis (PSA) is performed using Monte Carlo simulation with 5,000 random iterations from the distributions assigned to the model parameters. This is only undertaken for the ICUR, since a stated threshold for the cost per LYG is not available. A beta distribution is chosen for transition probabilities and utility values. If count data are not available, beta distributions are fitted using the ‘methods of moments’ based on the mean value and standard errors reported in the studies [257]. If neither standard error is available, a variance equal to 10% of the mean is arbitrarily assumed. No distributions are assigned to the cost values based on official tariffs (e.g. salvage surgery). However, when the cost assigned to a health state is obtained through a combination of different items, the proportion of users (e.g. patients receiving home-care palliative care) is varied using a beta distribution or the cost itself (e.g. the cost of palliative chemotherapy) is varied using a gamma distribution. A cost-effectiveness scatterplot illustrates the uncertainty surrounding the base-case ICUR; cost-effectiveness acceptability curves (CEACs) are also calculated to quantify the

probability that the more intensive follow-up (i.e. arm B) would be deemed cost-effective at different threshold values.

5.2.10 Expected value of perfect information

The decisions based on existing information are inevitably uncertain, and this uncertainty can lead to the ‘wrong’ decision being made; in the present study, this would mean the adoption of a follow-up program that is not cost-effective in HNC. The expected value of perfect information (EVPI) represents the ‘value’ of obtaining perfect knowledge of the ‘true’ values of all parameters used in a model. In other words, it can be interpreted as *the maximum amount the decision-maker is willing to pay to obtain perfect information* [249]. Based on the non-parametric approach presented by Briggs et al. [257] and Oostenbrink et al. [264], the per patient EVPI is calculated as the difference between the expected net monetary benefit (NMB) with perfect information and the NMB under the current condition of uncertainty. Thereafter, since information is a non-rival public good, the population EVPI is calculated for the total number of potential beneficiaries of reducing the model’s uncertainty. The total number of incident HNC cases is estimated at 13,000 yearly in Italy [265] and 60% of them are locally advanced patients treated with combined modality therapy [17]; of these, about 10% recur locally or distantly shortly after treatment [266]; thus, around 7000 patients are estimated to be free of disease for at least six months and enter a regular follow-up program. The population EVPI is discounted at 3% on annual basis over a 10-year period (i.e. the expected lifetime of the programs under evaluation) [126].

The decision model is implemented in Microsoft Excel 2013 with the support of @RISK software (Palisade Corp) for the sensitivity analyses.

5.3 Results

5.3.1 Base-case analysis

The cost-effectiveness results are presented in Table 5.7. The extra-cost per patient enrolled in the more intensive follow-up (arm B) is €1,903 compared to the less intensive option (arm A). At the same time, the addition of routine radiological investigations leads to an increase of 0.10 QALYs and 0.15 LYG per patient. Thus, the ICUR results in €19,951 per QALY gained and the ICER in €13,123 per LYG.

Table 5.7 Cost-effectiveness results (base-case analysis).

	Costs (€) *	Incremental costs (€)	QALYs*	Incremental QALYs	LYs*	Incremental LYs (LYG)	ICUR (€)	ICER (€)
Follow-up A	16,895		5.36		6.61			
Follow-up B	18,798	1,903	5.45	0.10	6.75	0.15	19,951	13,123

*Costs and outcomes (QALYs, LYs) are reported per patient. LYs: life years; LYG: life years gained; ICER: incremental cost-effectiveness ratio; ICUR: incremental cost-utility ratio; QALYs: quality-adjusted life years.

5.3.2 Sensitivity analyses

In univariate sensitivity analysis (Table 5.8), the ICUR is most sensitive to the overall risk of recurrence over the 5-year follow-up. Specifically, the recurrence risk is inversely proportional to the incremental cost per QALY, falling to €11,737 (ICER: €7,718) using a 50% value as reported by some studies [10] [243] and rising to €40,228 (ICER: €26,468) at the lowest risk of 20%. Other variables including secondary treatment-related mortality, cost of head and neck surgery and discounting have a limited effect on the results.

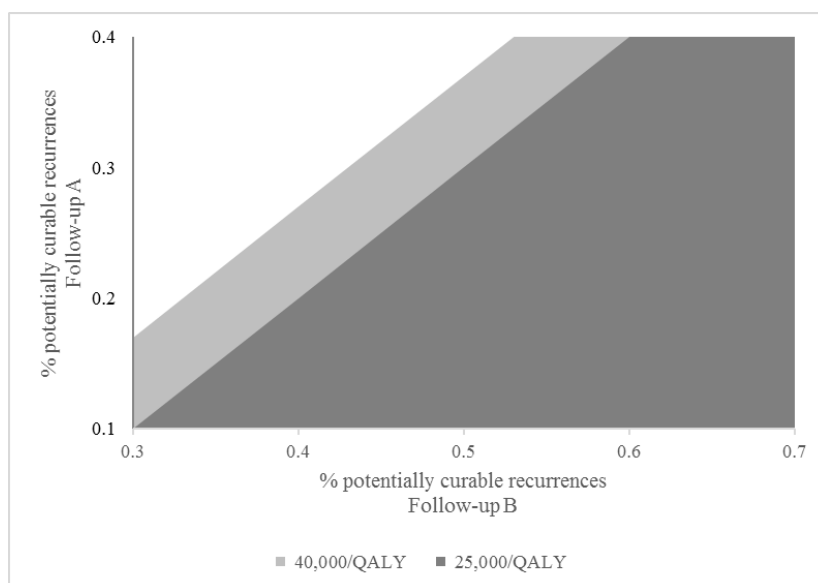
Table 5.8 One-way sensitivity analysis.

	ICUR (€)	ICER (€)
<i>Discount rate</i>		
3.0% (base-case)	19,951	13,123
3.5%* (NICE, [263])	19,985	13,179
<i>Risk of recurrences (over the 5-year follow-up)</i>		
20%	40,228	26,468
30%	24,474	16,100
35% (base-case)	19,951	13,123
40%	16,544	10,881
50%	11,737	7,718
<i>Treatment-related mortality (salvage surgery)</i>		
0% [251]	20,011	13,052
3.8% (base-case)	19,951	13,123
5% [267]	19,931	13,147
<i>Treatment-related mortality (re-irradiation) [26]</i>		
0%	19,971	13,099
4.9% (base-case)	19,951	13,123
11%	19,925	13,154
<i>Salvage surgery (cost, €)</i>		
5,281 (hospitals without emergency department)	19,849	13,056
5,444 (base-case)	19,951	13,123
5,607 (hospital with emergency department)	20,052	13,190
<i>Cost in 'no evidence of disease' state after salvage treatment (arm B)</i>		
12.1 (base-case)	19,951	13,123
91.9 (up to year 3th) and 15.3 (from the 4th year)	22,480	14,787

*corresponding to 0.287% monthly. ICER: incremental cost-effectiveness ratio; ICUR: incremental cost-utility ratio; NICE: National Institute for Health and Care Excellence.

Figure 5.2 presents the two-way sensitivity analysis (full details of the results are reported in Table 5.9). As expected, the cost-effectiveness of the intensive follow-up (arm B) increases with the positive difference between the “curability” of recurrences detected in arm B and arm A, respectively, reaching a maximum value of €6,330/QALY (€4,163/LYG) when this parameter is equal to 0.7 in arm B and 0.1 in arm A and a minimum value of €113,354/QALY (€74,561/LYG) when the difference between the two “curability” rates is only 0.05.

Figure 5.2 Two-way sensitivity analysis of the proportion of potentially salvageable recurrences (or second primaries) detected over the 5-year follow-up.



Legend: areas in dark grey represent the cost-effectiveness regions where the ICURs fall under the (upper and lower) thresholds recommended by Italian guidelines (i.e. the intensive follow-up is cost-effective compared to the less intensive one).

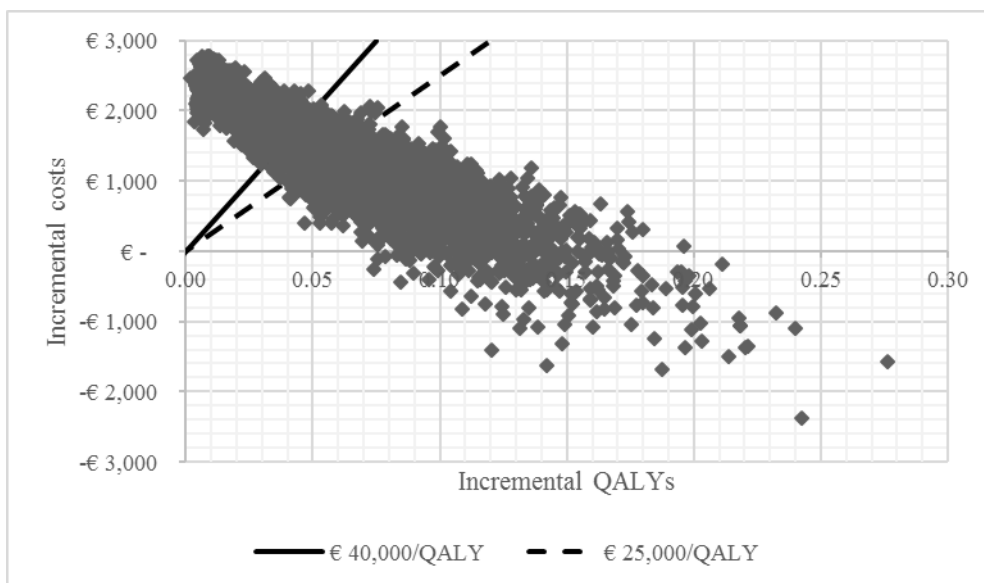
Table 5.9 Two-way sensitivity analysis of the proportion of potentially salvageable recurrences (or second primaries) detected over the 5-year follow-up.

		Intensive follow-up (arm B)								
		ICER	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65
Less intensive follow-up (arm A)	0.10	16,963	13,123	10,563	8,735	7,363	6,297	5,443	4,745	4,163
	0.15	23,363	16,963	13,123	10,563	8,735	7,363	6,297	5,443	4,745
	0.20	36,162	23,363	16,963	13,123	10,563	8,735	7,363	6,297	5,443
	0.25	74,561	36,162	23,363	16,963	13,123*	10,563	8,735	7,363	6,297
	0.30		74,561	36,162	23,363	16,963	13,123	10,563	8,735	7,363
	0.35	-79,033		74,561	36,162	23,363	16,963	13,123	10,563	8,735
	0.40	-40,635	-79,033		74,561	36,162	23,363	16,963	13,123	10,563
			Intensive follow-up (arm B)							
Less intensive follow-up (arm A)	ICUR	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70
	0.10	25,789	19,951	16,059	13,279	11,194	9,573	8,275	7,214	6,330
	0.15	35,518	25,789	19,951	16,059	13,279	11,194	9,573	8,275	7,214
	0.20	54,977	35,518	25,789	19,951	16,059	13,279	11,194	9,573	8,275
	0.25	113,354	54,977	35,518	25,789	19,951*	16,059	13,279	11,194	9,573
	0.30		113,354	54,977	35,518	25,789	19,951	16,059	13,279	11,194
	0.35	-120,154		113,354	54,977	35,518	25,789	19,951	16,059	13,279
	0.40	-61,777	-120,154		113,354	54,977	35,518	25,789	19,951	16,059

*Bold values correspond to the base-case results. ICER: incremental cost-effectiveness ratio; ICUR: incremental cost-utility ratio.

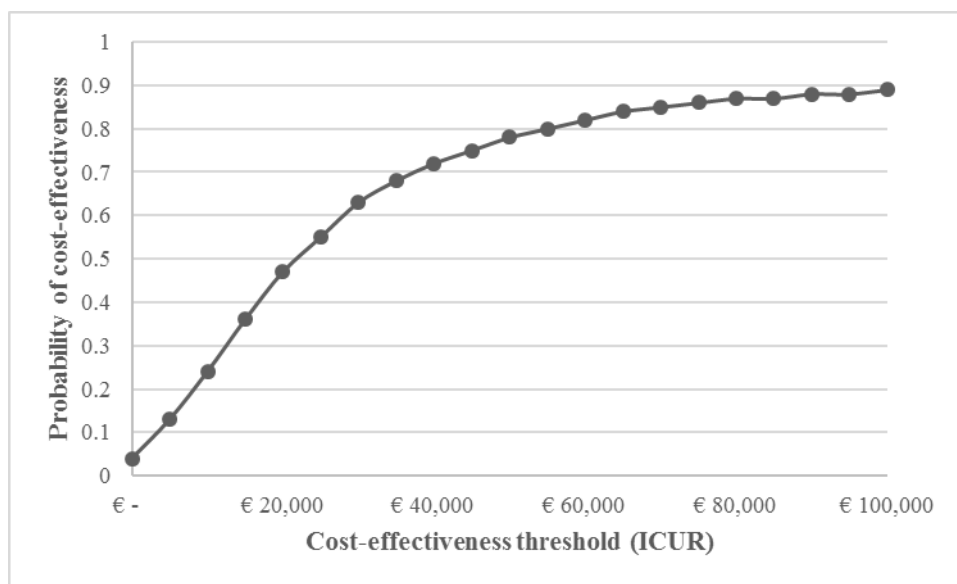
In the cost-effectiveness plane (Figure 5.3), most ICURs (72%) are to the right of the €40,000 threshold, suggesting that a more intensive follow-up (arm B) might represent a cost-effective option compared to a less intensive one (arm A). Even considering the lowest bound of the willingness-to-pay for Italy (i.e. €25,000), the intensive follow-up is cost-effective in more than 50% of simulations. None of the simulations fall in the left side of the graph (i.e. negative difference in QALYs); thus, under the model's assumptions, the more intensive follow-up (arm B) is always more effective than the symptom-driven surveillance (arm A). The CEAC (Figure 5.4) reports the probability of the intensive follow-up being cost-effective at different thresholds; at a willingness-to-pay equal to zero almost 5% of the simulations report a cost-saving result.

Figure 5.3 Cost-effectiveness plane with Monte Carlo simulations (5,000 runs).



Legend: the lines represent, respectively, the upper (€40,000) and lower (€25,000) cost-effectiveness thresholds for Italy; each dot represents one ICUR resulting from the 5,000 Monte Carlo simulations; points to the right of the straight lines are considered cost-effective.

Figure 5.4 Cost-effectiveness acceptability curve (CEAC).

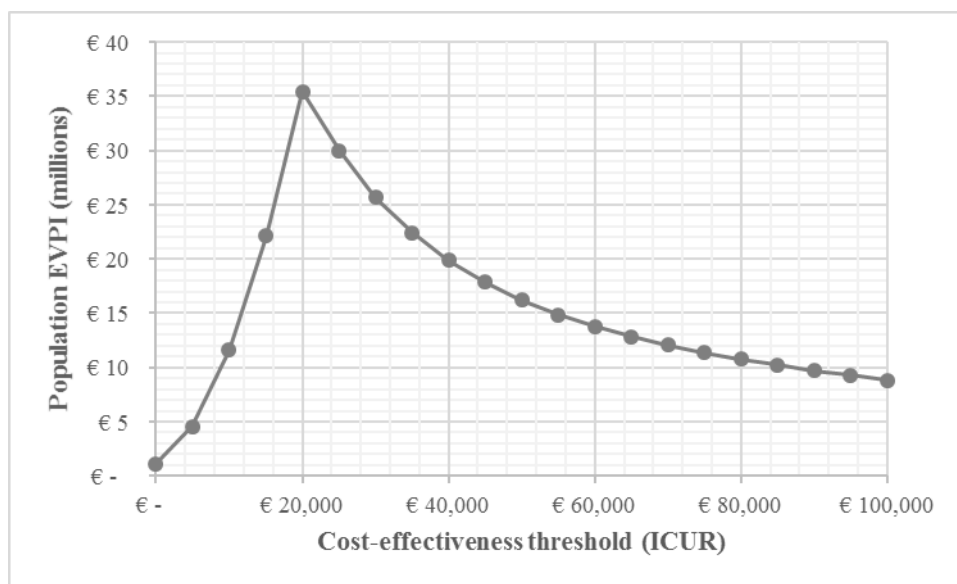


Legend: the probability of the intensive follow-up (arm B) to be cost-effective is plotted against the willingness-to-pay per QALY.

5.3.3 Expected value of perfect information

The PSA shows that at the threshold of €40,000 (€25,000) per QALY the probability of the more intensive follow-up to be cost-effective over the less intensive one is equal to 0.72 (0.55); thus, a considerable uncertainty exists in making decisions using the currently available data. The EVPI reveals that the value of undertaking additional research is worth €298.10 and €449.12 per patient at the higher (i.e. €40,000) and lower (i.e. €25,000) cost-effectiveness thresholds for Italy, respectively. This results in a population EVPI of around €20 million (about €30 million at the lowest threshold). The population EVPI using different cost-effectiveness thresholds is reported in Figure 5.5; the decision uncertainty (and, therefore, the EVPI) reaches a peak at around €20,000, and then slowly decrease at higher cost-effectiveness ratios. If the population EVPI exceeds the expected cost of additional research (performing the HETeCo trial costs at least €200,000), then it is potentially valuable to conduct further research in this area.

Figure 5.5 Expected value of perfect information (EVPI).



Legend: the expected value of perfect information (EVPI) at population level is shown at different threshold values for Italy.

5.4 Discussion

5.4.1 Synthesis of results

In a time of financial constraints, economic evaluations are increasingly used to inform decisions about the efficient allocation of healthcare resources. In cancer care, these methods have been mainly applied to evaluate drug treatments, while less evidence has been generated for other interventions such as screening, medical devices, surgical treatments, and follow-up programs [57]. There is no agreement on a common follow-up strategy in HNC across the guidelines proposed by the different cancer associations worldwide. Among them, the NCCN guidelines do not recommend routine imaging in the absence of symptoms [34]. Whereas, clinical practice in Italy usually involves regular radiological assessments over the post-treatment period, thus resulting in more intensive programs. The addition of routine MRI, CT and PET scans to the scheduled clinical examinations might increase the detection accuracy of recurrent HNC in

patients, especially the asymptomatic ones. However, the effectiveness (and cost-effectiveness) of these more intensive follow-up schemes has never been shown with rigorous methods.

Although RCTs provide valuable information for cost-effectiveness analyses, especially in terms of relative intervention effects, their observation periods are usually shorter than the required time horizon to estimate incremental costs and effects. Thus, model-based economic evaluations are increasingly being performed using data other than from trials [268]. Models present the advantage of gathering information from a variety of sources and explicitly accounting for the uncertainty associated with input parameters. For example, modelling gives decision-makers the ability to change parameters to examine how outcomes vary under different policy or clinical scenarios. Several methods have been developed to quantify uncertainty in decision models. First, the results of Monte Carlo simulations can be plotted in a cost-effectiveness plane with incremental outcomes (e.g. QALYs) measured on the horizontal axis and incremental costs on the vertical axis; alternatively, the CEAC illustrates the probability that the intervention is acceptable for a range of cost-effectiveness thresholds. Second, the EVPI reports the highest value of undertaking further research to resolve the uncertainty surrounding the decision at different levels of the cost-effectiveness threshold [269].

The current model explores the potential costs and outcomes in terms of QALYs and survival gains of two alternative follow-up programs in HNC, corresponding to the arms of an ongoing RCT. However, the trial was mainly used to generate a research question, while most of the data to populate the model were obtained from other sources. In this model, more intensive follow-up is cost-effective with a cost per QALY gained of €19,951. In the PSA, more than two-thirds of the simulations are below the willingness-to-pay of €40,000 and, at this threshold value, the EVPI is equal to around €300 per patient (€20 million at population level in Italy). Moreover, the two-way sensitivity

analysis shows that a difference in the proportion of potentially salvageable recurrences of about 0.15 between the two programs is sufficient to obtain an acceptable ICUR for arm B. The intermediate results can be compared with data reported in the literature. For example, the 5-year survival is equal to 58% and 60.5% in arm A and arm B, respectively, which is consistent with the epidemiological data [1] [6] [61]. Moreover, recurrent patients treated only with palliative intent spend an average 13-14 months in this state, which is aligned with the median overall survival (i.e. 10 months) reported by Vermorken [247], by assuming a positive skewness of the survival time distribution due to the presence of long survivors.

5.4.2 Comparison with other studies

In Chapter II, a systematic literature review of economic evaluations of post-treatment programs in oncology is presented. Some of the studies retrieved [82] [99] [126] [130] reported very high ICURs for intensive surveillance compared to less intensive options, whilst others [87] [96] [119] [127] concluded that the standard follow-up was cost-effective compared to a simplified one; in other cases, [85] [86] the cost-effectiveness of the intensive program depended on the patient's characteristics. The only study dealing with HNC follow-up [112] compared interventions differing only for the contact mode between patients and healthcare professionals (i.e. web-chat vs. traditional phone calls), and the time horizon was too short (i.e. 6 months) to capture meaningful endpoints. Thus, an economic evaluation study comparing follow-up programs of different intensity over a lifetime horizon was still lacking in the HNC literature.

Since no relevant HNC studies were retrieved from this systematic search, and most of the studies do not report the incremental costs per QALY or LYG, comparisons with other economic evaluations in cancer follow-up are not straightforward. However, in this model, the incremental cost per salvageable recurrence detected is equal to €20,249,

which is perfectly coincident with the value reported (i.e. £18,077) by a modelling study comparing an intensive versus a standard surveillance for colorectal cancer in UK [113]. Moreover, the estimated EVPI per patient is broadly similar to one reported in a model-based cost-utility analysis comparing strategies with different imaging methods in lung cancer follow-up (i.e. €282 per patient) [126].

5.4.3 Limitations

This study presents a few limitations. As a modelling study, the analysis inevitably represents a simplification of the real world. Furthermore, as common in Markov models, the model has no memory, meaning that the probability of moving to future states (e.g. from ‘palliatively treated recurrence’ to ‘death’) only depends on the present state and not on the sequence of events (i.e. no evidence of disease, salvage surgery or re-irradiation) which preceded it.

Second, there is no consensus regarding several of the model parameters. The risk of recurrence over the follow-up period varies in the literature between 20% and 50%. Based on a retrospective review of HNC patients undergoing a 5-year follow-up in Italy [248], a 35% risk of overall relapse (i.e. loco-regional and distant recurrences and second primaries) is assumed, 84% of which occur in the first 3 years. Another retrospective study, conducted in Norway, reports a similar figure (31% in 3 years) [270]. The rate of curability of loco-regional recurrences or second primaries according to different follow-up schemes has never been assessed in any comparative study; thus, the parameters used in this model (i.e. 50% vs. 25%) are based on previous literature [250] and clinical assumptions. The thirty-day postoperative mortality (i.e. 3.8%) is estimated from a single large study [252] which analysed the main causes of death in patients undergoing head and neck surgery. A comparable surgery-related mortality (i.e. 5%) is reported by a review article [267]. Similarly, treatment-related death following

re-irradiation is reported at 5% using data from a single study [253] but representing an ‘average’ effect estimate across the values reported in a recent review [26].

Third, cancer-related deaths are assumed to occur only during the active disease and patients without any cancer relapse after (primary or salvage) treatments experience the same risk of dying as the general population at the same age. However, an extra-mortality risk for HNC survivors has been shown by previous studies [271], mainly due to chronic liver disease and suicide. The model does not explore different ages at which patients can enter the ‘no evidence of disease’ state; the average age of 62 in the RCT is consistent with most studies reporting 60-65 years old HNC patients in the post-treatment follow-up or secondary treatment phases. Moreover, overall mortality is reported by 5-year age intervals in official statistics for Italy, thus varying this parameter in the sensitivity analysis would not significantly affect the results.

Fourth, the model does not account for any second primary sites other than the head and neck region, such as lung, oesophagus and colon as reported by previous studies [24] [248] [272]. Related to this, the HETeCo trial protocol is currently under review, with a low dose chest CT being included annually in heavy smokers according to the lung cancer screening NCCN guidelines [273]. Moreover, the possibility of a combined salvage treatment (e.g. surgery followed by re-irradiation), as reported in the literature [274] [275], is disregarded, as well as the use of radiotherapy with palliative intent [276] in recurrent HNC. Additionally, any re-treatment failures other than death, such as non-fatal toxicities leading to hospitalization or further treatments [26] [277] or residual disease after re-treatment, are considered in the analyses.

Fifth, this study does not account for patient’s anxiety and discomfort in undergoing intensive radiological investigations, nor the potential toxicities related to the use of PET and CT, which may reduce the cost-effectiveness of the intensive follow-up.

Moreover, the risk of false-positive imaging leading to further investigations or unnecessary treatments [248] [278] is not considered. The model represents an “ideal” post-treatment scenario where delays in performing visits and exams, and non-attendance at the recommended appointments on behalf of the patients are disregarded, although they clearly occur in clinical practice.

On the cost side, the analysis is conducted from the perspective of the Lombardy regional healthcare system; thus, the cost-effectiveness results might change if reimbursement tariffs of other Italian regions are adopted. However, these differences are expected to be minimal since the ICER and ICUR calculated using the national tariff fixed by the Ministry of Health (and to which Regions must abide in establishing their reimbursement tariffs) for the most expensive items in the model (i.e. head and neck surgery, CT, MRI, brain PET) are €12,265 and €18,647, respectively; IMRT has not been included yet in the national tariff set [279]. The study results are less generalizable to other countries, where clinical practice and healthcare costs may differ more considerably. Moreover, cost-effectiveness thresholds vary from country to country; for example, in an analogous economic evaluation study conducted in the Netherlands and comparing different follow-up strategies in lung cancer, the threshold was set at €80,000 per QALY gained [126], which is twice that recommended by Italian guidelines [262].

Furthermore, the analysis considers only direct healthcare costs, thus disregarding patient’s out-of-pocket costs (e.g. for frequent travel to the hospitals) and indirect costs (e.g. productivity losses due to travel time) that may be relevant during follow-up and salvage treatments. The cost of informal care by relatives and friends is expected to be considerable during the terminal disease stages; a survey on Italian cancer patients in their last three months of life estimated that 91% of them were cared for at home, but only 14% received assistance from domiciliary palliative care personnel [258].

Deterministic and probabilistic sensitivity analyses explore the robustness of the base-case results to deal with the inevitable modelling assumptions. A two-way sensitivity analysis evaluates the joint effect of varying the curability rate of recurrent events in arm A and arm B, since robust estimates of these parameters, representing the relative intervention effect, will only be available upon completion of the HETeCo trial. However, owing to the ICUR being considerably less than the cost-effectiveness threshold, it is unlikely that study conclusions would change even under alternative scenarios not evaluated in this study.

5.5 Conclusions

Until now, no comparative study has assessed the superiority of follow-up schemes using intensive radiological examinations over symptom-driven surveillance in HNC, in terms of clinical effectiveness and cost-effectiveness. This study presents an exploratory model to investigate alternative strategies for monitoring patients after completion of treatments for primary cancer from a health economics perspective, showing that an intensive surveillance scheme may well be cost-effective in Italy. Further research is needed to check these results in empirical studies or real-world settings, and a definite answer awaits, at a minimum, the completion of the HETeCo trial. The results are also influenced by the choice of the analysis's perspective; for example, using epidemiological and cost data from other European countries might yield alternative results in terms of ICER/ICUR and EVPI. The health and economic benefits of risk-adapted follow-up schemes based on clinical or demographic factors may also be assessed in future studies.

6 ELICITING PREFERENCES FOR CLINICAL FOLLOW-UP IN HEAD AND NECK CANCER PATIENTS USING BEST-WORST SCALING

6.1 Introduction

As reported in previous chapters, a follow-up program is essential shortly after the completion of treatment for HNC in order to identify potentially curable relapses. However, the optimal timing of visits and radiological assessments following treatment is debated by oncologists. Published recommendations are mostly informed by retrospective studies, expert opinions, and clinical practice rather than trial-based evidence [16] [32]. Until now, no consensus has been reached on the most appropriate follow-up modalities and timing in HNC patients. In addition to this clinical uncertainty, the patient's perspective has traditionally been neglected in designing cancer programs and elaborating clinical guidelines, whilst considering individual preferences might improve the feasibility, acceptability, and effectiveness of healthcare interventions [280].

Stated preference (discrete choice) techniques are a form of conjoint analysis which has been applied to a variety of settings to measure preferences for both market and non-market goods. This technique differs from the revealed preference method which uses observed data on actual choices made by individuals to measure preferences. The discrete choice technique relies on respondents making choices over hypothetical scenarios, which are described by a set of attributes and levels generated from an experimental design; in its original version, respondents are asked to choose the 'best' alternative within a set of two (or more) scenarios [281]. The use of stated preference methods was introduced in the early 1990s to value aspects of healthcare beyond health

outcomes and has been growing in recent years [282] [283] [284]. However, as for economic evaluations, limited evidence is available yet about preferences elicitation in oncological post-treatment surveillance.

The objective of this study is to quantify preferences for post-treatment surveillance in a large sample of patients treated for primary HNC, within the framework of the ongoing HETeCo trial which inspired the thesis.

6.2 Methods

This study used BWS to elicit patient's preferences for different aspects of follow-up after primary treatment for HNC.

6.2.1 Experimental design

The BWS choice experiment is a type of stated preference technique that is becoming increasingly popular in health economics. There exist three types of BWS studies in the literature, which differ in terms of the complexity of the alternatives under evaluation: the object case (case 1), the profile case (case 2), and the multi-profile case (case 3) [285]. The first is used to obtain the 'best' and 'worst' options within a list of items, which are not divided into levels and could be otherwise evaluated using a rating scale. Conversely, the second and third are considered as alternatives to traditional DCEs in which respondents evaluate one (case 2) or several (case 3) profiles at a time [285] [286]; as in DCEs, indeed, these methods require the identification of key elements (i.e. the attributes) each of which is split into two or more levels to create a series of scenarios described by different attribute-level combinations. However, instead of selecting a single profile in a choice set of two or more, in BWS (case 2) participants are asked to indicate which attribute-level they consider to be the 'best' and which to be the 'worst' (i.e. the 'BW pair') within each scenario. In the multi-profile case, respondents

are required to choose the best profile, as well as the worst one, in each choice set. In other words, in BWS, participants choose *‘the pair that exhibits the largest perceptual difference on an underlying continuum of interest’* [287] [288]. The present study used the profile case (case 2), which has been primarily applied in healthcare [285]; case 2, indeed, is particularly suitable in eliciting preferences for alternative features of a healthcare service but avoiding the complexity of case 3 that requires evaluation of multiple profiles at the same time.

The analysis was limited to the process-related aspects of the follow-up [289], as clinical outcomes of post-treatment surveillance in HNC are still under debate in the scientific community and under evaluation in the HETeCo trial. Relevant attributes and levels were established from literature review and expert opinion. In detail, common databases (PubMed and EMBASE) were searched using key terms such as “cancer” AND “follow up” AND “discrete choice experiment” (OR “best worst”) in title/abstract in order to identify studies that assessed patient’s preferences around post-treatment programs in oncology using stated preference methods. Moreover, the choice of attributes and levels was influenced by the features distinguishing the two surveillance strategies under investigation in the HETeCo trial to quantify the patient’s preferences for each of them and inform about the ‘value’ of these programs beyond effectiveness and cost-effectiveness, as traditionally defined.

Interviews with six patients during routine hospital visits were used to refine terminology and evaluate the comprehension and the acceptability of the BWS instrument.

6.2.2 Literature review

The few studies retrieved from the literature and quantifying preferences for cancer follow-up using state preference techniques are reported in Table 6.1. In details, two

binary DCE studies [290] [291] explored women's preferences for breast cancer follow-up services in the Netherlands and Australia, respectively. Face-to-face contacts were strongly preferred to telephone ones and a more intensive program of visits (every 3-6 months) was preferred over the less intensive options; moreover, women liked to be followed-up by a medical specialist and at specialized breast cancer clinics. A further binary experiment [292] recently carried out in Scotland for different types of cancer showed that survivors had a strong preference to see a consultant during a face-to-face appointment when receiving follow-up care. Lastly, a BWS study of post-treatment surveillance for soft tissue sarcoma in the UK concluded that patients typically preferred visits routinely consisting of a clinical examination and a chest X-ray, and secondary care- rather than general practice-based programs [293].

6.2.3 Questionnaire development

The preliminary work described so far allowed to eventually identify four attributes (from A to D): frequency and setting (hospital or mixed with primary care) of physical investigations; frequency of radiological assessments (MRI or CT); frequency (and eligibility) of PET scans; telephone calls to monitor the occurrence of new symptoms. Levels (from 0 to 2) were presented in order of increasing intensity of care and resources consumption for each attribute and connected to descriptions of the two HETeCo trial strategies (Table 6.2).

Table 6.1 Previous studies using stated preference techniques in cancer follow-up.

First author (year)	Country	Technique	Attributes (<i>Levels</i>)	Main findings
	Cancer	Sample size		
Murchie (2016) [292]	Scotland	DCE	<ul style="list-style-type: none"> - Healthcare provider (<i>consultant; registrar/trainee doctor; general practitioner; specialist nurse</i>) - Continuity of care (<i>yes; no</i>) - Contact mode (<i>face-to-face at hospital; face-to-face at general practitioner; telephone; videoconferencing/webcam/skype</i>) - Duration of appointments (<i>5 min, 10 min, 20 min, 30 min</i>) - Frequency of appointments (<i>every: 3, 6, 9, 12 months</i>) - Length of follow-up (<i>1 year; 2 years; 5 years; 10 years</i>) - Counselling (<i>no counselling; individual counselling; group counselling; family counselling</i>) - Additional services (<i>no additional services; personalized information pack; advice on complementary medicine; dietary advice</i>) 	Respondents overall preferred continuous, face-to-face consultant-led follow-up; they may accept non-consultant follow-up if compensated with changes elsewhere, notably greater continuity of care.
	Various (melanoma, breast, prostate, and colorectal)	668		
Damery (2014) [293]	England	BWS	<ul style="list-style-type: none"> - Length of follow-up (<i>5 years; 10 years; lifelong</i>) - Frequency of visits (<i>every: 3; 6; 12 months</i>) - Type of investigations (<i>clinical examinations; clinical examinations and x-ray; clinical examinations and MRI/CT scan</i>) - Healthcare provider (<i>general practitioner; specialist hospital nurse; specialist hospital doctor</i>) 	Patients typically prefer appointments routinely consisting of clinical examination and chest X-ray, and for follow-up to remain in secondary care rather than general practice.
	Soft tissue sarcoma	132		

Table 6.1 (cont.) Previous studies eliciting patients' preferences in cancer follow-up using stated preference techniques.

First author (year)	Country	Technique	Attributes (<i>Levels</i>)	Main findings
	Cancer	Sample size		
Bessen (2014) [290]	Australia	DCE	<ul style="list-style-type: none"> - Healthcare provider (<i>breast physician; general practitioner; breast cancer nurse</i>) - Frequency of visits (<i>every: 6; 9; 12 months</i>) - Location (<i>hospital clinic; general practice; local breast cancer follow-up clinic</i>) - Contact mode (<i>face-to-face; telephone; alternate between face-to-face and telephone</i>) - Drop-in clinics (<i>treatment side effects clinic; psychosocial support clinic; secondary prevention clinic</i>) 	In the absence of specialist follow-up, the most preferred scenario is a face-to-face local breast cancer follow-up clinic held every 6 months and led by a breast physician.
	Breast	722		
Kimman (2010) [291]	Netherlands	DCE	<ul style="list-style-type: none"> - Educational group program (<i>yes; no</i>) - Frequency of visits (<i>every: 3; 4; 6; 12 months</i>) - Waiting time (minutes) (<i>5; 30; 60; 90 minutes</i>) - Contact mode (<i>face-to-face; telephone</i>) - Healthcare provider (<i>medical specialist; breast care nurse/nurse practitioner; general practitioner; breast care nurse and medical specialist</i>) 	The medical specialist is the most preferred to perform the follow-up, but a combination of medical specialist and breast cancer nurse alternating is also acceptable to patients. Face-to-face contact is strongly preferred to telephone contact. Follow-up visits every three months are preferred over visits every four, six, or 12 months.
	Breast	331		

BWS: best-worst scaling; CT: computed tomography; DCE: discrete choice experiment; MRI: magnetic resonance imaging.

Table 6.2 Attributes and levels (in relation to the HETeCo trial arms).

Attributes	Levels	RCT
Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations (A)	0. Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	Neither
	1. Every 2-3 months for 2 years, every 5-6 months for 3 more years	Both arms
	2. Every 2-3 months for 5 years	Neither
Frequency of MRI/CT scans (B)	0. Only at the occurrence of new symptoms	Neither
	1. One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	Arm A
	2. Once or twice a year	Arm B
Frequency (and eligibility) of PET scans (C)	0. No PET scan during follow-up	Arm A
	1. Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	Arm B
	2. Yearly PET scan for all patients	Neither
Telephone calls to monitor occurrence of new symptoms (D)	0. No inter-visit calls from the hospital	Arm B
	1. Inter-visit calls by the nurse	Arm A
	2. Inter-visit calls by the oncologist	Neither

CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography.

A balanced study design was adopted in which each study attribute ($K=4$) had the same number of levels ($L_K = 3$). If an alternative contains K attributes, there are $K(K-1) = 4(4-1) = 12$ possible BW pairs the participant can choose within each scenario. As a full factorial design generating all possible attribute-level combinations ($3^4 = 81$ scenarios) was not feasible, a subset of 9 orthogonal scenarios (fractional factorial, main-effects design) was derived using the Hann and Shapiro catalogue, Master Plan 3 [294]. The total number of BW pairs in the orthogonal design was 108 ($12 \cdot 9$). This sub-group of selected scenarios preserved the properties of orthogonality (i.e. each attribute-level appears an equal number of times in combination with all other attribute-levels) and balance (i.e. each level within an attribute appears an equal number of times) (Table 6.3) [283] [295].

Table 6.3 Orthogonal scenarios.

Level	Attributes				
	A	B	C	D	All
0	3	3	3	3	12
1	3	3	3	3	12
2	3	3	3	3	12
All	9	9	9	9	36

By using an orthogonal design instead of a full factorial one, it is possible to avoid potential bias resulting from scenarios where, for example, all attributes are at the “top” or at the “bottom” level, as suggested by other authors [296]. Moreover, the two attribute-level combinations representing, respectively, the less intensive (i.e. arm A) and the more intensive (i.e. arm B) surveillance strategies in the HETeCo trial were purposely included among the nine orthogonal scenarios to subsequently calculate their overall utility based on the BWS data analysis.

The original questionnaire in Italian (and an English translation) is provided in the appendix (Annexes A6.1-2).

6.2.4 Recruitment and setting

Inclusion criteria for the experiment were broader than for the HETeCo trial. Patients aged 18 years and over, with a diagnosis of HNC in any anatomical site (except for the skin) in the last 5 years, who had completed any curative treatment at the NCI in Milan were eligible to participate. Patients were excluded if they were unable to comply with the study in the opinion of the clinical investigators, or they could not provide their informed consent. Moreover, this study excluded patients who underwent minor surgery for early stage cancer and subsequently did not attend a regular follow-up program in a multidisciplinary setting, i.e. with the contemporary presence of the head and neck surgeon, the radiation oncologist, and the medical oncologist. At the NCI, the routine follow-up program consists of outpatient visits every 2-3 months for the first 2 years after the end of treatment, then every 5-6 months for 3 more years. Radiological evaluations with MRI/CT scan are performed once 3 months after the end of treatment, then annually. PET is requested only in the case of doubtful imaging; no scheduled inter-visit contact is planned during the follow-up period.

The study was described to a consecutive sample of eligible patients during a routine follow-up appointment. Patients were reassured that responses to the questionnaire would not affect the care they were receiving at the hospital [297]. Those who agreed to participate were asked to sign a consent form and received the survey. Socio-demographic and clinical information were collected for each study participant. The questionnaire included a short rationale for the study and an explanation of the task required. After completing the BWS survey, patients were asked to answer some questions regarding their experience in performing the task. The experiment was approved by the NCI Ethical Committee in March 2015 and subsequently carried out as a cross-sectional survey between May and October of the same year.

6.2.5 Statistical analysis

Data on patients' characteristics were summarized through descriptive statistics; categorical variables were presented as percentages while continuous variables were presented as means and standard deviations. In regression analyses, missing demographic data were imputed using logical rules and information from related variables or, whenever this approach was not feasible, the most common value (i.e. the mode) [298]. Missing BW responses were imputed with the items most frequently selected as best and worst respectively within each scenario. The number of times each item was chosen as 'best' or 'worst' by the study participants was calculated. A best-minus-worst score was calculated by subtracting the number of times a feature was chosen as worst from the number of times it was chosen as best [280] [297].

Regression analysis was performed using a conditional logit model (*clogit* command in Stata) with cluster-adjusted (robust) standard errors [299]. BW pairs were treated as single variables and plotted as one data point at the individual level [293]. For each possible pair, the attribute-level was coded as 1 for the best and -1 for the worst; all

remaining attribute-levels were coded as 0. An example of data reporting is provided for one choice set scenario (i.e. A2B0C0D0, scenario 4) presented in the experiment (Table 6.4).

Table 6.4 Dataset example.

Patient	Scenario	Pair	Choice	A0	A1	A2	B0	B1	B2	C0	C1	C2	D0	D1	D2
1	1	1	0	0	0	1	-1	0	0	0	0	0	0	0	0
1	1	2	0	0	0	1	0	0	0	-1	0	0	0	0	0
1	1	3	1	0	0	1	0	0	0	0	0	0	-1	0	0
1	1	4	0	0	0	-1	1	0	0	0	0	0	0	0	0
1	1	5	0	0	0	-1	0	0	0	1	0	0	0	0	0
1	1	6	0	0	0	-1	0	0	0	0	0	0	1	0	0
1	1	7	0	0	0	0	1	0	0	-1	0	0	0	0	0
1	1	8	0	0	0	0	1	0	0	0	0	0	-1	0	0
1	1	9	0	0	0	0	0	0	0	1	0	0	-1	0	0
1	1	10	0	0	0	0	0	0	0	-1	0	0	1	0	0
1	1	11	0	0	0	0	-1	0	0	1	0	0	0	0	0
1	1	12	0	0	0	0	-1	0	0	0	0	0	1	0	0

The dependent variable took the value of 1 for the BW pair selected and 0 otherwise. In order to avoid a saturated model, the attribute-level that showed the lowest utility was used as the reference level; the omitted item took the value of zero on the utility scale and all estimates of the model were interpreted in relation to that. Therefore, each attribute-level can be positioned on an underlying preference scale (0; $+\infty$) starting with the reference item [285] [300]. Statistically significant coefficients indicated the importance of the attribute-level in determining overall utility [301].

A covariate-adjusted analysis was also performed to investigate sub-group preferences according to socio-demographic and clinical characteristics. A conditional logit model was run as previously described; however, interaction factors between selected covariates and choice outcomes (i.e. attribute-levels) were also added as independent variables [299]. In this model, interaction coefficients represent the additional utility of each attribute-level for the covariate [280]. A preliminary univariate regression analysis was performed to identify the demographic variables to be included in the final covariate-adjusted model as those displaying significant interaction terms ($p < 0.05$).

Variables with three (or more) categories in the questionnaire were dichotomized to increase the sample size within each group. Two age classes were generated around the median value (59 years) [280]. With regards to the clinical variables, the number of treatments received (i.e. one vs. more than one) was chosen as a ‘proxy’ of disease severity that, according to the clinicians involved in the study, might influence patient’s preferences in follow-up; the time from the end of treatments was equally considered clinically relevant. Any other clinical information was disregarded in this analysis. A covariate-adjusted regression was also separately run using the patient-reported difficulty level in performing the BWS task as an interaction factor. To run all covariate-adjusted models, additional columns with socio-demographic variables were added to the original database (Table 6.4).

Lastly, following the approach of a previous study [302], an overall utility for each hypothetical follow-up scheme deriving from the experiment was obtained by summing up the level coefficients from conditional logit regression.

All data were analysed using Stata version 14 (Stata Corp, 2015).

6.3 Results

6.3.1 Sample characteristics

A total of 162 consecutive patients who met the inclusion criteria were approached to participate in the survey; however, sixteen declined resulting in a response rate of 90%. Three questionnaires were excluded from data analysis, as they were not completed correctly or in full. Therefore, the final sample comprised 143 patients, of whom 74% were male. Socio-demographic and clinical features of the participants are presented in Table 6.5. The mean age of participants was 57.6 (± 12.1) years and more than one third

of patients were retired (34.2%). The great majority of patients (85.3%) lived with family and 64.3% less than 100 km from the hospital.

A variety of primary tumour diagnoses were observed in the sample, with the most common being oropharyngeal (38.4%), nasopharyngeal (28.0%) and laryngeal cancer (11.2%), mostly in a locally advanced stage (III and IV; 93.7%). Most patients (38.5%) received a combination of chemotherapy and radiation as primary treatment for HNC, or chemotherapy followed by the combined therapy (30.0%). Participants were equally distributed according to time since the end of treatments as follows: ≤ 2 years, 51.0%; > 2 years, 49.0%.

6.3.2 Best and worst choice counts

Frequency counts provide summary estimates of best and worst choices made by participants (Table 6.6). Of a total of 2,574 expected BW responses, only 12 (0.5%) were missing and imputed as previously explained. The highest ranked attribute-level was “physical investigations performed every 2-3 months for 2 years, then every 5-6 months for 3 more years”. The lowest rated feature is less clearly identifiable. According to the best-minus-worst score, the lowest valued attribute-level was “inter-visit calls by the nurse” to monitor patient’s health status. “No PET scan during follow-up” was the item least frequently chosen as “best”, while “primary care-based follow-up during the last 2 years” was that most often indicated as “worst”.

Table 6.5 Sample's characteristics.

Variable	Class	Frequency	
		N	%
<i>Socio-demographic data</i>			
Gender	Male	106	74.1
	Female	37	25.9
Age (years)		Mean (\pm SD)	57.6 \pm 12.1
Age at diagnosis (years)		Mean (\pm SD)	54.9 \pm 12.3
Employment status	Full-time employed	42	29.4
	Part-time employed	8	5.6
	Self-employed	28	19.6
	Retired	49	34.2
	Unemployed	7	4.9
	Other	7	4.9
	<i>Missing</i>	2	1.4
Educational level	Primary School	46	32.1
	Secondary School	57	39.9
	University	25	17.5
	Post-University	9	6.3
	<i>Missing</i>	6	4.2
Living status	Alone	15	10.5
	With family	122	85.3
	<i>Missing</i>	6	4.2
Distance from home (Km)	<100	92	64.3
	100-500	15	10.5
	>500	33	23.1
	<i>Missing</i>	3	2.1
<i>Clinical data</i>			
Site of primary tumour	Oropharynx	55	38.4
	Nasopharynx	40	28.0
	Larynx	16	11.2
	Oral cavity	12	8.4
	Sino nasal cavity	5	3.5
	Salivary glands	4	2.8
	Other	11	7.7
Grade of primary tumour	Early stage	9	6.3
	Locally advanced	134	93.7
HPV status of disease	Positive	47	32.9
	Negative	10	7.0
	Not applicable	86	60.1
Treatment(s)	CTRT	55	38.5
	CT + CTRT	40	30.0
	Surgery + RT	13	9.1
	Surgery + CTRT	12	8.4
	RT	9	6.3
	CT + RT	4	2.8
	Other	10	7.0
Number of treatments	1	66	46.1
	2	71	49.7
	3 or 4	6	4.2
Time from treatment end	\leq 2 years	73	51.0
	>2 years	70	49.0

CT: chemotherapy; CTRT: combined chemo-radiotherapy; HPV: human papilloma virus; RT: radiotherapy; SD: standard deviation.

Table 6.6 Frequency of best and worst selections across scenarios.

Attribute	Level	Best	Worst	Best-Worst
Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations	Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	109	194	-85
	Every 2-3 months for 2 years, every 5-6 months for 3 more years	345	14	331
	Every 2-3 months for 5 years	278	35	243
Frequency of MRI/CT scans	Only at the occurrence of new symptoms	64	120	-56
	One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	61	111	-50
	Once or twice a year	182	16	166
Frequency (and eligibility) of PET scans	No PET scan during follow-up	7	167	-160
	Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	74	78	-4
	Yearly PET scan for all patients	103	84	19
Telephone calls to monitor occurrence of new symptoms	No inter-visit calls from the hospital	9	172	-163
	Inter-visit calls by the nurse	21	185	-164
	Inter-visit calls by the oncologist	34	111	-77

CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography.

6.3.3 Conditional logistic regression analysis

The logistic regression results are presented in Table 6.7. The attribute-level with the lowest utility coefficient was “not performing any PET scan during follow-up” and was assumed as the reference level. The regression coefficients of BW pairs show the additional utility of each attribute-level over the reference case. As already observed in frequency counts, the feature showing the highest utility was “physical investigations performed every 2-3 months for 2 years and every 5-6 months for 3 more years”. A more intensive frequency of visits (“every 2-3 months for 5 years”) ranked second, and “MRI/CT scan performed once or twice a year” ranked third. In contrast, the attribute-levels with the lowest utility were “follow-up based at primary care during the last 2 years”, “inter-visit calls by the nurse” and “no inter-visit calls from the hospital” in that order; however, none of them was statistically significant compared to the reference level. For each individual attribute, the distance between the most and the least preferred levels is an indication of the relative importance of that attribute to respondents [285] [300]. In this survey, the “frequency and setting of physical and larynx/pharynx endoscopic investigations” is the item with the largest difference

between level coefficients (2.482 i.e. 2.523 minus 0.041) and, thus, the greatest impact on patients' utilities.

Table 6.7 Utility coefficients from paired conditional logistic regression analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.041	0.194	0.834	-0.340	0.422
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.523	0.115	<0.001	2.297	2.749
Every 2-3 months for 5 years	2.155	0.141	<0.001	1.877	2.432
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.526	0.112	<0.001	0.3055	0.746
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.565	0.122	<0.001	0.326	0.804
Once or twice a year	1.885	0.099	<0.001	1.691	2.079
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-	-	-	-	-
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.962	0.109	<0.001	0.748	1.176
Yearly PET scan for all patients	1.111	0.138	<0.001	0.840	1.382
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.092	0.093	0.322	-0.091	0.276
Inter-visit calls by the nurse	0.087	0.112	0.439	-0.133	0.307
Inter-visit calls by the oncologist	0.601	0.115	<0.001	0.375	0.826

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

6.3.4 Covariate-adjusted regression analysis

Table 6.8 provides results from the conditional logistic regression analysis after adjusting for selected clinical and demographic data. Educational level (more educated i.e. university, post-university =1; less educated i.e. primary school, secondary school =0), employment status (employed i.e. full-time employed, part-time employed, self-employed =1; not employed i.e. retired, unemployed, other =0), living status (with family=1; alone =0), time in follow-up (>2 years =1; ≤ 2 years =0) and number of treatments (≥ 1 i.e. 2, 3 or 4 =1; one only =0) which displayed significant interactions in univariate regression analysis (Tables A6.1-8) were included in the final model. Conversely, no significant interaction coefficients were found with respect to age (age ≥ 59 =1; age < 59 =0), gender (female =1; male =0) and distance from home (≥ 100 km =1; < 100 km =0).

Table 6.8 Utility coefficients from covariate-adjusted conditional logistic regression analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.178	0.935	0.849	-1.654	2.010
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.303	0.494	<0.001	1.336	3.271
Every 2-3 months for 5 years	1.810	0.600	0.003	0.634	2.985
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.382	0.467	0.414	-0.534	1.297
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.392	0.383	0.306	-0.359	1.142
Once or twice a year	1.590	0.492	0.001	0.625	2.556
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.556	0.317	0.079	-1.177	0.064
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.631	0.457	0.167	-0.264	1.527
Yearly PET scan for all patients	0.877	0.516	0.089	-0.134	1.887
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.164	0.331	0.621	-0.813	0.486
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.414	0.269	0.124	-0.113	0.941
Interactions (Educational level)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	1.700	0.559	0.002	0.605	2.795
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.285	0.383	0.458	-0.467	1.036
Every 2-3 months for 5 years	0.472	0.397	0.235	-0.306	1.250
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.081	0.300	0.786	-0.507	0.670
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.214	0.293	0.465	-0.788	0.360
Once or twice a year	0.756	0.339	0.026	0.092	1.420
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.757	0.263	0.004	0.240	1.273
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.372	0.264	0.159	-0.146	0.891
Yearly PET scan for all patients	0.342	0.412	0.407	-0.466	1.150
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.337	0.251	0.180	-0.156	0.829
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.157	0.213	0.461	-0.261	0.576

Table 6.8 (cont.) Utility coefficients from covariate-adjusted conditional logistic regression analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
Interactions (Job)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.132	0.418	0.752	-0.951	0.688
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.652	0.278	0.019	0.106	1.198
Every 2-3 months for 5 years	0.755	0.321	0.019	0.126	1.384
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.654	0.271	0.016	0.124	1.185
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.338	0.266	0.203	-0.183	0.860
Once or twice a year	0.741	0.255	0.004	0.241	1.240
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.164	0.225	0.464	-0.276	0.605
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.463	0.252	0.067	-0.031	0.957
Yearly PET scan for all patients	0.418	0.329	0.205	-0.228	1.063
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.412	0.211	0.051	-0.002	0.827
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.063	0.185	0.732	-0.299	0.425
Interactions (Living status)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.400	0.820	0.626	-1.209	2.008
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.352	0.498	0.480	-0.624	1.327
Every 2-3 months for 5 years	0.465	0.586	0.427	-0.683	1.613
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.077	0.394	0.846	-0.696	0.850
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.450	0.309	0.146	-0.156	1.056
Once or twice a year	0.181	0.471	0.700	-0.742	1.105
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.732	0.282	0.009	0.180	1.284
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.134	0.395	0.735	-0.640	0.908
Yearly PET scan for all patients	0.578	0.490	0.238	-0.382	1.538
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.066	0.302	0.826	-0.526	0.659
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.337	0.213	0.114	-0.081	0.756

Table 6.8 (cont.) Utility coefficients from covariate-adjusted conditional logistic regression analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
Interactions (Time in follow-up)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.870	0.404	0.031	-1.662	-0.079
Every 2-3 months for 2 years, every 5-6 months for 3 more years	-0.466	0.251	0.063	-0.958	0.025
Every 2-3 months for 5 years	-0.452	0.298	0.129	-1.035	0.131
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	-0.197	0.256	0.442	-0.700	0.306
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.136	0.252	0.590	-0.631	0.359
Once or twice a year	-0.592	0.238	0.013	-1.058	-0.125
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.481	0.223	0.031	-0.919	-0.043
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	-0.239	0.243	0.326	-0.716	0.238
Yearly PET scan for all patients	-0.764	0.311	0.014	-1.374	-0.154
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.361	0.192	0.061	-0.739	0.016
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	-0.649	0.173	<0.001	-0.988	-0.309
Interactions (Number of treatments)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.805	0.421	0.056	-1.632	0.021
Every 2-3 months for 2 years, every 5-6 months for 3 more years	-0.490	0.241	0.042	-0.962	-0.017
Every 2-3 months for 5 years	-0.672	0.282	0.017	-1.225	-0.119
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	-0.492	0.267	0.066	-1.016	0.032
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.675	0.253	0.008	-1.171	-0.180
Once or twice a year	-0.306	0.236	0.196	-0.770	0.158
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.404	0.221	0.068	-0.838	0.030
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	-0.132	0.244	0.588	-0.612	0.347
Yearly PET scan for all patients	-0.487	0.310	0.116	-1.095	0.121
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.050	0.188	0.789	-0.419	0.319
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.081	0.175	0.643	-0.261	0.423

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

The interpretation of regression results is facilitated through the example of education. There were statistically significant differences between education groups with respect to three of the four attributes in the experiment. The total utility of “MRI/CT scan performed once or twice a year” for highly educated patients is the sum of the attribute-level coefficient (1.590) and its interaction term coefficient with educational level (0.756), which gives 2.346. The corresponding utility for less educated people is the coefficient without interaction (1.590). Thus, it is possible to infer that all patients like a more intensive radiological investigations program; however, this preference is stronger for those with more education. Furthermore, highly educated patients are more likely to prefer a primary care-based follow-up than those with a lower education level. The total utility of this item, indeed, is equal to 1.878 for the former, while not significantly different from the reference value for the latter (0.178). The last significant interaction is with “no PET scan during follow-up”; the overall utility for more educated patients is 0.201 whilst not significantly different from zero for the less educated ones.

In a similar way, it is possible to calculate separate utilities for different groups of patients according to the remaining four covariates selected within the univariate analysis.

6.3.5 Scenario’s utilities

The calculation of scenario’s utilities consisted in summing up the coefficients for each attribute-level combination resulting from the (unadjusted) conditional logit model in the overall sample. The same approach can be used to obtain utilities for specific subgroups using results from the covariate-adjusted model. By summing the level coefficients taken one at a time within each attribute, the most preferred hypothetical scenario (overall utility: 6.120) across the sample would be a hospital-based follow-up with frequency of visits decreasing over time (i.e. every 2-3 months for 2 years, every 5-

6 months for the next 3 years), radiological assessments (i.e. MRI/CT) performed once or twice a year, yearly PET scan for all patients (irrespective of individual risk of recurrences) and inter-visit calls by the oncologist to monitor the occurrence of new symptoms. On the contrary, the least desirable option is a mixed hospital-/primary care-based surveillance with MRI/CT scan performed only at the occurrence of new symptoms, no PET scan scheduled during follow-up period and inter-visit calls by the nurse to check the patient's health (scenario utility: 0.654).

Among the nine scenarios included in the experiment (i.e. the fractional factorial design, Annexes A6.1-2), the utilities associated with the two follow-up programs (i.e. arm A and arm B) under evaluation in the HETeCo trial are of interest. The one corresponding to arm B of the trial (scenario 3) obtains the highest utility (i.e. 5.462) in the overall sample, whilst the estimated utility for the less intensive follow-up (arm A, scenario 2) is substantially lower (i.e. 3.175). The lowest valued scenario (i.e. 1.616) is the one with primary-care based physical investigations during the last two program years, radiological assessments (MRI/CT) performed only at the occurrence of new symptoms, yearly PET scan only for high-risk patients and inter-visit calls by the nurse to monitor the occurrence of new symptoms (Table 6.9).

Table 6.9 Utilities of the survey scenarios (including the HETeCo trial arms).

	Attribute-levels	Coefficient	Utility
SCENARIO 1	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for the first 2 years and every 5-6 months for the last 3 years	2.523	4.761
	Radiological assessments (MRI/CT) <i>only</i> at the occurrence of new symptoms	0.526	
	Yearly PET scan for <i>all</i> patients (irrespective of age or other risk factors)	1.111	
	Inter-visits call by the <i>oncologist</i> to monitor new symptoms occurrence	0.601	
SCENARIO 2 (ARM A)	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for the first 2 years and every 5-6 months for the last 3 years	2.523	3.175
	One radiological assessment (MRI/CT) only at the beginning of follow-up (later only at occurrence of new symptoms)	0.565	
	No PET scan during follow-up	0.000 (ref.)	
	Inter-visits call by the nurse to monitor new symptoms occurrence	0.087	
SCENARIO 3 (ARM B)	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for the first 2 years and every 5-6 months for the last 3 years	2.523	5.462
	Radiological examinations (MRI/CT) once or twice a year	1.885	
	Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.962	
	No inter-visit calls from the hospital to monitor new symptoms occurrence	0.092	
SCENARIO 4	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 5 years	2.155	2.773
	Radiological assessments (MRI/CT) only at the occurrence of new symptoms	0.526	
	No PET scan during follow-up	0.000 (ref.)	
	No inter-visit calls from the hospital to monitor new symptoms occurrence	0.092	
SCENARIO 5	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 5 years	2.155	4.283
	One radiological assessment (MRI/CT) only at the beginning of follow-up (later only at occurrence of new symptoms)	0.565	
	Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.962	
	Inter-visits call by the oncologist to monitor new symptoms occurrence	0.601	
SCENARIO 6	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 5 years	2.155	5.238
	Radiological examinations (MRI/CT) once or twice a year	1.885	
	Yearly PET scan for all patients (irrespective of age or other risk factors)	1.111	
	Inter-visits call by the nurse to monitor new symptoms occurrence	0.087	
SCENARIO 7	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 3 years. Primary care-based follow-up for the last 2 years.	0.041	1.616
	Radiological assessments (MRI/CT scan) only at the occurrence of new symptoms	0.526	
	Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.962	
	Inter-visits call by the nurse to monitor new symptoms occurrence	0.087	
SCENARIO 8	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 3 years. Primary care-based follow-up for the last 2 years.	0.041	1.809
	One radiological assessment (MRI/CT scan) only at the beginning of follow-up (later only at occurrence of new symptoms)	0.565	
	Yearly PET scan for all patients (irrespective of age or other risk factors)	1.111	
	No inter-visit calls from the hospital to monitor new symptoms occurrence	0.092	
SCENARIO 9	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 3 years. Primary care-based follow-up for the last 2 years.	0.041	2.527
	Radiological examinations (MRI/CT scan) once or twice a year	1.885	
	No PET scan during follow-up	0.000 (ref.)	
	Inter-visits call by the oncologist to monitor new symptoms occurrence	0.601	

CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography.

6.3.6 Patients' evaluation of the experiment

Table 6.10 presents data on patients' self-reported difficulties in understanding and completing the questionnaire. The average completion time was 9.2 (\pm 3.1) minutes. Nearly half of the participants rated the BWS task very easy to perform (i.e. level 1; 45.4%) and did not need any support from healthcare professionals or family members (44.7%). More than one-third (37.1%) reported no difficulties during completion among the options available; "understanding the task" was the most common difficulty (21.7%) followed by "length of the questionnaire" (6.3%) and "technical/scientific language" (5.6%). A further 14% indicated other difficulties mainly related to indecision in selecting the BW pair and the feeling that scenarios were too repetitive. The difficulty level (i.e. equal to 1 (low)=0; >1=1) used as interaction factor with BW attribute-levels did not yield any significant results ($p < 0.05$) in univariate regression analysis, suggesting that its impact on participants' responses was negligible (Tables A6.9).

Table 6.10 Self-reported difficulties in performing the exercise.

		N	%
Completion time	\leq 5 minutes	33	23.1
	5-10 minutes	87	60.8
	> 10 minutes	14	9.8
	<i>Missing</i>	9	6.3
Difficulty level	1 (low)	65	45.4
	2	28	19.6
	3	32	22.4
	4	8	5.6
	5 (high)	7	4.9
	<i>Missing</i>	3	2.1
Support needed (e.g. nurse, caregiver)	Yes	63	44.1
	No	64	44.7
	<i>Missing</i>	16	11.2
Main difficulty	None	53	37.1
	Length of the questionnaire	9	6.3
	Understanding the task	31	21.7
	Technical/scientific language	8	5.6
	Other	20	14.0
	<i>Missing</i>	22	15.3

6.4 Discussion

Few studies have explored patient's preferences for delivery of post-treatment cancer programs and even fewer have attempted to derive utility estimates from them. A non-systematic literature review identified a total of four studies using stated preference techniques, of which only one adopts the BWS methodology and none addresses HNC; overall, these studies confirm that patients tend to prefer an intensive hospital-based post-treatment surveillance including frequent face-to-face visits with specialized doctors and radiological examinations.

A further study [280] adopting the BWS methodology was identified even if not strictly related to follow-up but addressing a symptom supporting care intervention in lung cancer patients after completion of first line therapies. With respect to HNC, a non-DCE survey only on patients' view of their follow-up regimen was conducted in UK. The study revealed that most patients felt their follow-up visits too frequent and were in favour of a less intensive, symptom-driven follow-up [303].

The BWS method is argued to have several advantages over traditional DCEs [280] [300]. First, respondents are provided with profiles one by one rather than two (or more) at a time; thus, BWS is considered less cognitively demanding for participants [288] [296] [299] [301]. These expectations were confirmed in this study by the self-reported judgement on the choice task, which was graded as simple and quick by the majority of respondents. Moreover, BWS may elicit more information than traditional DCEs, as respondents make choices within profiles rather than between profiles; in particular, in BWS a single attribute-level combination acts as benchmark, instead of a whole scenario. In this way, it is possible to calculate utility coefficients for each item in the experiment, which may be useful in evaluating different elements of a healthcare service [288] [296] [299] [301]. Lastly, profile-based (case 2) BWS was selected in

preference to the traditional binary DCE because, in a life-threatening condition like HNC, it was anticipated that patients would always select the option which they thought would maximize survival and consequently less information would be generated on how other attribute/levels are valued.

This study is the first stated preference survey of HNC follow-up and, in Italy, of any cancer surveillance. The survey aimed at providing insights into patients' views on post-treatment monitoring in this cancer population using BWS methodology. Moreover, a covariate-adjusted analysis was performed to investigate socio-demographic or clinical characteristics related to the choice of attribute-levels. It was not surprising to find that patients' preferences for HNC follow-up were generally aligned with the scheme currently adopted by NCI where the study was conducted. This tendency has been described as the 'lure of the familiar' [293], meaning that individuals are likely to stick with they have already experienced, even if potentially unsatisfactory. Participants in this study revealed clear preferences for follow-up to remain in secondary care, even during the last phases of the program. Intensive radiological examinations (once or twice a year) were strongly preferred. Inter-visit telephone calls were generally disliked, especially when performed by healthcare professionals other than medical doctors. These results are in contrast with those found by a previous study on patients' preferences in HNC follow-up [303]; however, that survey was conducted in a different geographical setting (highly deprived areas of London) and without relying on stated preference methods. Differences in preferences according to individual characteristics were also found. Overall, highly educated patients were more likely to prefer primary-care based follow-up and intensive MRI/CT radiological investigations but avoiding PET scan. Patients with a job tended to prefer more frequent visits to the hospital but no inter-visit calls, while those living with family revealed a stronger preference for not performing any PET scan during follow-up. Patients two years (or more) following

treatment expressed a lower utility for more intensive MRI/CT investigations and were keener to avoid inter-visit telephone calls with their clinicians. Patients who had received more than one treatment option (e.g. surgery followed by radiotherapy) were less keen to accept a symptom-driven radiological surveillance and to travel frequently to the hospital for physical investigations. Conversely to a previous study [280], no differences were found in age or gender with regards to preferences for delivering a post-treatment intervention in cancer care; however, the program under evaluation was considerably different. In terms of hypothetical follow-up utilities, the experiments showed an overall preference for more intensive scenarios, including arm B of the ongoing HETeCo trial.

This study has a number of limitations. First, the data collection was restricted to only one centre that, due to some distinctive features (i.e. high specialization, commitment to research, innovative technologies), may not be representative of a typical cancer clinic in Italy. Moreover, patients attending the NCI, especially those coming from afar, are likely to be more educated, wealthy or health conscious than the general HNC population. However, the NCI is the leading centre of the HETeCo trial with experience in cancer follow-up research, and the referral to a single centre reduced bias related to different ways of administering the survey and providing support during the completion of the questionnaire. Second, the cognitive ability of each participant with respect to completing the task was not evaluated, and on some occasions the patient was supported by an accompanying person. Moreover, the attribute-level descriptions were sometimes long, included multiple concepts and involved technical terms. Nevertheless, given the very low number (n=3) of questionnaires excluded from the analysis and the limited self-reported difficulties, it is likely that the task was feasible for most participants; moreover, regression analyses did not show any significant interactions between BW choices and difficulty level. The final limitations concern the restricted range of

hypothetical follow-up programs that can be valued owing to the small number of attributes-levels and the assumption of no interaction between BWS items. However, the small number of items included in the experiment, as well as the use of a small factorial main-effects design (i.e. 9 out of 81 scenarios), was justified by feasibility considerations.

6.5 The use of best-worst scaling utilities for health economic comparison: fact or fiction?

The interest in collecting information regarding patient's preferences in healthcare using rigorous stated preference methods is confirmed by the increasing number of studies that can be found in the health economics literature; a recent study reported 53 BWS applications in health and healthcare [304]. Until now, DCEs have been mainly applied to elicit patient's preferences and quantify trade-off among alternative treatments described by hypothetical scenarios. However, there is little guidance on how DCE/BWS preference and cost data can be combined in cost-effectiveness analysis to inform healthcare decisions [289].

Conversely to HSUVs derived from EQ-5D (and other generic HRQoL measures) or direct techniques (e.g. SG and TTO), which range between 0 (death) and 1 (perfect health), there exist no standard scales on which DCE-based utilities are measured [289]. In this chapter, utilities for each attribute-level are obtained from a conditional logit model (*clogit* in Stata), which calculates them by summing up the number of times each item is selected as best or worst [301]. In theory, BWS-derived utility coefficients can vary between zero (i.e. the reference case representing the lowest valued item) and $+\infty$. However, the highest utility obtained in the (unadjusted) conditional logit model is 2.523, corresponding to the most preferred item of the experiment (i.e. a hospital-based program of physical investigations with frequency decreasing over time). A similar

study [293] quantifying preferences for soft tissue sarcoma follow-up using BWS and conditional logit regression obtained utilities confined in a comparable range (0; 2.503).

BWS studies can be classified into two main categories according to their area of application, i.e. evaluation of HSUVs/HRQoL and evaluation of healthcare interventions. In the first case, profiles are derived from different combinations of responses to a HRQoL questionnaire (e.g. EQ-5D-5L identifies a total of 3,125 unique health state profiles) and regression analyses furnish utility coefficients for each attribute-level combination (e.g. mobility, level 2). Few examples are available in the BWS literature; among them, Ratcliffe et al. [302] obtained HSUVs from the Child Health Utility-9D (CHU9D), which is a newly developed generic preference-based instrument to measure HRQoL in children and adolescents. In this study, utility coefficients from conditional logit regression are rescaled onto the 0-1 range of HSUVs and summed up to generate utilities for each possible health state defined by the CHU9D. This study presents the feasibility of BWS approach to evaluate HSUVs for QALY calculation and cost-utility analyses, although further research is needed to explore the advantages and limitations of this approach.

The way of combining utility estimates from the second group of BWS studies (i.e. those assessing preferences for healthcare interventions) and costs for economic evaluation remains unclear. In 2006, McIntosh [305] proposes an initial framework to perform cost-benefit analyses using DCE-derived utilities. A subsequent study from Benning and colleagues [289] presents a methodology to combine individual-specific preference data obtained from a traditional (binary) DCE with cost data to inform about the cost-effectiveness of customized care compared to standardized care. A more recent study [284] attempts to incorporate patients' preferences into an economic evaluation and compare results with the standard cost per QALY approach using data from a RCT;

two other studies [306] [307] included DCE within a trial, but without collecting comparable QALY data.

A simpler strategy can be used here to allow a cost-utility comparison of follow-up programs (A vs. B) under investigation in the HETeCo trial combining scenario utilities from the BWS experiment (Table 6.9) and the programs' 5-year cost estimates calculated in Chapter V (Table 5.5), as shown in Table 6.11.

Table 6.11 Cost-utility comparison using the survey results.

	Costs (€)	Δ costs (€)	BWS utility	Δ BWS utility	Δ costs (€)/Δ BWS utility
Follow-up A	729.0		3.175		
Follow-up B	3,676.5	2,947.5	5.462	2.287	1,288.8

This approach presents several limitations mainly related to the interpretation of results. First, there exists no standard scale for DCE utilities, thus they can be compared within the same experiment only, and not across different studies. Second, conversely to traditional cost-utility analyses where country-specific guidelines exist to state the value of a QALY gained, this approach does not furnish a decision rule (i.e. a threshold) to state the cost-effectiveness of an intervention compared to the others in case differences in costs and utilities are both positive or negative (i.e. non-dominant comparisons). Third, in DCE/BWS surveys, the patient is rarely asked to value health-related aspects, but more frequently process-related aspects of an intervention (e.g. frequency of radiological imaging), although it is not stated how this information should be considered in medical decision-making and allocation of healthcare budgets [38]. Moreover, some authors argue that 'value' should depend on 'outcome' and not on the 'process' of care that led to the outcome achievement [35]. Forth, utilities derived from DCEs/BWS experiments are generally obtained from cross-sectional surveys, thus preventing long-term comparison of costs and effects. These limitations suggest that

currently no valuable alternative can replace the cost per QALY approach, which still represents the milestone of cost-utility comparisons in healthcare.

6.6 Conclusions

In recent years, there has been growing interest in using preference elicitation methods to inform health policy and medical decision-making. Incorporating patient's preferences into the treatment and follow-up strategies may help in tailoring healthcare to the patient and increasing adherence to treatment [18]. In HNC follow-up, patients seem to be reassured by a regular follow-up with scheduled imaging and expertise of specialists, as already reported in other experiences [308]. The present study highlighted patients' limited interest in alternative ways of delivering post-treatment services, such as symptom-driven surveillance, telephone monitoring or non-specialist follow-up. Healthcare professionals (e.g. general practitioners or nurses) other than specialist doctors were probably considered not skilled enough to conduct cancer follow-up. There might be a resistance to change from established to new types of service without adequate reassurance from the clinicians. In particular, patients with less education may benefit least from a patient-initiated follow-up owing to difficulties in understanding medical instructions. Overall, there is a need for improved communication for cancer patients to evaluate consciously the post-treatment phase and to promote self-managed symptoms monitoring [309]. Patients likely prefer intensive radiological assessment because of fear of disease recurrence; however, tests should be performed for clinical reasons and not (only) for patient's reassurance. The long-term effects associated with frequent and prolonged radiological scans should also be considered. In this regard, more efforts should be spent in order to identify the most cost-effective follow-up scheme in HNC, thus providing the scientific community and patients with evidence-based programs. The HETeCo trial comparing health and economic outcomes in this

setting is ongoing and, in terms of patient's preferences, the analysis presented in this chapter seems to advantage the more intensive follow-up scheme (arm B). Finally, differences in preferences were found according to the intensity of treatments received and the time already spent in follow-up; these results might justify a provision of different surveillance schemes based on these clinical variables, as already suggested by guidelines in the field [20]. Similarly, inter-visit calls appear to be more valuable in the initial phases of the follow-up than in the final ones, when patients may feel more confident of beating cancer.

Overall, this study provided useful insights into individual preferences for several aspects of post-treatment surveillance in HNC in Italy. Additional elements might be explored in the future, such as the level of scientific evidence, co-payment for extra-investigations and late side effects of intensive investigations. Currently, there is evidence of heterogeneity in preferences with respect to a limited number of patient's characteristics. More research also considering the costs of different follow-up regimens is required to justify the provision of customized follow-up programs in HNC patients. From a methodological perspective, further research is needed to evaluate the feasibility of incorporating preferences information into traditional economic evaluations.

7 DISCUSSION AND CONCLUSIONS

7.1 Introduction

This thesis values follow-up programs in HNC from a health economic perspective. Post-treatment surveillance is recognized to be a valuable service in oncology, although the best way of monitoring patients after the primary treatment is completed is often uncertain. In HNC, several guidelines exist, although most of them are based on routine practice or expert opinion. Additionally, the clinical studies present conflicting results about the appropriate frequency of radiological investigations that should be prescribed in the post-treatment phase [20]. In Italy, follow-up is usually quite intensive, according to the recommendations of national scientific societies and individual cancer centres initiatives. However, the added ‘value’ of intensive follow-up programs has never been proved with rigorous methods.

More generally, what is ‘value’ in healthcare and, especially, in oncology, is still debated in the literature. The QALY is a measure of health effectiveness that is widely used to inform the allocation of healthcare resources. In the traditional QALY approach, value is assessed in terms of preference or desirability for a given health state [36]. Several HTA agencies, including NICE in England, have endorsed the QALY as a standard measure to promote comparability in cost-utility analyses across different disease areas and treatments. However, the measurement of QALY as a ‘proxy’ of value in cancer poses some challenges reported in a recent paper by Devlin and Lorgelly [38]. The authors highlight that an overall survival estimate is required to calculate QALYs, whilst oncology trials rarely continue for long enough to capture that; thus, modelling has become an unavoidable approach to conduct cost-utility analyses in many cancers that have become chronic diseases. Moreover, as a measure combining length and

quality of life, the QALY requires utility weights that are infrequently directly available from clinical studies, where cancer-specific HRQoL tools, not provided with a preference-based algorithm, are usually preferred. Thus, “mapping” is growing in popularity to obtain HSUVs from non-preference-based measures. In addition, the QALY concept endorses the preference for a given health state, expressed as HSUV, but formally disregards preferences for the process of care; although information on patients’ desires are increasingly recognized as important for policy-makers, there is little consensus on how HTA can incorporate these data alongside QALYs [38].

In oncology, a few alternative value frameworks [51] [52] [53] [54] [55] have recently been developed, with the intent of evaluating cancer drugs through a set of meaningful attributes. Much emphasis is given to safety and clinical efficacy, whilst HRQoL and patients’ preferences are usually not included or given much weight. Most of these tools, indeed, have been developed by referring to RCTs only, thus disregarding the evidence generated by other study types. Moreover, there is no agreement across the different frameworks regarding which dimensions should be considered, and how they should be incorporated and weighted into the tool [56]. Lastly, aiming at specifically evaluating cancer pharmaceuticals, these frameworks do not allow a value comparison across different types of healthcare interventions. Thus, despite some limitations recognized in the literature, the cost per QALY remains central to value comparisons in healthcare and is endorsed by the main body of this thesis. Since it is not established how to enlarge the definition of QALY to incorporate additional elements of ‘value’ such as patients’ preferences, these are explored in an independent piece of work.

This PhD thesis consists of a series of chapters, some of which have already been published as research papers. A synthesis of the main findings from each chapter is reported in the following section. Thereafter, a broader discussion of the thesis’s

limitations, its contributions to research and policy implications, and areas for future research are presented.

7.2 Main findings

- Chapter II. For the purposes of this thesis, the first task is to systematically verify the absence of economic evaluation studies in the HNC follow-up. The search is then extended to all cancer types to learn about the methodology and the findings in evaluating any surveillance programs in oncology. This work results in an independent piece of work, but also provides key insights for the model-based economic evaluation performed later in the thesis. In synthesis, following the PRISMA statement [72], a systematic literature search was undertaken for studies published, in a first instance, since 2000 until the end of 2014, and subsequently updated until June 2017. The inclusion criteria imply selection of full economic evaluations of any type (i.e. cost-effectiveness, cost-utility, cost-benefit, cost-minimization, cost-consequences) assessing follow-up programs in adult cancer patients who have successfully completed the primary treatment. The original systematic review, published in *Expert Review of Pharmacoeconomics & Outcomes Research* in October 2015, includes 39 articles [310]; the updated review identifies 14 additional studies, for a total of 53 studies included in the thesis's chapter. Several types of follow-up interventions are compared in the included studies, but the attention of this thesis is focussed on those comparing programs of different 'intensity', defined as either increased frequency of standard examinations or add-on of new diagnostic procedures. Overall, most economic evaluations discourage the adoption of intensive surveillance, which turns out to be not cost-effective compared to minimal programs. Greater heterogeneity is observed in more recent studies

(published after 2014), with some showing a favourable cost-effectiveness profile for less intensive programs, and others recommending instead the more intensive ones. However, some of the results are reported as incremental cost per recurrence detected, which are not comparable with standard cost-effectiveness thresholds; moreover, the added value of an early diagnosis of cancer relapse is related to the availability of secondary treatments that are able to increase the quantity (and quality) of life. Thus, intermediate outcomes (e.g. increased detection of recurrences) are of limited value in the absence of an established relationship with final outcomes (i.e. survival). Other studies are cost-consequence analyses reporting significantly lower costs for less intensive options but no significant differences in clinical outcomes, mainly expressed as recurrence detection rates; even in these cases, a cost-effectiveness ratio is unavailable for comparison with country-specific thresholds. Quality scoring of the included studies is performed using the 24-item CHEERS checklist [73], with 31 (out of 53) papers being classified as ‘high quality’ studies, and rigour of economic evaluations improving over time. However, some critical issues emerge also from the most recent studies, including the wide use of intermediate outcomes, the short analysis timeframe preventing estimation of the final survival endpoint, and a limited use of cost-effectiveness and cost-utility analyses combining costs and outcomes in a ratio comparable with other disease areas. Moreover, whilst the original literature search confirms the absence of published economic evaluations in HNC follow-up until 2014, a study examining HNC is found by updating the search; however, this study compares a web-based follow-up to a standard telephone-based one in a cost-consequences framework and over a short period. Thus, there is still a research need for a

sound cost-utility analysis assessing the value of follow-up involving alternative programs of radiological investigations with a lifetime horizon.

- Chapter III. The availability of HSUVs is essential for model-based economic evaluations using QALYs. In this thesis, utility parameters are required for post-treatment health states composing the Markov state-transition model presented in Chapter V; thus, a systematic review of HSUVs in HNC is carried out and gives rise to an independent research paper, published in *Health and Quality of Life Outcomes* in September 2017 [311]. In brief, common electronic databases (PubMed, EMBASE, the Cochrane Library) are searched using a combination of relevant free-text terms. Other searches are conducted in the Tufts Cost-Effectiveness Analysis Registry and the School of Health and Related Research Health Utilities Database (SchARRHUD) specifically containing health utilities, in addition to the HERC database of mapping studies. Studies are considered for inclusion if reporting original HSUVs obtained through established techniques including direct methods (i.e. TTO and SG), MAUIs (e.g. EQ-5D), and ‘mapping’. The studies are qualitatively assessed using a list of criteria provided by recent guidelines on the topic [160]. Overall, a total of 28 studies qualify for data extraction and 346 unique HSUVs are retrieved from them. Three studies obtain utility values using mapping functions, but only one presents an original algorithm with HNC data. The remaining 25 studies are almost equally distributed between those using direct and indirect (i.e. MAUIs) techniques; EQ-5D is the most frequently adopted tool among MAUIs. A few critical elements are identified in reviewing these studies, including small sample sizes, limited reporting of missing values (and methods for dealing with missingness), and poor description of patients’ characteristics, especially in studies addressing HNC together with other cancer types. Moreover, most

studies elicit HSUVs during treatment or in the post-treatment disease-free state, whilst limited evidence is available for more advanced stages including the recurrent and terminal ones.

- Chapter IV. This chapter presents a mapping study providing a set of algorithms to obtain EQ-5D-5L utility values from a widely used HNC-specific HRQoL tool (i.e. EORTC QLQ-C30 with H&N35 module) which is being administered regularly to the patients in the ongoing HETeCo trial. The EQ-5D-5L responses obtained from a sample of patients currently enrolled in the RCT are valued using the currently available tariff sets (i.e. England, Netherlands, Canada, Uruguay, Korea, Japan, China) reported on the EuroQol website (last updated on 18th April 2017). Three different techniques are applied including a linear mixed-effects model, random-effects Tobit, and ALDVMM; all the developed models consider that multiple observations from each patient are likely to be correlated. Separate models for QLQ-C30 and H&N35 scales/items are developed and backward selection applied to identify the significant variables ($p < 0.05$). Overall, HNC patients in follow-up report a substantial HRQoL impairment, especially caused by insomnia, fatigue, dry mouth, sticky saliva, and financial problems. The average EQ-5D-5L utility value ranges between 0.786 and 0.905 according to the country set adopted; these values are aligned with a study [186] retrieved by the systematic review (Chapter III) and enrolling patients with similar characteristics. In models using the core set of questions (QLQ-C30), which are applicable to all cancer types, the scales/items significantly affecting the EQ-5D-5L utility are global health status, physical functioning, emotional functioning, nausea and vomiting, pain, constipation, diarrhoea, and financial difficulties; among the H&N35 scales/items, the most significant variables are pain (localised in the head and neck region), trouble

with social contacts, and felt ill. Some of these variables correspond to the EQ-5D dimensions, such as physical functioning with dimensions 1-3 (mobility, self-care, usual activities), pain with dimension 4 (pain/discomfort), and emotional functioning with dimension 5 (anxiety/depression). Overall, the linear random-effects and the multi-component ALDVM models show comparable goodness-of-fit in terms of AIC/BIC statistics, although the latter provide more precise estimates for the poorest health states; models using QLQ-C30 perform better than those using H&N35 for all the tariff sets adopted.

- Chapter V. The ‘core’ of the thesis is an exploratory cost-utility analysis of two alternative follow-up strategies in HNC, which coincide with the arms of the ongoing HETeCo trial. Since the clinical trial is expected to be completed by 2020, it generated a research question only, while the analyses are conducted in a modelling framework. Briefly, a Markov model with mutually exclusive health states is developed to predict the lifetime outcomes and costs of an intensive follow-up (corresponding to arm B in the trial) as conceived by the clinical investigators compared to a less intensive, symptom-driven surveillance (arm A) based on the NCCN guidelines. A variety of sources are used to inform the model structure and parameters, including the trial protocol, published and unpublished literature, and expert opinion; the systematic review reported in Chapter III provides the utility parameters, except for the ‘no evidence of disease’ state, whose value is derived from preliminary trial data. In the base-case analysis, the more intensive follow-up results in an incremental cost per QALY of €19,951, which is below the recommended €40,000/QALY threshold for Italy [262], and its cost-effectiveness is confirmed in sensitivity analyses. The key efficacy parameters in the model (i.e. the proportion of potentially salvageable recurrences in the two groups) are particularly uncertain, being

derived from published studies and clinical opinion pending completion of the trial. The value of undertaking additional research is estimated at around €300 per patient using the EVPI technique. Although this work is only exploratory, it suggests that an intensive follow-up program is likely to be a valuable option from a healthcare system perspective in Italy. The model's findings are poorly comparable with those identified by the systematic literature review presented in Chapter II, since most reviewed studies are cost-consequences analyses not reporting the ratio between incremental costs and incremental outcomes or expressing results in terms of cost per recurrence detected. A couple of studies [82] [99] performing a model-based cost-utility comparison of intensive versus less intensive programs obtained higher ICURs than in this thesis, but even these results are scarcely comparable due to different cancer sites and programs' specifications. A research paper based on this work has been published in *European Journal of Cancer* in April 2018 [312].

- Chapter VI. This work reports the findings of a BWS experiment carried out during summer 2015 at the NCI (Milan) and published in *Value in Health* in June 2017 [313]. This study represents the first experiment of this type in HNC follow-up, and the second one in cancer follow-up in general, after that published by Damery et al [293]. It investigates patients' preferences for several aspects of post-treatment surveillance using a stated preference technique. A balanced study design (i.e. four attributes, three levels each) is built and nine orthogonal scenarios (from a total of 81) are presented to a sample of patients currently in follow-up after being treated for primary HNC. The choice of attributes and levels for the experiment is informed by the features distinguishing arm A and arm B in the ongoing HETeCo trial. For each scenario reported in the BWS survey, participants are required to indicate the "best" and

the “worst” attribute-level combination (the ‘BW pair’). Responses are analysed through descriptive statistics and conditional logit regression; a covariate-adjusted model is also estimated to investigate sub-group preferences according to socio-demographic and clinical characteristics. The study findings reveal a general preference towards a follow-up program like the one already experienced at the NCI, a phenomenon that has been described in the literature as the ‘lure of the familiar’ [293]. In the overall sample, indeed, the item obtaining the highest utility coefficient is a program of physical examinations resembling that already performed at the NCI, whilst the one showing the lowest utility is not performing any PET scan during follow-up. In addition, patients typically dislike ways of delivering follow-up other than specialist-led visits, such as symptom-driven, telephone or primary care-based surveillance. These findings are broadly consistent with previous studies adopting stated preference techniques in other cancer types and reporting general preferences for follow-up to remain in secondary care and to include frequent face-to-face appointments and radiological examinations. The covariate-adjusted analyses reveal little evidence of preference heterogeneity, although some patient-specific variables (i.e. education level, employment and living status, time already spent in follow-up and number of treatments received) are significantly associated with the choice of BW items; conversely, age and gender have no significant effect on preferences. By summing the attribute-level coefficients resulting from the (unadjusted) conditional logit model, the scenario corresponding to arm B in the HETeCO trial obtains the highest utility in the overall sample, thus suggesting an alignment between elicited preferences and cost-effectiveness results (Chapter V) in HNC follow-up.

7.3 Limitations

The different studies comprising this thesis present some limitations that have been extensively described in the corresponding chapters. Here, a discussion of the overall weaknesses of the thesis is provided. The original thesis's research question was generated from a multicentre RCT that was to be completed by 2017. However, the study experienced severe delays, mainly due to a reluctance to participate on behalf of patients who are used to receive a quite intensive follow-up at Italian hospitals, broadly like that scheduled for arm B, and are afraid to be assigned to the symptom-driven one (arm A). Moreover, the clinical study faced some logistical and resource constraints that hampered the work at individual centres. Due to this slow recruitment process, more centres have joined the study over time, and the deadline for its completion has been delayed to 2020.

This issue mainly affected two chapters of this thesis. In Chapter IV, a limited number of observations were available to conduct the mapping exercise, thus affecting the precision of model estimates, and hindering the use of alternative regression techniques (e.g. response mapping). Moreover, the data collected at different time points were pooled together, thus preventing a longitudinal study design that considers also the visit number, and not only the patient, as a cluster variable. Therefore, it is likely that this paper will be updated once more patients are recruited to the RCT.

In Chapter V, the original idea was to perform an RCT-based economic evaluation over a 3-year period, corresponding to the trial length, with a subsequent extrapolation of lifetime outcomes and costs based on the trial's results using established survival analysis techniques. In the absence of adequate clinical information from the trial, the economic evaluation of the two follow-up strategies was performed instead using alternative data sources. The proportion of potentially salvageable recurrences that is

detected by the two alternative surveillance schemes, which represents a key parameter in the model, was retrieved from previous studies and the assumptions reported in the HETeCo trial protocol. Thus, an empirical confirmation of this data upon the trial completion is of interest.

Further limitations concern the two systematic literature reviews that are included in this thesis. In the first review (Chapter II), only the main databases (i.e. PubMed, EMBASE, and the Cochrane Library) were searched, despite there being others (e.g. EconLit) could also have been considered. In the second review (Chapter III), more databases were searched, including some specifically aimed at collecting HSUVs; however, a broad unpublished literature might be available on this topic that was disregarded. Additionally, in both reviews, the range of keywords was limited to those most frequently adopted in the field, thus potentially excluding studies using alternative terms (e.g. ‘monitoring’ instead of ‘follow-up’ or ‘surveillance’).

7.4 Contributions to research

Despite their limitations, each chapter (from II to VI) contributes to fill some knowledge “gaps” around different aspects defining the ‘value’ of follow-up programs in HNC. The two systematic reviews (Chapters II and III) provide a comprehensive and critical synthesis of relevant studies to inform the development of future research projects in HNC follow-up; moreover, they test the suitability of the available guidelines (i.e. the CHEERS checklist [73] and a set of recommendations from Papaioannou et al. [160]) to provide a quality scoring of the available literature. The mapping chapter (Chapter IV) is a technical work that furnishes a set of mapping functions to generate HSUVs in studies where only cancer-specific instruments (i.e. EORTC QLQ-C30 and H&N35) are collected. Moreover, the study represents a knowledge advancement in the recent mapping literature by exploring alternative

modelling techniques to the traditional linear regression, and a range of EQ-5D-5L value sets. The model presented in Chapter V adopts the well-established technique of cost per QALY to explore the value of using intensive radiological investigations in HNC follow-up. More intensive follow-up was shown to potentially increase patient's overall survival and quality-adjusted survival and be good 'value for money' for a regional healthcare system in Italy; moreover, conducting additional research in this field can be worthwhile according to the EVPI analysis. This chapter, despite intrinsic limitations related to the unavailability of reliable efficacy parameters, fills a relevant knowledge "gap" in HNC literature where no economic evaluations comparing surveillance programs of different intensity were available to date. Lastly, the BWS study (Chapter VI) explores the topic of patients' preferences that has received limited attention from health economics and HTA decision-making so far. The work also tests the suitability of a survey instrument, the BWS, which has been applied much less often than the better-known binary DCE. The analysis of self-reported difficulties by the patients confirms that BWS is a quick and easy tool for eliciting patients' preferences in vulnerable people like cancer patients.

7.5 Policy implications

Cost-effectiveness analyses are increasingly used to inform policy-makers about the efficient allocation of limited healthcare resources. This is particularly true in countries where health services are largely tax-financed, such as Italy, and healthcare expenditure is growing dramatically over time. In oncology, these methods have been mainly applied to drug treatments such as chemotherapy and radiotherapy, whilst much less evidence is available for non-pharmaceutical interventions including population screening, surgery, medical devices, and follow-up programs [57]. This thesis has the merit of stimulating a scientific debate around the topic of post-treatment surveillance in

cancer and can inform health and policy decision-making in several ways. The systematic review of economic evaluations in cancer follow-up (Chapter II) is useful to oncologists and policy-makers who wish to update surveillance programs in their organizations or countries based on scientific evidence. The quality assessment based on the CHEERS checklist allows identification of the most reliable evaluations. However, the search identified only one study [112] related to HNC and adopting a too short horizon to capture meaningful clinical outcomes. Thus, the first economic evaluation of follow-up interventions in this cancer population has been provided by this thesis (Chapter V). The cost-effectiveness model's results revealed that intensive follow-up with frequent radiological assessments (MRI, CT, PET) over time is likely to generate a quality-adjusted (and -unadjusted) survival gain at acceptable additional cost compared to less intensive, symptom-driven surveillance. In sensitivity analyses, a difference of 0.15 in 'curability' between arm B and arm A of the HETeCo trial is estimated to be sufficient for obtaining the cost-effectiveness of the more intensive intervention. Moreover, the healthcare costs of HNC follow-up have not been deeply studied yet [23]. In Chapter V, an estimate of 5-year costs associated to the two alternative surveillance strategies is provided, with details on resource consumption and cost items for the Lombardy region, which may represent a useful template to cost other follow-up programs in Italy and elsewhere. These results can influence the healthcare planning and financing in the next few years, as well as the clinical practice, at least in Italy.

Discrete choice preference methods, including BWS, allow analysis of innovative policies and complex interventions with multiple features [281]; moreover, they have a potential to contribute to an efficient allocation of scarce resources, although their role in the healthcare decision process has not yet been codified. DCE studies may be useful to account for trade-offs among different aspects of cancer care and to subsequently

prioritise and rank interventions from a patient's perspective. Overall, the awareness of patients' preferences and expectations can improve the adherence to the existing programs, incentivise their improvement or replacement, and strengthen the doctor-patient relationship. Although this thesis does not provide data on adherence to follow-up, the clinicians involved in this study report skipped appointments and poor compliance with physician's recommendations, especially during the later years of surveillance.

7.6 Future research

This thesis contributes to the literature on the health economic aspects of HNC follow-up. However, additional areas might be explored in the future, using this work as a starting point. The first systematic review (Chapter II) highlights the need for sound economic evaluations in any cancer follow-up, since the studies published so far present some methodological weaknesses, mainly related to the use of intermediate endpoints. Moreover, the studies mostly address common neoplasms such as breast, colorectal, and cervical, whilst other cancers, including HNC, received much less attention. The updated search conducted in June 2017 identified many new studies published in a limited time, thus carrying on this review work over the years would be valuable to collect systematically and to synthesize new evidence in the field.

Additionally, the area of health-related utility assessment in HNC deserves further attention, since the collection of HSUVs (Chapter III) retrieved only few data for recurrent and palliative stages, when patients are likely to be too sick and unable to self-complete questionnaires or take part in TTO/SG tasks. Thus, future research on HSUVs elicitation techniques that allows, for example, a systematic involvement of the caregiver or healthcare professionals, or the use of user-friendly mobile-app technologies to administer EQ-5D directly to the patient, might be further explored in

this cancer population. Moreover, additional work is needed in the mapping area to derive reliable functions to convert the most common HNC-specific questionnaires, including the FACT-H&N, into utility values that are usable in economic evaluation studies, and to establish the most efficient regression techniques for this task. As new EQ-5D-5L tariff sets are available, the development of model functions using algorithms other than those already included in the Chapter IV should be explored.

More research is needed also on the clinical effectiveness of alternative follow-up programs, which is essential to inform cost-utility analyses such as that reported in Chapter V. Until now, no prospective studies have been conducted on the topic, and the evidence generated from experimental trials is completely lacking [20]. Thus, as confirmed by the EVPI results, RCTs beyond the one mentioned in this thesis are strongly encouraged to understand the impact of varying follow-up intensity on patient's survival and QALYs. Moreover, even in publicly funded healthcare systems such as Italy, the private expenditure and indirect costs over the course of the disease might be substantial in cancer; this issue is particularly relevant in HNC, where most patients have traditionally belonged to middle-low socioeconomic groups [19]. Thus, future economic evaluations of HNC follow-up programs might consider a broader societal perspective. Cost-effectiveness analysis of follow-up strategies stratified by risk of recurrences and cancer site is a further research opportunity, since different post-treatment patterns have been identified according to age, smoking status and positivity or negativity to HPV. Lastly, how the use of mapped utility values, instead of original ones, impacts on QALY calculation and, consequently, on cost-utility analyses results might be addressed in future research.

In the BWS study (Chapter VI), the preference analysis is limited to the process-related aspects of the follow-up, as clinical outcomes are still uncertain and under investigation in the RCT. Additional elements might be considered in future studies, such as the

survival gains, the toxicities related to intensive radiological surveillance, the level of scientific evidence, and co-payments for additional investigations. Moreover, how HTA should systematically consider patients' preferences alongside the traditional QALY approach remains largely understudied [284]. The next, challenging step is combining BWS estimates with program costs to inform healthcare decisions about 'customised care', which has been defined as *healthcare tailored on a patient-by-patient basis* and has the potential to achieve cost-effectiveness when many individuals prefer a less expensive program over a more intensive one [289].

7.7 Conclusions

The number of people living with cancer has increased substantially over time and is expected to grow further given advancements in medical and surgical treatment, diagnostic tools, and an aging population [314]. After being treated for their primary cancer, patients usually enter a program of post-treatment follow-up which may last for several years. Routine surveillance is aimed at detecting recurrences, metastases or second primaries at the earliest opportunity to administer potentially salvage treatments; however, these schemes often lack a sound scientific base and may impose a significant economic burden to healthcare systems and societies [32]. In HNC, recent advances in primary treatments and rising incidence of HPV-related cancers are likely to increase the number of patients who complete the standard 5-year follow-up.

Until now, post-treatment follow-up in oncology has received little attention, since most of the research efforts have been oriented towards the development of effective anti-cancer therapies. This "gap" is particularly evident for some malignancies like HNC. This thesis contributes to the current debate around post-treatment surveillance in HNC, by taking advantage of the collaboration with an HNC oncology department located at the NCI in Milan, Italy. Moreover, in conducting the BWS experiment, the candidate

had the opportunity to hear the patients' 'voice', who could value the main features making up a follow-up program in HNC. She has also been invited to join a focus group organized by the Italian Association of Medical Oncology (AIOM) and involving several healthcare professionals with the objective to standardise the follow-up practices in HNC on the national territory based on sound clinical and economic evidence. Other stakeholders including policy-makers, patients' representatives, caregivers, and pharmaceutical companies might contribute to this emerging discussion in the future.

The first RCT is ongoing in Italy and it is hoped that many other studies, with a range of designs, will be carried out in Italy and elsewhere on this topic. However, a relevant step forward in the knowledge of HNC follow-up has been taken with this thesis.

REFERENCES

- [1] B. J. Braakhuis, C. R. Leemans e O. Visser, «Incidence and survival trends of head and neck squamous cell carcinoma in the Netherlands between 1989 and 2011,» *Oral Oncol*, vol. 50, pp. 670-75, 2014.
- [2] G. Gatta, L. Botta, M. J. Sánchez, L. A. Anderson, D. Pierannunzio, L. Licitra e the EUROCCARE Working Group, «Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCCARE-5 population-based study,» *Eur J Cancer*, vol. 51, pp. 2130-43, 2015.
- [3] Associazione Italiana di Oncologia Medica (AIOM), «Linee guida per i tumori della testa e del collo,» 2013. [Online]. Available: www.aiom.it.
- [4] A. Karligkiotis, N. Machouchas, C. Bozzo, A. Melis, A. Cossu, M. Budroni e et al., «Head and neck cancer epidemiology in North Sardinia, Italy,» *Acta Medica Mediterr*, vol. 30, n. 1, pp. 41-7, 2014.
- [5] AIRTUM Working Group, «I tumori in Italia. Rapporto 2016. La sopravvivenza dei pazienti oncologici in Italia,» 2017. [Online]. Available: www.epiprev.it/materiali/2017/EP2/AIRTUM_2017_full.pdf.
- [6] Associazione Italiana Registro Tumori (AIRTUM); Associazione Italiana Oncologia Medica (AIOM), «I numeri del cancro in Italia 2015,» 2015. [Online]. Available: http://www.registri-tumori.it/PDF/AIOM2015/I_numeri_del_cancro_2015.pdf.
- [7] ISTAT, «Il consumo di alcol in Italia. Anno 2016,» 2017. [Online]. Available: https://www.istat.it/it/files/2017/04/Consumo_alcol_in_Italia_2016.pdf?title=Consumo+di+alcol+++12%2Fapr%2F2017+++Testo+integrale++nota+metodologica.pdf.
- [8] EpiCentro, «Rapporto nazionale Passi 2011: abitudine al fumo,» 2011. [Online]. Available: <http://www.epicentro.iss.it/passi/rapporto2011/FumoAttivo.asp>.
- [9] S. Marur e A. A. Forastiere, «Head and neck squamous cell carcinoma: update on epidemiology, diagnosis and treatment,» *Mayo Clin Proc*, vol. 91, n. 3, pp. 386-96, 2016.
- [10] D. A. de Andrade e J. P. Machiels, «Treatment options for patients with recurrent or metastatic squamous cell carcinoma of the head and neck, who progress after platinum-based chemotherapy,» *Curr Opin Oncol*, vol. 24, pp. 211-7, 2012.
- [11] O. Guntinas-Lichius, T. G. Wendt, N. Kornetzky, J. Buentzel, D. Esser, D. Böger e et al., «Trends in epidemiology and treatment and outcome for head and neck cancer: A population-based long-term analysis from 1996 to 2011 of the Thuringian cancer registry,» *Oral Oncol*, vol. 50, pp. 1157-64, 2014.

- [12] K. A. Price e E. E. Cohen, «Current treatment options for metastatic head and neck cancer,» *Curr Treat Options Oncol*, vol. 13, pp. 35-46, 2012.
- [13] E. Wissinger, I. Griebisch, J. Lungershausen, T. Foster e C. L. Pashos, «The economic burden of head and neck cancer: A systematic literature review,» *Pharmacoeconomics*, vol. 32, n. 9, pp. 865-82, 2014.
- [14] M. A. Tejani, R. B. Cohen e R. Mehra, «The contribution of cetuximab in the treatment of recurrent and/or metastatic head and neck cancer,» *Biologics*, vol. 4, pp. 173-85, 2010.
- [15] L. Sweeny, E. L. Rosenthal, L. Clemons, T. M. Stevens, E. R. McIntosh e W. R. Carroll, «Outcomes after surgical salvage for recurrent oropharyngeal squamous cell carcinoma,» *Oral Oncol*, vol. 60, pp. 118-24, 2016.
- [16] A. Kawecki e R. Krajewski, «Follow-up in patients treated for head and neck cancer,» *Memo*, vol. 7, pp. 87-91, 2014.
- [17] J. B. Vermorken e P. Specenier, «Optimal treatment for recurrent/metastatic head and neck cancer,» *Ann Oncol*, vol. 21, n. 7, pp. vii252-61, 2010.
- [18] J. A. de Souza e T. Y. Seiwert, «A value framework in head and neck cancer,» *Am Soc Clin Oncol Educ Book*, pp. e304-9, 2014.
- [19] H. J. Lu e P. M. Chang, «How to treat recurrent/metastatic head and neck cancer: the economic issue in real-world practice,» *Curr Med Res Opin*, vol. 33, n. 4, pp. 779-80, 2017.
- [20] F. De Felice, D. Musio e V. Trombolini, «Follow-up in head and neck cancer: a management dilemma,» *Adv Otolaryngol*, 2015.
- [21] K. S. Choe, D. J. Haraf, A. Solanki, E. E. Cohen, T. Y. Seiwert, K. M. Stenson e et al., «Prior chemoradiotherapy adversely impacts outcomes of recurrent and second primary head and neck cancer treated with concurrent chemotherapy and reirradiation,» *Cancer*, vol. 117, n. 20, pp. 4671-8, 2011.
- [22] D. Posch, H. Fuchs, G. Kornek, A. Grah, J. Pammer, M. B. Aretin e et al., «Docetaxel plus cetuximab biweekly is an active regimen for the first line treatment of patients with recurrent/metastatic head and neck cancer,» *Sci Rep*, vol. 6, pp. 1-7, 2016.
- [23] N. Denaro, M. C. Merlano e E. G. Russi, «Follow-up in head and neck cancer: do more does it mean do better? A systematic review and our proposal based on our experience,» *Clin Exp Otorhinolaryngol*, vol. 9, n. 4, pp. 287-97, 2016.
- [24] J. H. González-Botas, P. V. Vázquez e C. V. Barro, «Second primary tumours in head and neck cancer,» *Acta Otorrinolaringol Esp*, vol. 67, n. 3, pp. 123-9, 2016.
- [25] A. S. Ho, D. H. Kraus, I. Ganly, N. Y. Lee, J. P. Shah e L. G. Morris, «Decision making in the management of recurrent head and neck cancer,» *Head Neck*, vol. 36, n. 1, pp. 144-51, 2014.

- [26] P. Strojan, J. Corry, A. Eisbruch, J. B. Vermorken, W. M. Mendenhall e A. W. Lee, «Recurrent and second primary squamous cell carcinoma of the head and neck: when and how to reirradiate,» *Head Neck*, vol. 37, n. 1, pp. 134-50, 2015.
- [27] S. J. Wong, D. E. Heron, K. Stenson, D. C. Ling e J. A. Vargo, «Locoregional recurrent or second primary head and neck cancer: management strategies and challenges,» *Am Soc Clin Oncol Educ Book*, vol. 35, pp. e284-92, 2016.
- [28] A. Pagh, T. Vedtofte, C. D. Lynggaard, N. Rubek, M. Lonka, J. Johansen e et al., «The value of routine follow-up after treatment for head and neck cancer: A National Survey from DAHANCA,» *Acta Oncol*, vol. 52, n. 2, pp. 277-84, 2013.
- [29] M. Piemonte e M. G. Ruggiu, «Il follow-up dei tumori maligni della testa e del collo. Quaderni monografici di aggiornamento,» 2013. [Online]. Available: https://www.aooi.it/contents/attachment/c11/QMA_31_2013.pdf.
- [30] F. De Felice, M. de Vincentiis, V. Valentini, D. Musio, S. Mezi, L. Lo Mele e et al., «Follow-up program in head and neck cancer,» *Crit Rev Oncol Hematol*, vol. 113, pp. 151-5, 2017.
- [31] S. J. Wang, «Surveillance radiologic imaging after treatment of oropharyngeal cancer: a review,» *World J Surg Oncol*, vol. 13, pp. 1-6, 2015.
- [32] K. Manikantan, S. Khode, R. C. Dwivedi, R. Palav, C. M. Nutting, P. Rhys-Evans e et al., «Making sense of post-treatment surveillance in head and neck cancer: when and what to follow-up,» *Cancer Treat Rev*, vol. 35, pp. 744-53, 2009.
- [33] Fondazione IRCCS Istituto Nazionale dei Tumori, Milan (Italy), «Health and economic outcomes of two different follow-up strategies in effectively cured advanced head and neck cancer (HETeCo),» [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT02262221>. [Consultato il giorno 2016].
- [34] D. G. Pfister, S. Spencer, D. M. Brizel, B. Burtness, P. M. Busse, J. J. Caudell e et al., «Head and neck cancers, Version 2.2014. Clinical practice guidelines in oncology,» *J Natl Compr Canc Netw*, vol. 12, n. 10, pp. 1454-87, 2014.
- [35] M. E. Porter, «What is value in healthcare?,» *N Engl J Med*, vol. 363, n. 26, pp. 2477-81, 2010.
- [36] M. C. Weinstein, G. Torrance e A. McGuire, «QALYs: the basics,» *Value Health*, vol. 12, n. 1, pp. S5-9, 2009.
- [37] B. H. Goulart, «Value: the next frontier in cancer care,» *Oncologist*, vol. 21, n. 6, pp. 651-3, 2016.
- [38] N. J. Devlin e P. K. Lorgelly, «QALYs as a measure of value in cancer,» *J Cancer Policy*, vol. 11, pp. 19-25, 2017.
- [39] M. Garau, K. K. Shah, A. R. Mason, Q. Wang, A. Towse e M. F. Drummond, «Using QALYs in cancer,» *Pharmacoeconomics*, vol. 29, n. 8, pp. 673-85, 2011.

- [40] R. Llewellyn-Bennett, L. Bowman e R. Bulbulia, «Post-trial follow-up methodology in large randomized controlled trials: a systematic review protocol,» *Syst Rev*, vol. 5, n. 1, pp. 1-7, 2016.
- [41] S. Michiels, E. D. Saad e M. Buyse, «Progression-Free Survival as a Surrogate for Overall Survival in Clinical Trials of Targeted Therapy in Advanced Solid Tumors,» *Drugs*, vol. 77, n. 7, pp. 713-9, 2017.
- [42] P. M. Rothwell, «Factors that can affect the external validity of randomised controlled trials,» *PLoS Clin Trials*, vol. 1, n. 1, pp. 1-5, 2006.
- [43] U. A. Matulonis, A. M. Oza, T. W. Ho e J. A. Ledermann, «Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials,» *Cancer*, vol. 121, n. 11, pp. 1737-46, 2015.
- [44] A. E. Attema e W. B. Brouwer, «On the (not so) constant proportional trade-off in TTO,» *Qual Life Res*, vol. 19, n. 4, pp. 489-97, 2010.
- [45] P. Dolan, G. Kavetsos e A. Tsuchiya, «Sick but satisfied: the impact of life and health satisfaction on choice between health scenarios,» *J Health Econ*, vol. 32, n. 4, pp. 708-14, 2013.
- [46] S. Dixon, L. Longworth and A. Wailoo, «Assessing technologies at the end of life: a review of empirical evidence,» 2009. [Online]. Available: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/DSU-EOL-lit-review.FINAL_.pdf.
- [47] R. Grieve, M. Grishchenko e J. Cairns, «SF-6D versus EQ-5D: reasons for differences in utility scores and impact on reported cost-utility,» *Eur J Health Econ*, vol. 10, pp. 15-23, 2009.
- [48] D. N. Lakdawalla, J. A. Romley, Y. Sanchez, J. R. Maclean, J. R. Penrod e T. Philipson, «How cancer patients value hope and the implications for cost-effectiveness assessments of high-cost cancer therapies,» *Health Aff (Millwood)*, vol. 31, n. 4, pp. 676-82, 2012.
- [49] M. Collins e N. Latimer, «NICE's end of life criteria: who gains, who loses?,» *BMJ*, vol. 346, 2013.
- [50] J. E. Brazier, S. Dixon e J. Ratcliffe, «The role of patient preferences in cost-effectiveness analysis: a conflict of values?,» *Pharmacoeconomics*, vol. 27, n. 9, pp. 705-12, 2009.
- [51] L. E. Schnipper, N. E. Davidson, D. S. Wollins, D. W. Blayney, A. P. Dicker, P. A. Ganz e et al., «Updating the American Society of Clinical Oncology Value Framework: Revisions and Reflections in Response to Comments Received,» *J Clin Oncol*, vol. 34, n. 24, pp. 2925-34, 2016.

- [52] N. I. Cherny, R. Sullivan, U. Dafni, J. M. Kerst, A. Sobrero, C. Zielinski e et al., «A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS),» *Ann Oncol*, vol. 26, n. 8, pp. 1547-73, 2015.
- [53] R. W. Carlson e E. Jonasch, «NCCN Evidence Blocks,» *J Natl Compr Canc Netw*, vol. 14, n. 5, pp. 616-9, 2016.
- [54] R. Nelson, «New abacus calculates price of cancer drug based on value,» Medscape, 2015. [Online]. Available: <http://www.medscape.com/viewarticle/847755>.
- [55] L. Pizzi, «The Institute for Clinical and Economic Review and its growing influence on the US healthcare,» *Am Health Drug Benefits*, vol. 9, n. 1, pp. 9-10, 2016.
- [56] J. Maervoet, P. Moise e S. Naidoo, «Overview and comparison of frameworks for the valuation of oncology drugs,» in *ISPOR 21st Annual International Meeting*, Washington, DC, USA, 2016.
- [57] M. Barbieri, H. L. Weatherly, R. Ara, H. Basarir, M. Sculpher, R. Adams e et al., «What is the quality of economic evaluations of non-drug therapies? A systematic review and critical appraisal of economic evaluations of radiotherapy for cancer,» *Appl Health Econ Health Policy*, vol. 12, pp. 497-510, 2014.
- [58] I. Boutron, L. Guittet, C. Estellat, D. Moher, A. Hróbjartsson e P. Ravaud, «Reporting methods of blinding in randomized trials assessing nonpharmacological treatments,» *PLoS Med*, vol. 4, n. 2, pp. 370-80, 2007.
- [59] A. Lanfranconi, L. Pylkkänen, S. Deandrea, A. Bramesfeld, D. Lerda, L. Neamtii e et al., «Intensive follow-up for women with breast cancer: review of clinical, economic and patient's preference domains through evidence to decision framework,» *Health Qual Life Outcomes*, vol. 15, n. 1, pp. 1-18, 2017.
- [60] C. Jackson, J. Stevens, S. Ren, N. Latimer, L. Bojke, A. Manca e et al., «Extrapolating survival from randomized trials using external data: a review of methods,» *Med Decis Making*, vol. 37, n. 4, pp. 377-90, 2017.
- [61] Y. Schenker, R. M. Arnold, J. E. Bauman, D. E. Heron e J. T. Johnson, «An enhanced role for palliative care in the multidisciplinary approach to high-risk head and neck cancer,» *Cancer*, vol. 122, n. 3, pp. 340-3, 2016.
- [62] S. H. Kim, M. W. Jo, H. J. Kim e J. H. Ahn, «Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients,» *Health Qual Life Outcomes*, vol. 10, pp. 1-6, 2012.
- [63] E. R. Marriott, G. van Hazel, P. Gibbs e A. J. Hatswell, «Mapping EORTC-QLQ-C30 to EQ-5D-3L in patients with colorectal cancer,» *J Med Econ*, vol. 20, n. 2, pp. 193-9, 2017.

- [64] D. Rowen, J. Brazier, T. Young, S. Gaugris, B. M. Craig, M. T. King e et al., «Deriving a preference-based measure for cancer using the EORTC QLQ-C30,» *Value Health*, vol. 14, n. 5, pp. 721-31, 2011.
- [65] R. A. Lewis, R. D. Neal, N. H. Williams, B. France, M. Hendry, D. Russell e et al., «Follow-up of cancer in primary care versus secondary care: systematic review,» *Br J Gen Pract*, vol. 59, n. 564, pp. e234-47, 2009.
- [66] R. Dickinson, S. Hall, J. E. Sinclair, C. Bond e P. Murchie, «Using technology to deliver cancer follow-up: a systematic review,» *BMC Cancer*, vol. 14, pp. 1-16, 2014.
- [67] M. Jeffery, B. E. Hickey e P. N. Hider, «Follow-up strategies for patients treated for non-metastatic colorectal cancer (Review),» *Cochrane Libr*, vol. 1, pp. 1-37, 2007.
- [68] S. Pita-Fernández , M. Alhayek-Ai , C. González-Martín , B. López-Calviño, T. Seoane-Pillado e S. Pértega-Díaz, «Intensive follow-up strategies improve outcomes in nonmetastatic colorectal cancer patients after curative surgery: a systematic review and meta-analysis,» *Ann Oncol*, vol. 26, n. 4, pp. 644-56, 2015.
- [69] A. Dignonnet, M. Hamoir, G. Andry, M. J. Haigentz, R. P. Takes, C. E. Silver e et al., «Post-therapeutic surveillance strategies in head and neck squamous cell carcinoma,» *Eur Arch Otorhinolaryngol*, vol. 270, n. 5, pp. 1569-80, 2013.
- [70] V. Nama, A. Nordin e A. Bryant, «Patient-reported outcome measures for follow-up after gynecological cancer treatment (Review),» *Cochrane Libr*, vol. 11, pp. 1-15, 2013.
- [71] C. Robertson, S. K. Arcot Ragupathy, C. Boachie, J. M. Dixon, C. Fraser, R. Hernández e et al., «The clinical effectiveness and cost-effectiveness of different surveillance mammography regimens after the treatment for primary breast cancer: systematic reviews registry database analyses and economic evaluation,» *Health Technol Assess*, vol. 15, n. 34, pp. 1-322, 2011.
- [72] D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew e et al., «Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement,» *Syst Rev*, vol. 4, n. 1, pp. 1-9, 2015.
- [73] D. Husereau, M. Drummond, S. Petrou, C. Carswell, D. Moher, D. Greenberg e et al., «Consolidated Health Economic Evaluation Reporting Standards (CHEERS)-- explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force,» *Value Health*, vol. 16, n. 2, pp. 231-50, 2013.
- [74] L. G. Gordon e D. Rowell, «Health system costs of skin cancer and cost-effectiveness of skin cancer prevention and screening: a systematic review,» *Eur J Cancer Prev*, vol. 24, pp. 141-9, 2015.
- [75] S. Maru, J. Byrnes, M. J. Carrington, S. Stewart e P. A. Scuffham, «Systematic review of trial-based analyses reporting the economic impact of heart failure management programs compared with usual care,» *Eur J Cardiovasc Nurs*, vol. 15, n. 1, pp. 82-90, 2016.

- [76] C. Mihalopoulos e M. L. Chatterton, «Economic evaluations of interventions designed to prevent mental disorders: a systematic review,» *Early Interv Psychiatry*, vol. 9, n. 2, pp. 85-92, 2015.
- [77] S. J. Palfreyman e P. W. Stone, «A systematic review of economic evaluations assessing interventions aimed at preventing or treating pressure ulcers,» *Int J Nurs Stud*, vol. 52, n. 3, pp. 769-88, 2015.
- [78] B. Almog, R. Gamzu, J. Bornstein, I. Levin, O. Fainaru, J. Niv e et al., «Clinical and economic benefit of HPV-load testing in follow-up and management of women postcone biopsy for CIN2-3,» *Br J Cancer*, vol. 89, n. 1, pp. 109-12, 2003.
- [79] K. A. Armstrong, J. L. Semple e P. C. Coyte, «Replacing ambulatory surgical follow-up visits with mobile app home monitoring: modeling cost-effective scenarios,» *J Med Internet Res*, vol. 16, n. 9, pp. 1-34, 2014.
- [80] S. A. Assoumou, K. H. Mayer, L. Panther, B. P. Linas e J. Kim, «Cost-effectiveness of surveillance strategies after treatment for high-grade anal dysplasia in high-risk patients,» *Sex Transm Dis*, vol. 40, n. 4, pp. 298-303, 2013.
- [81] K. M. Augestad, J. Norum, S. Dehof, R. Aspevik, U. Ringberg, T. Nestvold e et al., «Cost-effectiveness and quality of life in surgeon versus general practitioner-organised colon cancer surveillance: a randomised controlled trial,» *BMJ Open*, vol. 3, n. 4, pp. 1-15, 2013.
- [82] P. Auguste, P. Barton, C. Meads, C. Davenport, S. Malysiak, M. Kowalska e et al., «Evaluating PET-CT in routine surveillance and follow-up after treatment for cervical cancer: a cost-effectiveness analysis,» *BJOG*, vol. 121, n. 4, pp. 464-76, 2014.
- [83] J. M. Baena-Cañada, P. Ramírez-Daffós, C. Cortés-Carmona, P. Rosado-Varela, J. Nieto-Vera e E. Benítez-Rodríguez, «Follow-up of long-term survivors of breast cancer in primary care versus specialist attention,» *Fam Pract*, vol. 30, n. 5, pp. 525-32, 2013.
- [84] K. Beaver, W. Hollingworth, R. McDonald, G. Dunn, D. Tysver-Robinson, L. Thomson e et al., «Economic evaluation of a randomized clinical trial of hospital versus telephone follow-up after treatment for breast cancer,» *Br J Surg*, vol. 96, n. 12, pp. 1406-15, 2009.
- [85] T. Bessen e J. Karnon, «A patient-level calibration framework for evaluating surveillance strategies: a case study of mammographic follow-up after early breast cancer,» *Value Health*, vol. 17, pp. 669-78, 2014.
- [86] T. Bessen, D. M. Keefe e J. Karnon, «Does one size fit all? Cost utility analyses of alternative mammographic follow-up schedules, by risk of recurrence,» *Int J Technol Assess Health Care*, vol. 31, n. 5, pp. 281-8, 2015.
- [87] F. Borie, C. Combescure, J. P. Daurès, B. Trétarre e B. Millat, «Cost-effectiveness of two follow-up strategies for curative resection of colorectal cancer: comparative study using a Markov model,» *World J Surg*, vol. 28, n. 6, pp. 563-9, 2004.

- [88] V. M. Coupé, J. Berkhof, R. H. Verheijen e C. J. Meijer, «Cost-effectiveness of human papillomavirus testing after treatment for cervical intraepithelial neoplasia,» *BJOG*, vol. 114, n. 4, pp. 416-24, 2007.
- [89] D. Coyle, E. Grunfeld, K. Coyle, G. Pond, J. A. Julian e M. N. Levine, «Cost effectiveness of a survivorship care plan for breast cancer survivors,» *J Oncol Pract*, vol. 10, n. 2, pp. e86-e92, 2014.
- [90] I. Cromwell, M. Gaudet, S. J. Peacock e C. Aquino-Parsons, «Cost-effectiveness analysis of anal cancer screening in women with cervical neoplasia in British Columbia, Canada,» *BMC Health Serv Res*, vol. 16, pp. 1-10, 2016.
- [91] S. Damude, J. E. Hoekstra-Weebers, A. B. Francken, S. Ter Meulen, E. Bastiaannet e H. J. Hoekstra, «The MELFO-Study: prospective, randomized, clinical trial for the evaluation of a stage-adjusted reduced follow-up schedule in cutaneous melanoma patients-results after 1 year,» *Ann Surg Oncol*, vol. 23, n. 9, pp. 2762-71, 2016.
- [92] V. Dansk, P. U. Malmström, M. Bläckberg e M. Malmenäs, «Hexaminolevulinate hydrochloride blue-light flexible cystoscopy in the detection and follow-up of nonmuscle-invasive bladder cancer: cost consequences during outpatient surveillance in Sweden,» *Future Oncol*, vol. 12, n. 8, pp. 1025-38, 2016.
- [93] P. Das, A. K. Ng, C. C. Earle, P. M. Mauch e K. M. Kuntz, «Computed tomography screening for lung cancer in Hodgkin's lymphoma survivors: decision analysis and cost-effectiveness analysis,» *Ann Oncol*, vol. 17, n. 5, pp. 785-93, 2006.
- [94] E. W. de Bekker-Grob, M. N. van der Aa, E. C. Zwarthoff, M. J. Eijkemans, B. W. van Rhijn, T. H. van der Kwast e et al., «Non-muscle-invasive bladder cancer surveillance for which cystoscopy is partly replaced by microsatellite analysis of urine: a cost-effective alternative?,» *BJU Int*, vol. 104, n. 1, pp. 41-7, 2009.
- [95] J. D. Emery, M. Jefford, M. King, D. Hayne, A. Martin, J. Doorey e et al., «ProCare Trial: a phase II randomized controlled trial of shared care for follow-up of men with prostate cancer,» *BJU Int*, vol. 119, n. 3, pp. 381-9, 2017.
- [96] F. S. Erenay, O. Alagoz, R. Banerjee, A. Said e R. R. Cima, «Cost-effectiveness of alternative colonoscopy surveillance strategies to mitigate metachronous colorectal cancer incidence,» *Cancer*, vol. 122, n. 16, pp. 2560-70, 2016.
- [97] F. Forni, G. Ferrandina, F. Deodato, G. Macchia, A. G. Morganti, D. Smaniotto e et al., «Squamous cell carcinoma antigen in follow-up of cervical cancer treated with radiotherapy: evaluation of cost-effectiveness,» *Int J Radiat Oncol Biol Phys*, vol. 69, n. 4, pp. 1145-9, 2007.
- [98] S. Gilbert, K. R. Reid, M. Y. Lam e D. Petsikas, «Who should follow up lung cancer patients after operation?,» *Ann Thorac Surg*, vol. 69, n. 6, pp. 1696-700, 2000.

- [99] B. A. Guadagnolo, R. S. Punglia, K. M. Kuntz, P. M. Mauch e A. K. Ng, «Cost-effectiveness analysis of computerized tomography in the routine follow-up of patients after primary treatment for Hodgkin's disease,» *J Clin Oncol*, vol. 24, n. 25, pp. 4116-22, 2006.
- [100] C. Hassan, P. J. Pickhardt, A. Zullo, E. Di Giulio, A. Laghi, D. H. Kim e et al., «Cost-effectiveness of early colonoscopy surveillance after cancer resection,» *Dig Liver Dis*, vol. 41, n. 12, pp. 881-5, 2009.
- [101] N. Hatam, N. Ahmadloo, M. Vazirzadeh, A. Jafari e M. Askarian, «Cost-effectiveness of intensive vs. standard follow-up models for patients with breast cancer in Shiraz, Iran,» *Asian Pac J Cancer Prev*, vol. 17, n. 12, pp. 5309-14, 2016.
- [102] F. Helgesen, S. O. Andersson, O. Gustafsson, E. Varenhorst, B. Gobén, S. Carnock e et al., «Follow-up of prostate cancer patients by on-demand contacts with a specialist nurse: a randomized study,» *Scand J Urol Nephrol*, vol. 34, n. 1, pp. 55-61, 2000.
- [103] U. R. Hengge, A. Wallerand, A. Stutzki e N. Kockel, «Cost-effectiveness of reduced follow-up in malignant melanoma,» *J Dtsch Dermatol Ges*, vol. 5, n. 10, pp. 898-907, 2007.
- [104] A. M. Kamat, J. A. Karam, H. B. Grossman, A. K. Kader, M. Munsell e C. P. Dinney, «Prospective trial to identify optimal bladder cancer surveillance protocol: reducing costs while maximizing sensitivity,» *BJU Int*, vol. 108, n. 7, pp. 1119-23, 2011.
- [105] M. S. Kent, P. Korn, J. L. Port, P. C. Lee, N. K. Altorki e R. J. Korst, «Cost effectiveness of chest computed tomography after lung cancer resection: a decision analysis model,» *Ann Thorac Surg*, vol. 80, n. 4, pp. 1215-22, 2005.
- [106] M. L. Kimman, C. D. Dirksen, A. C. Voogd, P. Falger, B. C. Gijsen, M. Thuring e et al., «Economic evaluation of four follow-up strategies after curative treatment for breast cancer: results of an RCT,» *Eur J Cancer*, vol. 47, n. 8, pp. 1175-85, 2011.
- [107] I. Koinberg, G. B. Engholm, A. Genell e L. Holmberg, «A health economic evaluation of follow-up after breast cancer surgery: results of an rct study,» *Acta Oncol*, vol. 48, n. 1, pp. 99-104, 2009.
- [108] R. Kokko, M. Hakama e K. Holli, «Follow-up cost of breast cancer patients with localized disease after primary treatment: a randomized trial,» *Breast Cancer Res Treat*, vol. 93, n. 3, pp. 255-60, 2005.
- [109] A. Lanceley, C. Berzuini, M. Burnell, S. Gessler, S. Morris, A. Ryan e et al., «Ovarian cancer follow-up: a preliminary comparison of 2 approaches,» *Int J Gynecol Cancer*, vol. 27, n. 1, pp. 59-68, 2017.
- [110] R. Legood, M. Smith, J. B. Lew, R. Walker, S. Moss, H. Kitchener e et al., «Cost effectiveness of human papillomavirus test of cure after treatment for cervical intraepithelial neoplasia in England: economic analysis from NHS Sentinel Sites Study,» *BMJ*, vol. 345, pp. 1-10, 2012.

- [111] W. Lu, M. J. Greuter, M. Schaapveld, K. M. Vermeulen, T. Wiggers e G. H. de Bock, «Safety and cost-effectiveness of shortening hospital follow-up after breast cancer treatment,» *Br J Surg*, vol. 99, n. 9, pp. 1227-33, 2012.
- [112] K. X. Lyu, J. Zhao, B. Wang, G. X. Xiong, W. Q. Yang, Q. H. Liu e et al., «Smartphone application WeChat for clinical follow-up of discharged patients with head and neck tumors: a randomized controlled trial,» *Chin Med J (Engl)*, vol. 129, n. 23, pp. 2816-23, 2016.
- [113] D. A. Macafee, D. K. Whynes e J. H. Scholefield, «Risk-stratified intensive follow up for treated colorectal cancer - realistic and cost saving?,» *Colorectal Dis*, vol. 10, n. 3, pp. 222-30, 2008.
- [114] J. Melnikow, S. Kulasingam, C. Slee, L. J. Helms, M. Kuppermann, S. Birch e et al., «Surveillance after treatment for cervical intraepithelial neoplasia: outcomes, costs, and cost-effectiveness,» *Obstet Gynecol*, vol. 116, n. 5, pp. 1158-70, 2010.
- [115] J. Monteil, A. Vergnenègre, F. Bertin, F. Dalmay, S. Gaillard, F. Bonnaud e et al., «Randomized follow-up study of resected NSCLC patients: conventional versus 18F-DG coincidence imaging,» *Anticancer Res*, vol. 30, n. 9, pp. 3811-6, 2010.
- [116] S. Moore, J. Corner, J. Haviland, M. Wells, E. Salmon, C. Normand e et al., «Nurse led follow up and conventional medical follow up in management of patients with lung cancer: randomised trial,» *BMJ*, vol. 325, n. 7373, pp. 1-7, 2002.
- [117] B. G. Ok, Y. S. Ji, Y. H. Ko e P. H. Song, «Usefulness of urine cytology as a routine work-up in the detection of recurrence in patients with prior non-muscle-invasive bladder cancer: practicality and cost-effectiveness,» *Korean J Urol*, vol. 55, n. 10, pp. 650-5, 2014.
- [118] A. Oltra, A. Santaballa, B. Munárriz, M. Pastor e J. Montalar, «Cost-benefit analysis of a follow-up program in patients with breast cancer: a randomized prospective study,» *Breast J*, vol. 13, n. 6, pp. 571-4, 2007.
- [119] N. T. Phippen, L. J. Havrilesky, J. C. Barnett, C. A. Hamilton, M. P. Stany e W. J. Lowery, «Does Routine Posttreatment PET/CT Add Value to the Care of Women With Locally Advanced Cervical Cancer?,» *Int J Gynecol Cancer*, vol. 26, n. 5, pp. 944-50, 2016.
- [120] S. Polinder, E. M. Verschuur, P. D. Siersema, E. J. Kuipers e E. W. Steyerberg, «Cost comparison study of two different follow-up protocols after surgery for oesophageal cancer,» *Eur J Cancer*, vol. 45, n. 12, pp. 2110-5, 2009.
- [121] G. B. Secco, R. Fardelli, D. Gianquinto, P. Bonfante, E. Baldi, Ravera G e et al., «Efficacy and cost of risk-adapted follow-up in patients after colorectal cancer surgery: a prospective, randomized and controlled trial,» *Eur J Surg Oncol*, vol. 28, n. 4, pp. 418-23, 2002.

- [122] A. Siddika, D. Tolia-Shah, T. E. Pearson, N. G. Richardson e A. H. Ross, «Remote surveillance after colorectal cancer surgery: an effective alternative to standard clinic-based follow-up,» *Colorectal Dis*, vol. 17, n. 10, pp. 870-5, 2015.
- [123] E. Strand, I. Nygren, L. Bergkvist e K. Smedh, «Nurse or surgeon follow-up after rectal cancer: a randomized trial,» *Colorectal Dis*, vol. 13, n. 9, pp. 999-1003, 2011.
- [124] A. I. Tergas, L. J. Havrilesky, A. N. Fader, S. R. Guntupalli, W. K. Huh, L. S. Massad e et al., «Cost analysis of colposcopy for abnormal cytology in post-treatment surveillance for cervical cancer,» *Gynecol Oncol*, vol. 130, n. 3, pp. 421-5, 2013.
- [125] C. W. Tzeng, D. E. Abbott, S. B. Cantor, J. B. Fleming, J. E. Lee, P. W. Pisters e et al., «Frequency and intensity of postoperative surveillance after curative treatment of pancreatic cancer: a cost-effectiveness analysis,» *Ann Surg Oncol*, vol. 20, n. 7, pp. 2197-203, 2013.
- [126] J. van Loon, J. P. Grutters, R. Wanders, L. Boersma, A. M. Dingemans, G. Bootsma e et al., «18FDG-PET-CT in the follow-up of non-small cell lung cancer patients after radical radiotherapy with or without chemotherapy: an economic evaluation,» *Eur J Cancer*, vol. 46, n. 1, pp. 110-9, 2010.
- [127] C. J. Verberne, T. Wiggers, I. Grossmann, G. H. de Bock e K. M. Vermeulen, «Cost-effectiveness of a carcinoembryonic antigen (CEA) based follow-up programme for colorectal cancer (the CEA Watch trial),» *Colorectal Dis*, vol. 18, n. 3, pp. O91-6, 2016.
- [128] E. M. Verschuur, E. W. Steyerberg, H. W. Tilanus, S. Polinder, M. L. Essink-Bot, K. T. Tran e et al., «Nurse-led follow-up of patients after oesophageal or gastric cardia cancer surgery: a randomised trial,» *Br J Cancer*, vol. 100, pp. 70-6, 2009.
- [129] A. Visser, H. W. van Laarhoven, P. H. Govaert, M. S. Schlooz, L. Jansen, T. van Dalen e et al., «Group medical consultations in the follow-up of breast cancer: a randomized feasibility study,» *J Cancer Surviv*, vol. 9, n. 3, pp. 450-61, 2015.
- [130] J. X. Wu, C. E. Beni, K. A. Zanocco, C. Sturgeon e M. W. Yeh, «Cost-effectiveness of long-term every three-year versus annual postoperative surveillance for low-risk papillary thyroid cancer,» *Thyroid*, vol. 25, n. 7, pp. 797-803, 2015.
- [131] M. J. Edelman, F. J. Meyers e D. Siegel, «The utility of follow-up testing after curative cancer therapy. A critical review and economic analysis,» *J Gen Intern Med*, vol. 12, n. 5, pp. 318-31, 1997.
- [132] N. Hex e C. Bartlett, «Cost-effectiveness of follow-up services for childhood cancer survivors outside the cancer setting,» *Curr Opin Support Palliat Care*, vol. 7, n. 3, pp. 314-7, 2013.
- [133] M. Yang, D. S. Patel, W. Tufail e A. M. Issa, «The quality of economic studies of cancer pharmacogenomics: a quantitative appraisal of the evidence,» *Expert Rev Pharmacoecon Outcomes Res*, vol. 13, n. 5, pp. 597-611, 2013.

- [134] M. F. Drummond e T. O. Jefferson, «Guidelines for authors and peer reviewers of economic submissions to the BMJ,» *BMJ*, vol. 313, pp. 275-83, 1996.
- [135] P. J. Neumann, S. J. Goldie e M. C. Weinstein, «Preference-based measures in economic evaluation in health care,» *Annu Rev Public Health*, vol. 21, pp. 587-611, 2000.
- [136] S. E. Wolowacz, A. Briggs, V. Belozeroff, P. Clarke, L. Doward, R. Goeree e et al., «Estimating health-state utility for economic models in clinical studies: an ISPOR Good Research Practices Task Force Report,» *Value Health*, vol. 19, n. 6, pp. 704-19, 2016.
- [137] A. A. Rashidi, A. H. Anis e C. A. Marra, «Do visual analogue scale (VAS) derived standard gamble (SG) utilities agree with Health Utilities Index utilities? A comparison of patient and community preferences for health status in rheumatoid arthritis patients,» *Health Qual Life Outcomes*, vol. 4, pp. 1-10, 2006.
- [138] A. G. Witney, G. J. Treharne, M. Tavakoli, A. C. Lyons, K. Vincent, D. L. Scott e et al., «The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis,» *Rheumatology (Oxford)*, vol. 45, n. 8, pp. 975-81, 2006.
- [139] P. Dolan e M. Sutton, «Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values,» *Soc Sci Med*, vol. 44, n. 10, pp. 1519-30, 1997.
- [140] S. J. Whitehead e S. Ali, «Health outcomes in economic evaluation: the QALY and utilities,» *Br Med Bull*, vol. 96, n. 1, pp. 5-21, 2010.
- [141] P. Krabbe, *The measurement of health and health status*, Academic Press, 2016.
- [142] M. W. Post, «Definitions of quality of life: what has happened and how to move on,» *Top Spinal Cord Inj Rehabil*, vol. 20, n. 3, pp. 167-80, 2014.
- [143] W. Wiedermann e U. Frick, «Using surveys to calculate disability-adjusted life-year,» *Alcohol Res*, vol. 35, n. 2, pp. 128-33, 2013.
- [144] A. E. Attema, Y. Edelaar-Peeters, M. M. Versteegh e E. A. Stolk, «Time trade-off: one methodology, different methods,» *Eur J Health Econ*, vol. 14, n. Suppl 1, pp. S53-64, 2013.
- [145] J. Richardson, M. A. Khan, A. Iezzi e A. Maxwell, «Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments,» *Med Decis Making*, vol. 35, n. 3, pp. 276-91, 2015.
- [146] A. Beaudet, J. Clegg, P. O. Thuresson, A. Lloyd e P. McEwan, «Review of utility values for economic modeling in type 2 diabetes,» *Value Health*, vol. 17, n. 4, pp. 462-70, 2014.

- [147] European Network for Health Technology Assessment (EUnetHTA), «Methods for health economic evaluations - A guideline based on current practices in Europe,» May 2015. [Online]. Available: http://www.eunetha.eu/wp-content/uploads/2018/03/Methods_for_health_economic_evaluations.pdf.
- [148] Y. Hao, V. Wolfram e J. Cook, «A structured review of health utility measures and elicitation in advanced/metastatic breast cancer,» *Clinicoecon Outcomes Res*, vol. 8, pp. 293-303, 2016.
- [149] K. S. Boye, L. S. Matza, D. H. Feeny, J. A. Johnston, L. Bowman e J. B. Jordan, «Challenges to time trade-off utility assessment methods: when should you consider alternative approaches?,» *Expert Rev Pharmacoecon Outcomes Res*, vol. 14, n. 3, pp. 437-50, 2014.
- [150] J. Horsman, W. Furlong, D. Feeny e G. Torrance, «The health utilities index (HUI®): concepts, measurement properties and applications,» *Health Qual Life Outcomes*, vol. 1, pp. 1-13, 2003.
- [151] B. Button, «Quality of well-being (QWB) scale,» in *Encyclopedia of Quality of Life and Well-Being Research*, Springer, Dordrecht, 2014.
- [152] G. Hawthorne, J. Richardson e R. Osborne, «The assessment of quality of life (AQoL) instrument: a psychometric measure of health-related quality of life,» *Qual Life Res*, vol. 8, n. 3, pp. 209-24, 1999.
- [153] A. J. Wailoo, M. Hernandez-Alava, A. Manca, A. Mejia, J. Ray, B. Crawford e et al., «Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR Good Practices for Outcomes Research Task Force Report,» *Value Health*, vol. 20, n. 1, pp. 18-27, 2017.
- [154] L. Longworth e D. Rowen, «Mapping to obtain EQ-5D utility values for use in NICE health technology assessments,» *Value Health*, vol. 16, n. 1, pp. 202-10, 2013.
- [155] S. M. Szabo, R. L. Dobson, B. M. Donato, G. L'Italien, S. J. Hotte e A. R. Levy, «The quality-of-life impact of head and neck cancer: preference values from the Canadian general public,» *Health Outcomes Res Med*, vol. 3, n. 1, pp. e11-23, 2012.
- [156] M. B. Hannouf, C. Sehgal, J. Q. Cao, J. D. Mocanu, E. Winqvist e G. S. Zaric, «Cost-effectiveness of adding cetuximab to platinum-based chemotherapy for first line treatment of recurrent or metastatic head and neck cancer,» *PLoS One*, vol. 7, n. 6, pp. 1-9, 2012.
- [157] N. L. Liberato, C. Rognoni, S. Rubrichi, S. Quaglini, M. Marchetti, T. Gorlia e et al., «Adding docetaxel to cisplatin and fluorouracil in patients with unresectable head and neck cancer: a cost-utility analysis,» *Ann Oncol*, vol. 23, n. 7, pp. 1825-32, 2012.
- [158] Y. Komatsuzaki, P. Gramegna, J. M. Stephens, M. F. Botteman, C. L. Pashos e A. Redaelli, «Preferences and utilities of health outcomes and treatments associated with head and neck cancer: a systematic review,» *Am J Cancer*, vol. 5, n. 1, pp. 27-34, 2006.

- [159] C. J. Sampson, J. C. Tosh, C. P. Cheyne, D. Broadbent e M. James, «Health state utility values for diabetic retinopathy: protocol for a systematic review and meta-analysis,» *Syst Rev*, vol. 4, pp. 1-6, 2015.
- [160] D. Papaioannou, J. Brazier e S. Paisley, «Systematic searching and selection of health state utility values from the literature,» *Value Health*, vol. 16, pp. 686-95, 2013.
- [161] K. Jeong e J. Cairns, «Systematic review of health state utility values for economic evaluation of colorectal cancer,» *Health Econ Rev*, vol. 6, n. 1, pp. 1-10, 2016.
- [162] A. Gheorghe, G. Moran, H. Duffy, T. Roberts, T. Pinkney e M. Calvert, «Health utility values associated with surgical site infection: a systematic review,» *Value Health*, vol. 18, n. 8, pp. 1126-37, 2015.
- [163] «CEA Registry,» 2016. [Online]. Available: <http://healtheconomics.tuftsmedicalcenter.org/cear4/home.aspx>.
- [164] «SchARRHUD Health Utilities Database,» 2016. [Online]. Available: <https://www.scharrhud.org/>.
- [165] H. Dakin, «Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database,» *Health Qual Life Outcomes*, vol. 11, pp. 1-6, 2013.
- [166] Health Economics Research Centre (HERC), University of Oxford, «HERC database of mapping studies, Version 5.0,» [Online]. Available: <http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies>.
- [167] J. Smith-Palmer, J. P. Bae, K. S. Boye, K. Norrbacka, B. Hunt e W. J. Valentine, «Evaluating health-related quality of life in type 1 diabetes: a systematic literature review of utilities for adults with type 1 diabetes,» *Clinicoecon Outcomes Res*, vol. 8, pp. 559-71, 2016.
- [168] G. W. Torrance, D. Feeny e W. Furlong, «Visual analog scales: do they have a role in the measurement of preferences for health states?,» *Med Decis Mak*, vol. 21, n. 4, pp. 329-34, 2001.
- [169] J. E. Fincham, «Response rates and responsiveness for surveys, standards, and the journal,» *Am J Pharm Educ*, vol. 72, n. 2, pp. 1-3, 2008.
- [170] N. Linendoll, T. Saunders, R. Burns, J. D. Nyce, K. B. Wendell, A. M. Evens e et al., «Health-related quality of life in Hodgkin lymphoma: a systematic review,» *Health Qual Life Outcomes*, vol. 14, n. 1, pp. 1-14, 2016.
- [171] A. Marcellusi, A. Capone, G. Favato, F. S. Mennini, G. Baio, K. Haeussler e et al., «Health utilities lost and risk factors associated with HPV-induced diseases in men and women: the HPV Italian collaborative study group,» *Clin Ther*, vol. 37, n. 1, pp. 156-67, 2015.

- [172] C. W. Noel, D. J. Lee, Q. Kong, W. Xu, C. Simpson, D. Brown e et al., «Comparison of health state utility measures in patients with head and neck cancer,» *JAMA Otolaryngol Head Neck Surg*, vol. 141, n. 8, pp. 696-703, 2015.
- [173] E. L. Conway, K. C. Farmer, W. J. Lynch, G. L. Rees, G. Wain e J. Adams, «Quality of life valuations of HPV-associated cancer health states by the general population,» *Sex Transm Infect*, vol. 88, n. 7, pp. 517-21, 2012.
- [174] D. W. Hamilton, J. E. Bins, P. McMeekin, A. Pedersen, N. Steen, A. De Soyza e et al., «Quality compared to quantity of life in laryngeal cancer: A time trade-off study,» *Head Neck*, vol. 38, n. Suppl 1, pp. E631-7, 2016.
- [175] C. S. Hollenbeak, V. J. Lowe e B. C. Stack, «The cost-effectiveness of fluorodeoxyglucose 18-F positron emission tomography in the N0 neck,» *Cancer*, vol. 92, n. 9, pp. 2341-8, 2001.
- [176] V. Jalukar, G. F. Funk, A. J. Christensen, L. H. Karnell e P. J. Moran, «Health states following head and neck cancer treatment: patient, health-care professional, and public perspectives,» *Head Neck*, vol. 20, n. 7, pp. 600-8, 1998.
- [177] J. Ringash, D. A. Redelmeier, B. O'Sullivan e A. Bezjak, «Quality of life and utility in irradiated laryngeal cancer patients,» *Int J Radiat Oncol Biol Phys*, vol. 47, n. 4, pp. 875-81, 2000.
- [178] M. H. Weiss, L. B. Harrison e R. S. Isaacs, «Use of decision analysis in planning a management strategy for the stage N0 neck,» *Arch Otolaryngol Head Neck*, vol. 120, n. 7, pp. 699-702, 1994.
- [179] J. R. de Almeida, N. L. Villanueva, A. J. Moskowitz, B. A. Miles, M. S. Teng, A. Sikora e et al., «Preferences and utilities for health states after treatment for oropharyngeal cancer: transoral robotic surgery versus definitive (chemo)radiotherapy,» *Head Neck*, vol. 36, n. 7, pp. 923-33, 2014.
- [180] H. A. Llewellyn-Thomas, H. J. Sutherland e E. C. Thiel, «Do patients' evaluations of a future health state change when they actually enter that state?,» *Med Care*, vol. 31, n. 11, pp. 1002-12, 1993.
- [181] J. van der Donk, P. C. Levendag, A. J. Kuijpers, F. H. Roest, J. D. Habbema, C. A. Meeuwis e et al., «Patient participation in clinical decision-making for treatment of T3 laryngeal cancer: a comparison of state and process utilities,» *J Clin Oncol*, vol. 13, n. 9, pp. 2369-78, 1995.
- [182] E. Del Barco Morillo, R. Mesía, J. C. Adansa Klain, S. Vázquez Fernández, J. Martínez-Galán, M. Pastor Borgoñon e et al., «Phase II study of panitumumab and paclitaxel as first-line treatment in recurrent or metastatic head and neck cancer. TTCC-2009-03/VECTITAX study,» *Oral Oncol*, vol. 62, pp. 54-9, 2016.

- [183] N. Ouattassi, N. Benmansour, S. El Fakir, C. Nejari e M. N. Alami, «Translation and validation of EORTC QLQ-H&N 35 into Moroccan Arabic for ENT head and neck cancer patients in Morocco,» *Eur Arch Otorhinolaryngol*, vol. 273, n. 9, pp. 2727-34, 2016.
- [184] C. Parrilla, A. Minni, H. Bogaardt, G. F. Macri, M. Battista, R. Roukos e et al., «Pulmonary rehabilitation after total laryngectomy: a multicenter time-series clinical trial evaluating the Provox XtraHME in HME-Naïve patients,» *Ann Otol Rhinol Laryngol*, vol. 124, n. 9, pp. 706-13, 2015.
- [185] S. N. Rogers, R. D. Miller, K. Ali, A. B. Minhas, H. F. Williams e D. Lowe, «Patients' perceived health status following primary surgery for oral and oropharyngeal cancer,» *Int J Oral Maxillofac Surg*, vol. 35, n. 10, pp. 913-9, 2006.
- [186] B. L. Ramaekers, M. A. Joore, J. P. Grutters, P. van den Ende, J. Jong, R. Houben e et al., «The impact of late treatment-toxicity on generic health-related quality of life in head and neck cancer patients after radiotherapy,» *Oral Oncol*, vol. 47, n. 8, pp. 768-74, 2011.
- [187] A. S. Pickard, R. Jiang, H. W. Lin, S. Rosenbloom e D. Cella, «Using patient-reported outcomes to compare relative burden of cancer: EQ-5D and functional assessment of cancer therapy-general in eleven types of cancer,» *Clin Ther*, vol. 38, n. 4, pp. 769-77, 2016.
- [188] M. T. Truong, Q. Zhang, D. I. Rosenthal, M. List, R. Axelrod, E. Sherman e et al., «Quality of life and performance status from a substudy conducted within a prospective phase 3 randomized trial of concurrent accelerated radiation plus cisplatin with or without cetuximab for locally advanced head and neck carcinoma: NRG Oncology Radiatio,» *Int J Radiat Oncol Biol Phys*, vol. 97, n. 4, pp. 687-99, 2017.
- [189] T. M. Govers, W. H. Schreuder, W. M. Klop, J. P. Grutters, M. M. Rovers, M. A. Merks e et al., «Quality of life after different procedures for regional control in oral cancer patients: cross-sectional survey,» *Clin Otolaryngol*, vol. 41, n. 3, pp. 228-33, 2016.
- [190] L. Pottel, M. Lycke, T. Boterberg, H. Pottel, L. Goethals, F. Duprez e et al., «G-8 indicates overall and quality-adjusted survival in older head and neck cancer patients treated with curative radio-chemotherapy,» *BMC Cancer*, vol. 15, pp. 1-11, 2015.
- [191] K. Aro, L. Bäck, V. Loimu, K. Saarilahti, S. Rogers, H. Sintonen e et al., «Trends in the 15D health-related quality of life over the first year following diagnosis of head and neck cancer,» *Eur Arch Otorhinolaryngol*, vol. 273, n. 8, pp. 2141-50, 2016.
- [192] K. M. Higgins, «What treatment for early-stage glottic carcinoma among adult patients: CO2 endolaryngeal laser excision versus standard fractionated external beam radiation is superior in terms of cost utility?,» *Laryngoscope*, vol. 121, n. 1, pp. 116-34, 2011.
- [193] E. E. Kent, A. Ambs, S. A. Mitchell, S. B. Clauser, A. W. Smith e R. D. Hays, «Health-related quality of life in older adult survivors of selected cancers: data from the SEER-MHOS linkage,» *Cancer*, vol. 121, n. 5, pp. 758-65, 2015.

- [194] V. Loimu, A. A. Mäkitie, L. J. Bäck, H. Sintonen, P. Räsänen, R. Roine e et al., «Health-related quality of life of head and neck cancer patients with successful oncological treatment,» *Eur Arch Otorhinolaryngol*, vol. 272, n. 9, pp. 2415-23, 2015.
- [195] K. K. Chan, A. R. Willan, M. Gupta e E. Pullenayegum, «Underestimation of uncertainties in health utilities derived from mapping algorithms involving health-related quality-of-life measures: statistical explanations and potential remedies,» *Med Decis Mak*, vol. 34, n. 7, pp. 863-72, 2014.
- [196] A. Parthan, M. R. Posner, C. Brammer, P. Beltran e J. P. Jansen, «Cost utility of docetaxel as induction chemotherapy followed by chemoradiation in locally advanced squamous cell carcinoma of the head and neck,» *Head Neck*, vol. 31, n. 10, pp. 1255-62, 2009.
- [197] P. F. Krabbe, L. Peerenboom, B. S. Langenhoff e T. J. Ruers, «Responsiveness of the generic EQ-5D summary measure compared to the disease-specific EORTC QLQ C-30,» *Qual Life Res*, vol. 13, n. 7, pp. 1247-53, 2004.
- [198] J. B. Vermorken, E. Remenar, C. van Herpen, T. Gorlia, R. Mesia, M. Degardin, et al. e EORTC 24971/TAX 323 Study Group., «Cisplatin, fluorouracil, and docetaxel in unresectable head and neck cancer,» *N Engl J Med*, vol. 357, n. 17, pp. 1695-704, 2007.
- [199] J. H. Yong, J. Beca, B. O'Sullivan, S. H. Huang, T. McGowan, P. Warde e et al., «Cost-effectiveness of intensity-modulated radiotherapy in oropharyngeal cancer,» *Clin Oncol (R Coll Radiol)*, vol. 24, n. 7, pp. 532-8, 2012.
- [200] M. B. Nichol, N. Sengupta e D. R. Globe, «Evaluating quality-adjusted life years: estimation of the health utility index (HUI2) from the SF-36,» *Med Decis Mak*, vol. 21, n. 2, pp. 105-12, 2001.
- [201] E. H. Pow, D. L. Kwong, A. S. McMillan, M. C. Wong, J. S. Sham, L. H. Leung e et al., «Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: initial report on a randomized controlled clinical trial,» *Int J Radiat Oncol Biol Phys*, vol. 66, n. 4, pp. 981-91, 2006.
- [202] J. Heiskanen, A. M. Tolppanen, R. P. Roine, J. Hartikainen, M. Hippeläinen, H. Miettinen e et al., «Comparison of EQ-5D and 15D instruments for assessing the health-related quality of life in cardiac surgery patients,» *Eur Heart J Qual Care Clin Outcomes*, vol. 2, n. 3, pp. 193-200, 2016.
- [203] T. Vainiola, V. Pettilä, R. P. Roine, P. Räsänen, A. M. Rissanen e H. Sintonen, «Comparison of two utility instruments, the EQ-5D and the 15D, in the critical care setting,» *Intensive Care Med*, vol. 36, n. 12, pp. 2090-3, 2010.
- [204] L. Lunde, «Can EQ-5D and 15D be used interchangeably in economic evaluations? Assessing quality of life in post-stroke patients,» *Eur J Health Econ*, vol. 14, pp. 539-50, 2013.

- [205] P. Vartiainen, P. Mäntyselkä, T. Heiskanen, N. Hagelberg, S. Mustola, H. Forssell e et al., «Validation of EQ-5D and 15D in the assessment of health-related quality of life in chronic pain,» *Pain*, vol. 158, n. 8, pp. 1577-85, 2017.
- [206] D. Arnold, A. Girling, A. Stevens e R. Lilford, «Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis,» *BMJ*, vol. 339, pp. 1-8, 2009.
- [207] S. L. Dennett, K. S. Boye e N. R. Yurgin, «The impact of body weight on patient utilities with or without type 2 diabetes: a review of the medical literature,» *Value Health*, vol. 11, n. 3, pp. 478-86, 2008.
- [208] C. M. McDonough e A. N. Tosteson, «Measuring preferences for cost-utility analysis: how choice of method may influence decision-making,» *Pharmacoeconomics*, vol. 25, n. 2, pp. 93-106, 2007.
- [209] J. Brazier e D. Rowen, «NICE DSU Technical Support Document 11: Alternatives to EQ-5D for Generating Health State Utility Values [Internet],» NICE Decision Support Unit Technical Support Documents, 2011.
- [210] National Institute for Health and Care Excellence, «Position statement on use of the EQ-5D-5L valuation set,» August 2017. [Online]. Available: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisal-guidance/eq5d5l_nice_position_statement.pdf. [Consultato il giorno August 2017].
- [211] P. M. Fayers, N. K. Aaronson, K. Bjordal, M. Groenvold, D. Curran, A. Bottomley e on behalf of the EORTC Quality of Life Group, The EORTC QLQ-C30 scoring manual, 3rd edition, Brussels: European Organization for Research and Treatment of Cancer (EORTC), 2001.
- [212] M. A. Weitzner, C. A. Meyers, C. K. Gelke, K. S. Byrne, D. F. Cella e V. A. Levin, «The functional assessment of cancer therapy (FACT) scale. Development of a brain subscale and revalidation of the general version (FACT-G) in patients with primary brain tumors,» *Cancer*, vol. 75, n. 5, pp. 1151-61, 1995.
- [213] R. Crott e A. Briggs, «Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences,» *Eur J Health Econ*, vol. 11, n. 4, pp. 427-34, 2010.
- [214] L. McKenzie e M. van der Pol, «Mapping the EORTC QLQ C-30 onto the EQ-5D instrument: the potential to estimate QALYs without generic preference data,» *Value Health*, vol. 12, n. 1, pp. 167-71, 2009.
- [215] L. Longworth, Y. Yang, T. Young, B. Mulhern, M. Hernández Alava, C. Mukuria e et al., «Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey,» *Health Technol Assess*, vol. 18, n. 9, pp. 1-224, 2014.

- [216] I. Khan e S. Morris, «A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches,» *Health Qual Life Outcomes*, vol. 12, pp. 1-16, 2014.
- [217] National Institute for Health and Care Excellence (NICE), «Guide to the methods of technology appraisal 2013,» 4 April 2013. [Online]. Available: <https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781>.
- [218] T. A. Young, C. Mukuria, D. Rowen, J. E. Brazier e L. Longworth, «Mapping functions in health-related quality of life: mapping from two cancer-specific health-related quality-of-life instruments to EQ-5D-3L,» *Med Decis Making*, vol. 35, n. 7, pp. 912-26, 2015.
- [219] R. Crott, «Mapping algorithms from QLQ-C30 to EQ-5D utilities: no firm ground to stand on yet,» *Expert Rev Pharmacoecon Outcomes Res*, vol. 14, n. 4, pp. 569-76, 2014.
- [220] A. Wailoo, M. Hernández, C. Philips, S. Brophy e S. Siebert, «Modeling health state utility values in ankylosing spondylitis: comparisons of direct and indirect methods,» *Value Health*, vol. 18, n. 4, pp. 425-31, 2015.
- [221] S. Rathod, J. Livergant, J. Klein, I. Witterick e J. Ringash, «A systematic review of quality of life in head and neck cancer treated with surgery with or without adjuvant treatment,» *Oral Oncol*, vol. 51, n. 10, pp. 888-900, 2015.
- [222] S. M. Parkar e M. N. Shah, «A relationship between quality-of-life and head and neck cancer: A systemic review,» *South Asian J Cancer*, vol. 4, n. 4, pp. 179-82, 2015.
- [223] S. Dunne, O. Mooney, L. Coffey, L. Sharp, D. Desmond e C. Timon, «Psychological variables associated with quality of life following primary treatment for head and neck cancer: a systematic review of the literature from 2004 to 2015,» *Psychooncology*, vol. 26, n. 2, pp. 149-60, 2017.
- [224] S. Petrou, O. Rivero-Arias, H. Dakin, L. Longworth, M. Oppe e R. Froud, «Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement,» *Int J Technol Assess Health Care*, vol. 31, n. 4, pp. 230-5, 2015.
- [225] S. Petrou, O. Rivero-Arias, H. Dakin, L. Longworth, M. Oppe, R. Froud e et al., «Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement,» *Qual Life Res*, vol. 25, n. 2, pp. 275-81, 2016.
- [226] I. Khan, S. Morris, N. Pashayan, B. Matata, Z. Bashir e J. Maguirre, «Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell lung cancer patients,» *Health Qual Life Outcomes*, vol. 14, pp. 1-15, 2016.
- [227] N. J. Devlin, K. K. Shah, Y. Feng, B. Mulhern e B. van Hout, «Valuing health-related quality of life: An EQ-5D-5L value set for England,» *Health Econ*, vol. 27, n. 1, pp. 7-22, 2018.

- [228] M. M. Versteegh, K. M. Vermeulen, S. M. Evers, G. A. de Wit, R. Prenger e E. A. Stolk, «Dutch tariff for the five-level version of EQ-5D,» *Value Health*, vol. 19, n. 4, pp. 343-52, 2016.
- [229] F. Xie, E. Pullenayegum, K. Gaebel, N. Bansback, S. Bryan, A. Ohinmaa, et al. e on behalf of the Canadian EQ-5D-5L Valuation Study, «A time trade-off-derived value set of the EQ-5D-5L for Canada,» *Med Care*, vol. 54, n. 1, pp. 98-105, 2016.
- [230] F. Augustovski, L. Rey-Ares, V. Irazola, O. U. Garay, O. Gianneo, G. Fernández e et al., «An EQ-5D-5L value set based on Uruguayan population preferences,» *Qual Life Res*, vol. 25, n. 2, pp. 323-33, 2016.
- [231] T. Shiroyiwa, S. Ikeda, S. Noto, A. Igarashi, T. Fukuda, S. Saito e et al., «Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan,» *Value Health*, vol. 19, n. 5, pp. 648-54, 2016.
- [232] S. H. Kim, J. Ahn, M. Ock, S. Shin, J. Park, N. Luo e et al., «The EQ-5D-5L valuation study in Korea,» *Qual Life Res*, vol. 25, n. 7, pp. 1845-52, 2016.
- [233] N. Luo, G. Liu, M. Li, H. Guan, X. Jin e K. Rand-Hendriksen, «Estimating an EQ-5D-5L Value Set for China,» *Value Health*, vol. 20, n. 4, pp. 662-9, 2017.
- [234] M. Hernández Alava, A. J. Wailoo e R. Ara, «Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values,» *Value Health*, vol. 15, n. 3, pp. 550-61, 2012.
- [235] M. Hernández Alava e A. J. Wailoo, «Fitting adjusted limited dependent variable mixture models to EQ-5D,» *The Stata Journal*, vol. 15, n. 3, pp. 737-50, 2015.
- [236] A. Wailoo, M. Hernandez-Alava e A. Escobar Martinez, «Modelling the relationship between the WOMAC Osteoarthritis Index and EQ-5D,» *Health Qual Life Outcomes* , vol. 12, pp. 1-6, 2014.
- [237] E. J. Kim, S. K. Ko e H. Y. Kang, «Mapping the cancer-specific EORTC QLQ-C30 and EORTC QLQ-BR23 to the generic EQ-5D in metastatic breast cancer patients,» *Qual Life Res*, vol. 21, n. 7, pp. 1193-203, 2012.
- [238] H. Dakin, L. Abel, R. Burns e Y. Yang, «Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement,» *Health Qual Life Outcomes*, vol. 16, n. 1, pp. 1-9, 2018.
- [239] Health Economics Research Centre (HERC), University of Oxford, «HERC database of mapping studies, Version 6.0,» 17 January 2017. [Online]. Available: <https://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies>.
- [240] H. Suzuki , A. Asakawa, H. Amitani, N. Nakamura e A. Inui, «Cancer cachexia— pathophysiology and management,» *J Gastroenterol*, vol. 48, n. 5, pp. 574-594, 2013.

- [241] B. Gupta, N. W. Johnson e N. Kumar, «Global epidemiology of head and neck cancers: A continuing challenge,» *Oncology*, vol. 91, n. 1, pp. 13-23, 2016.
- [242] J. Péron, V. Polivka, S. Chabaud, M. Poupert, P. Ceruse, A. Ramade e et al., «An effective and well-tolerated strategy in recurrent and/or metastatic head and neck cancer: successive lines of active chemotherapeutic agents.,» *BMC Cancer*, vol. 14, pp. 1-8, 2014.
- [243] K. P. Pendleton e J. R. Grandis, «Cisplatin-based chemotherapy options for recurrent and/or metastatic squamous cell cancer of the head and neck,» *Clin Med Insights Ther*, vol. 2013, n. 5, 2013.
- [244] K. Harrington, S. Temam, H. Mehanna, A. D'Cruz, M. Jain, I. D'Onofrio e et al., «Postoperative adjuvant lapatinib and concurrent chemoradiotherapy followed by maintenance lapatinib monotherapy in high-risk patients with resected squamous cell carcinoma of the head and neck: A phase III, randomized, double-blind, placebo-controlled stu,» *J Clin Oncol*, vol. 33, n. 35, pp. 4202-9, 2015.
- [245] R. Al-Wassia, S. Vakilian, C. Holly, K. Sultanem e G. Shenouda, «A retrospective study of head and neck re-irradiation for patients with recurrent or second primary head and neck cancer: the McGill University experience,» *J Otolaryngol Head Neck Surg*, vol. 44, pp. 1-8, 2015.
- [246] T. Kurzweg, N. Möckelmann, S. Laban e R. Knecht, «Current treatment options for recurrent/metastatic head and neck cancer: a post-ASCO 2011 update and review of last year's literature,» *Eur Arch Otorhinolaryngol*, vol. 269, n. 10, pp. 2157-67, 2012.
- [247] J. B. Vermorken, R. Mesia, F. Rivera, E. Remenar, A. Kawecki, S. Rottey e et al., «Platinum-based chemotherapy plus cetuximab in head and neck cancer,» *N Engl J Med*, vol. 359, n. 11, pp. 1116-27, 2008.
- [248] M. Imbimbo, P. Bossi, P. Potepan, C. Fallai, G. Scaramellini, L. Locati e et al., «OC-039: surveillance of head and neck cancer (HNC) patients: clinical and radiological exams or symptom driven follow up?,» *Radiother Oncol*, vol. 106, n. Suppl 1, p. S16, 2013.
- [249] V. P. Retèl, L. van der Molen, F. J. Hilgers, C. R. Rasch, A. A. L'Ortye, L. M. Steuten e et al., «A cost-effectiveness analysis of a preventive exercise program for patients with advanced head and neck cancer treated with concomitant chemo-radiotherapy,» *BMC Cancer*, vol. 11, pp. 1-9, 2011.
- [250] S. Temam, E. Pape, F. Janot, P. Wibault, M. Julieron, A. Lusinchi e et al., «Salvage surgery after failure of very accelerated radiotherapy in advanced head-and-neck squamous cell carcinoma,» *Int J Radiat Oncol Biol Phys*, vol. 62, n. 4, pp. 1078-83, 2005.
- [251] M. E. Zafereo, M. M. Hanasono, D. I. Rosenthal, E. M. Sturgis, J. S. Lewin, D. B. Roberts e et al., «The role of salvage surgery in patients with recurrent squamous cell carcinoma of the oropharynx,» *Cancer*, vol. 115, n. 24, pp. 5723-33, 2009.

- [252] N. Penel, E. Y. Amela, Y. Mallet, D. Lefebvre, S. Clisant, A. Kara e et al., «A simple predictive model for postoperative mortality after head and neck cancer surgery with opening of mucosa,» *Oral Oncol*, vol. 43, n. 2, pp. 174-80, 2007.
- [253] M. C. Biagioli, M. Harvey, E. Roman, L. E. Raez, A. H. Wolfson, S. Mutyala e et al., «Intensity-modulated radiotherapy with concurrent chemotherapy for previously irradiated, recurrent head and neck cancer,» *Int J Radiat Oncol Biol Phys*, vol. 69, n. 4, pp. 1067-73, 2007.
- [254] D. J. Sher, R. I. Haddad, C. M. Norris, M. R. Posner, L. J. Wirth, L. A. Goguen e et al., «Efficacy and toxicity of reirradiation using intensity-modulated radiotherapy for recurrent or second primary head and neck cancer,» *Cancer*, vol. 116, n. 20, pp. 4761-8, 2010.
- [255] Y. W. Won, Y. H. Park, M. J. Ahn, I. G. Do, Y. H. Ko e K. Park, «A phase II study of combination chemotherapy with capecitabine and cisplatin in patients with metastatic or recurrent squamous cell carcinoma of the head and neck,» *Ann Oncol*, vol. 22, n. 2, pp. 417-23, 2011.
- [256] Istituto Nazionale di Statistica (ISTAT), «Tavole di mortalità per classi di età,» 2015. [Online]. Available: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_MORTALITA1.
- [257] A. Briggs, M. Sculpher e K. Claxton, Decision modelling for health economic evaluation (Handbooks in health economic evaluation), 1st edition, Oxford: Oxford University Press, 2006.
- [258] M. Beccaro, M. Costantini, D. F. Merlo e I. S. Group, «Inequity in the provision of and access to palliative care for cancer patients. Results from the Italian survey of the dying of cancer (ISDOC),» *BMC Public Health*, vol. 7, pp. 1-13, 2007.
- [259] D. Z. Louis, S. E. Hegarty, M. Leoni, R. De Palma, S. Varga e R. Melotti, «Variation in hospital utilization at the end of life for patients with cancer in the Emilia-Romagna region of Italy,» *Tumori*, vol. 102, n. 6, pp. 614-20, 2016.
- [260] R. Pannuti, «Qualità di vita del malato terminale e sostenibilità delle cure palliative: l'integrazione tra sanità pubblica e privato sociale,» *Evidence*, vol. 7, n. 9, pp. 1-2, 2015.
- [261] S. Francisci, S. Guzzinati, M. Mezzetti, E. Crocetti, F. Giusti, G. Miccinesi e et al., «Cost profiles of colorectal cancer patients in Italy based on individual patterns of care,» *BMC Cancer*, vol. 13, pp. 1-11, 2013.
- [262] Gruppo di lavoro Associazione Italiana di Economia Sanitaria (AIES) coordinato da Giovanni Fattore, «Proposta di linee guida per la valutazione economica degli interventi sanitari in Italia,» *Pharmacoeconomics Ital Res Articles*, vol. 11, n. 2, pp. 83-93, 2009.
- [263] P. J. Neumann, T. G. Ganiats, G. D. Sanders, L. B. Russell e J. E. Siegel, Cost-effectiveness in health and medicine. 2nd ed, Oxford: Oxford University Press, 2017.

- [264] J. B. Oostenbrink, M. J. Al, M. Oppe e M. P. Rutten-van Mólken, «Expected value of perfect information: an empirical example of reducing decision uncertainty by conducting additional research,» *Value Health*, vol. 11, n. 7, pp. 1070-80, 2008.
- [265] Associazione Italiana Registro Tumori (AIRTUM), «Italian cancer figures - Report 2015. Epithelial tumours of head and neck.,» 2015. [Online]. Available: http://www.registri-tumori.it/PDF/AIRTUM2016/TUMORIRARI/AIRTUM_RARI_S001_headandneck.pdf.
- [266] J. E. Leeman, J. Li, X. Pei, P. Venigalla, Z. S. Zumsteg, E. Katsoulakis e et al., «Patterns of treatment failure and postrecurrence outcomes among patients with locally advanced head and neck squamous cell carcinoma after chemoradiotherapy using modern radiation techniques,» *JAMA Oncol*, vol. 3, n. 11, pp. 1487-94, 2017.
- [267] P. R. Patel e J. K. Salama, «Reirradiation for recurrent head and neck cancer,» *Expert Rev Anticancer Ther*, vol. 12, n. 9, pp. 1177-89, 2012.
- [268] M. Sculpher, «Clinical trials provide essential evidence, but rarely offer a vehicle for cost-effectiveness analysis,» *Value Health*, vol. 18, n. 2, pp. 141-2, 2015.
- [269] G. R. Barton, A. H. Briggs e E. A. Fenwick, «Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI),» *Value Health*, vol. 11, n. 5, pp. 886-97, 2008.
- [270] M. E. Boysen, U. K. Zätterström e J. F. Evensen, «Self-reported symptoms to monitor recurrent head and neck cancer-analysis of 1,678 cases,» *Anticancer Res*, vol. 36, n. 6, pp. 2849-54, 2016.
- [271] S. T. Massa, N. Osazuwa-Peters, K. M. Christopher, L. D. Arnold, M. Schootman, R. J. Walker e et al., «Competing causes of death in the head and neck cancer population,» *Oral Oncol*, vol. 65, pp. 8-15, 2017.
- [272] A. V. Priante, E. C. Castilho e L. P. Kowalski, «Second primary tumors in patients with head and neck cancer,» *Curr Oncol Rep*, vol. 13, n. 2, pp. 132-7, 2011.
- [273] D. E. Wood, E. Kazerooni, S. L. Baum, M. T. Dransfield, G. A. Eapen, D. S. Ettinger e et al., «Lung cancer screening, version 1.2015: featured updates to the NCCN guidelines,» *J Natl Compr Canc Netw*, vol. 13, n. 1, pp. 23-34, 2015.
- [274] A. S. Ho, Z. S. Zumsteg, A. Meyer, N. Riaz, R. Rahmati, D. H. Kraus e et al., «Impact of flap reconstruction on radiotoxicity after salvage surgery and reirradiation for recurrent head and neck cancer,» *Ann Surg Oncol*, vol. 23, n. Suppl 3, pp. 850-57, 2016.
- [275] F. Janot, D. de Raucourt, E. Benhamou, C. Ferron, G. Dolivet, R. J. Bensadoun e et al., «Randomized trial of postoperative reirradiation combined with chemotherapy after salvage surgery compared with salvage surgery alone in head and neck carcinoma,» *J Clin Oncol*, vol. 26, n. 34, pp. 5518-23, 2008.

- [276] V. Paleri e C. G. Kelly, «Re-irradiation with concurrent chemotherapy in recurrent head and neck cancer: a decision analysis model based on a systematic review,» *Clin Otolaryngol*, vol. 33, n. 4, pp. 331-7, 2008.
- [277] J. Cacicedo, A. Navarro, F. Alongi, A. Gómez de Iturriaga, O. Del Hoyo, E. Boveda e et al., «The role of re-irradiation of secondary and recurrent head and neck carcinomas. Is it a potentially curative treatment? A practical approach,» *Cancer Treat Rev*, vol. 40, n. 1, pp. 178-89, 2014.
- [278] S. F. Huntington, J. Svoboda e J. A. Doshi, «Cost-effectiveness analysis of routine surveillance imaging of patients with diffuse large B-cell lymphoma in first remission,» *J Clin Oncol*, vol. 33, n. 13, pp. 1467-74, 2015.
- [279] Ministero della Salute, «DECRETO 18 ottobre 2012. Remunerazione delle prestazioni di assistenza ospedaliera per acuti, assistenza ospedaliera di riabilitazione e di lungodegenza post acuzie e di assistenza specialistica ambulatoriale,» 28 January 2013. [Online]. Available: <http://www.gazzettaufficiale.it/eli/gu/2013/01/28/23/so/8/sg/pdf>.
- [280] A. Molassiotis, R. Emsley, D. Ashcroft, A. Caress, J. Ellis, R. Wagland e et al., «Applying Best-Worst scaling methodology to establish delivery preferences of a symptom supportive care intervention in patients with lung cancer,» *Lung Cancer*, vol. 77, n. 1, pp. 199-204, 2012.
- [281] R. L. Hicks, «A comparison of stated and revealed preference methods for fisheries,» in *American Agricultural Economics Association. Annual Meeting, July 28-31, 2002.*, Long Beach (California), 2002.
- [282] T. L. Mark e J. Swait, «Using stated preference and revealed preference modeling to evaluate prescribing decisions,» *Health Econ*, vol. 13, n. 6, pp. 563-73, 2004.
- [283] F. Reed Johnson, E. Lancsar, D. Marshall, V. Kilambi, A. Mühlbacher, D. A. Regier e et al., «Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force,» *Value Health*, vol. 16, n. 1, pp. 3-13, 2013.
- [284] M. Tinelli, M. Ryan e C. Bond, «What, who and when? Incorporating a discrete choice experiment into an economic evaluation,» *Health Econ Rev*, vol. 6, n. 1, pp. 1-9, 2016.
- [285] J. Marti, «A best-worst scaling survey of adolescents' level of concern for health and non-health consequences of smoking,» *Soc Sci Med*, vol. 75, n. 1, pp. 87-97, 2012.
- [286] K. L. Cheung, B. F. Wijnen, I. L. Hollin, E. M. Janssen, J. F. Bridges, S. M. Evers e et al., «Using best-worst scaling to investigate preferences in health care,» *Pharmacoeconomics*, vol. 34, n. 12, pp. 1195-209, 2016.
- [287] A. Flynn e J. J. Louviere, «Determining the appropriate response to evidence of public concern: the case of food safety,» *J Public Policy Mark*, vol. 11, pp. 12-25, 1992.
- [288] T. N. Flynn, J. J. Louviere, T. J. Peters e J. Coast, «Best-worst scaling: What it can do for health care research and how to do it,» *J Health Econ*, vol. 26, n. 1, pp. 171-89, 2007.

- [289] T. M. Benning, M. L. Kimman, C. D. Dirksen, L. J. Boersma e B. G. Dellaert, «Combining individual-level discrete choice experiment estimates and costs to inform health care management decisions about customized care: the case of follow-up strategies after breast cancer treatment,» *Value Health*, vol. 15, n. 5, pp. 680-9, 2012.
- [290] T. Bessen, G. Chen, J. Street, J. Elliott, J. Karnon, D. Keefe e et al., «What sort of follow-up services would Australian breast cancer survivors prefer if we could no longer offer long-term specialist-based care? A discrete choice experiment,» *Br J Cancer*, vol. 110, n. 4, pp. 859-67, 2014.
- [291] M. L. Kimman, B. G. Dellaert, L. J. Boersma, P. Lambin e C. D. Dirksen, «Follow-up after treatment for breast cancer: one strategy fits all? An investigation of patient preferences using a discrete choice experiment,» *Acta Oncol*, vol. 49, n. 3, pp. 328-37, 2010.
- [292] P. Murchie, P. F. Norwood, M. Pietrucin-Materek, T. Porteous, P. C. Hannaford e M. Ryan, «Determining cancer survivors' preferences to inform new models of follow-up care,» *Br J Cancer*, vol. 115, n. 12, pp. 1495-503, 2016.
- [293] S. Damery, M. Biswas, L. Billingham, P. Barton, H. Al-Janabi e R. Grimer, «Patient preferences for clinical follow-up after primary treatment for soft tissue sarcoma: a cross-sectional survey and discrete choice experiment,» *Eur J Surg Oncol*, vol. 40, n. 12, pp. 1655-61, 2014.
- [294] G. Kocur, T. Adler, W. Hyman e B. Aunet, «Guide to forecasting travel demand with direct utility assessment,» 1981.
- [295] W. J. Ungar, A. Hadioonzadeh, M. Najafzadeh, N. W. Tsao, S. Dell e L. D. Lynd, «Quantifying preferences for asthma control in parents and adolescents using best-worst scaling,» *Respir Med*, vol. 108, n. 6, pp. 842-51, 2014.
- [296] D. Potoglou, P. Burge, T. Flynn, A. Netten, J. Malley, J. Forder e et al., «Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data,» *Soc Sci Med*, vol. 72, n. 10, pp. 1717-27, 2011.
- [297] A. Ejaz, G. Spolverato, J. F. Bridges, N. Amini, Y. Kim e T. M. Pawlik, «Choosing a cancer surgeon: analyzing factors in patient decision making using a best-worst scaling methodology,» *Ann Surg Oncol*, vol. 21, n. 12, pp. 3732-8, 2014.
- [298] A. Duhamel, M. C. Nuttens, P. Devos, M. Picavet e R. Beuscart, «A preprocessing method for improving data mining techniques. Application to a large medical diabetes database,» *Stud Health Technol Inform*, vol. 95, pp. 269-74, 2003.
- [299] T. N. Flynn, J. J. Louviere, T. J. Peters e J. Coast, «Estimating preferences for a dermatology consultation using Best-Worst Scaling: comparison of various methods of analysis,» *BMC Med Res Methodol*, vol. 8, pp. 1-12, 2008.
- [300] F. Severin, J. Schmidtke, A. Mühlbacher e W. H. Rogowski, «Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments,» *Eur J Hum Genet*, vol. 21, n. 11, pp. 1202-8, 2013.

- [301] M. R. Franco, K. Howard, C. Sherrington, P. H. Ferreira, J. Rose, J. L. Gomes e et al., «Eliciting older people's preferences for exercise programs: a best-worst scaling choice experiment,» *J Physiother*, vol. 61, n. 1, pp. 34-41, 2015.
- [302] J. Ratcliffe, E. Huynh, G. Chen, K. Stevens, J. Swait, J. Brazier e et al., «Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm,» *Soc Sci Med*, vol. 157, pp. 48-59, 2016.
- [303] A. Trindade, P. Kothari, Z. Andreou, R. J. Hewitt e P. O'Flynn, «Follow-up in head and neck cancer: patients' perspective,» *Int J Health Care Qual Assur*, vol. 25, n. 2, pp. 145-9, 2012.
- [304] A. C. Mühlbacher, A. Kaczynski, P. Zweifel e F. R. Johnson, «Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview,» *Health Econ Rev*, vol. 6, n. 1, pp. 1-14, 2016.
- [305] E. McIntosh, «Using discrete choice experiments within a cost-benefit analysis framework: some considerations,» *Pharmacoeconomics*, vol. 24, n. 9, pp. 855-68, 2006.
- [306] S. Petrou e E. McIntosh, «Women's preferences for attributes of first-trimester miscarriage management: a stated preference discrete-choice experiment,» *Value Health*, vol. 12, n. 4, pp. 551-9, 2009.
- [307] M. van der Pol, A. Shiell, F. Au, D. Jonhston e S. Tough, «Eliciting individual preferences for health care: a case study of perinatal care,» *Health Expect*, vol. 13, n. 1, pp. 4-12, 2010.
- [308] R. A. Lewis, R. D. Neal, M. Hendry, B. France, N. H. Williams, D. Russell e et al., «Patients' and healthcare professionals' views of cancer follow-up: systematic review,» *Br J Gen Pract*, vol. 59, n. 564, pp. e248-59, 2009.
- [309] G. Frew, A. Smith, B. Zutshi, N. Young, A. Aggarwal, P. Jones e et al., «Results of a quantitative survey to explore both perceptions of the purposes of follow-up and preferences for methods of follow-up delivery among service users, primary care practitioners and specialist clinicians after cancer treatment,» *Clin Oncol (R Coll Radiol)*, vol. 22, n. 10, pp. 874-84, 2010.
- [310] M. Meregaglia e J. Cairns, «Economic evaluations of follow-up strategies for cancer survivors: a systematic review and quality appraisal of the literature,» *Expert Rev Pharmacoecon Outcomes Res*, vol. 15, n. 6, pp. 913-29, 2015.
- [311] M. Meregaglia e J. Cairns, «A systematic literature review of health state utility values in head and neck cancer,» *Health Qual Life Outcomes*, vol. 15, n. 1, pp. 1-13, 2017.
- [312] M. Meregaglia, J. Cairns, L. Licitra e P. Bossi, «The use of intensive radiological assessments in routine surveillance after treatment for head and neck cancer: An economic evaluation,» *Eur J Cancer*, vol. 93, pp. 89-98, 2018.

- [313] M. Meregaglia, J. Cairns, S. Alfieri, F. Favales, D. Mazzitelli, E. Orlandi e et al., «Eliciting preferences for clinical follow-up in patients with head and neck cancer using best-worst scaling,» *Value Health*, vol. 20, n. 6, pp. 799-808, 2017.
- [314] G. P. Guy, D. U. Ekwueme, K. R. Yabroff, E. C. Dowling, C. Li, J. L. Rodriguez e et al., «Economic burden of cancer survivorship among adults in the United States,» *J Clin Oncol*, vol. 31, n. 30, pp. 3749-57, 2013.

APPENDICES

APPENDIX TO CHAPTER 2

Table A2.1 PRISMA 2009 checklist.



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 43
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Not available
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Pages 43-44
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 44
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Not available
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Pages 45-46; Table 2.1
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Pages 45-46
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Not available
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Pages 45-46
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 46
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 46
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Not applicable
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Not applicable
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Not applicable

Table A2.1 (cont.) PRISMA 2009 checklist.



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Not applicable
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. <i>Additional analysis: study quality assessment.</i>	Page 46
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Page 47; Figure 2.1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 2.2; Pages 47-62
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Not applicable
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 2.2; Pages 62-65
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Not applicable
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Not applicable
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Tables 2.3-2.4; Pages 65-68
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Pages 73-75
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Pages 76-77
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Pages 78-79
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Not applicable

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

APPENDIX TO CHAPTER 3

Table A3.1 Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
Aro (2016) [191]	15D	Patients receiving treatment (i.e. surgery, CRT, or combined modality treatment) and followed-up for at least 12 months	Baseline	214	0.872							
			3 months	All	198	0.839	0.114					
				no PEG	109	0.862						
				PEG	88	0.810						
			6 months	202	0.857							
12 months	214	0.852										
Chan (2014) [195]	Mapping	Patients after treatment for HNC	Estimation sample	Actual	89	0.821		0.03				
				Predicted		0.821		0.02				
			Validation sample	Actual	48	0.801		0.02				
				Predicted		0.791		0.01				
Conway (2012) [173]	SG	Oropharyngeal cancer stages II-III treated with ND and CRT/RT and/or surgery		99	0.58			0.53-0.63	0.65		0.45-0.75	
de Almeida (2014) [179]	VAS	Scenarios describing treatment modalities for oropharyngeal cancer	TORS	Healthy subjects	50	0.67			0.61-0.73			
				Experts	9	0.82			0.75-0.89			
			TORS + adjuvant XRT	Healthy subjects	50	0.59			0.62-0.64			
				Experts	9	0.60			0.48-0.72			
			TORS + adjuvant CRT	Healthy subjects	50	0.53			0.47-0.58			
				Experts	9	0.45			0.33-0.57			
			XRT	Healthy subjects	50	0.54			0.49-0.60			
				Experts	9	0.59			0.48-0.70			
			CRT	Healthy subjects	50	0.48			0.43-0.54			
				Experts	9	0.42			0.29-0.54			

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR			
de Almeida (2014) [179] (cont.)	SG	Scenarios describing treatment modalities for oropharyngeal cancer	TORS	Healthy subjects	50	0.95			0.94-0.97					
				Experts	9	0.99			0.97-1.00					
			TORS + adjuvant XRT	Healthy subjects	50	0.89			0.85-0.93					
				Experts	9	0.97			0.94-1.00					
			TORS + adjuvant CRT	Healthy subjects	50	0.89			0.85-0.93					
				Experts	9	0.94			0.91-0.97					
			XRT	Healthy subjects	50	0.91			0.87-0.94					
				Experts	9	0.97			0.94-1.00					
			CRT	Healthy subjects	50	0.88			0.83-0.92					
				Experts	9	0.93			0.88-0.97					
			VAS	Scenarios describing treatment-related complications	Temporary tracheostomy	Healthy subjects	50	0.61			0.56-0.66			
						Experts	9	0.53			0.44-0.62			
	Permanent tracheostomy	Healthy subjects			50	0.44			0.38-0.52					
		Experts			9	-			-					
	Temporary gastrostomy	Healthy subjects			50	0.54			0.50-0.59					
		Experts			9	0.46			0.29-0.62					
	Permanent gastrostomy	Healthy subjects			50	0.36			0.29-0.43					
		Experts			9	-			-					
	Pharyngocutaneous fistula	Healthy subjects			50	0.53			0.47-0.60					
		Experts			9	0.46			0.36-0.57					
	Febrile neutropenia	Healthy subjects			50	0.70			0.65-0.75					
		Experts			9	0.77			0.63-0.91					
	Oesophageal stenosis	Healthy subjects	50	0.40			0.35-0.46							
		Experts	9	0.38			0.22-0.53							
Osteoradionecrosis	Healthy subjects	50	0.41			0.35-0.47								
	Experts	9	0.44			0.33-0.55								

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR
de Almeida (2014) [179] (cont.)	SG	Scenarios describing treatment-related complications	Temporary tracheostomy	Healthy subjects	50	0.94		0.92-0.97			
				Experts	9	0.98		0.96-0.99			
			Permanent tracheostomy	Healthy subjects	50	0.85		0.80-0.91			
				Experts	9	-		-			
			Temporary gastrostomy	Healthy subjects	50	0.89		0.85-0.94			
				Experts	9	0.98		0.96-0.99			
			Permanent gastrostomy	Healthy subjects	50	0.81		0.74-0.88			
				Experts	9	-		-			
			Pharyngocutaneous fistula	Healthy subjects	50	0.89		0.85-0.94			
				Experts	9	0.96		0.92-0.99			
			Febrile neutropenia	Healthy subjects	50	0.96		0.94-0.98			
				Experts	9	0.99		0.98-1.00			
			Oesophageal stenosis	Healthy subjects	50	0.85		0.80-0.90			
				Experts	9	0.96		0.94-0.98			
			Osteoradionecrosis	Healthy subjects	50	0.85		0.81-0.90			
				Experts	9	0.96		0.93-0.99			
	VAS	Scenarios describing remission and recurrence	Remission (after TORS)	Healthy subjects	50	0.80		0.76-0.85			
				Experts	9	0.87		0.81-0.94			
			Remission (after TORS/adjuvant XRT or after XRT)	Healthy subjects	50	0.75		0.70-0.79			
				Experts	9	0.80		0.75-0.85			
			Remission (after TORS/adjuvant CRT or after CRT)	Healthy subjects	50	0.72		0.67-0.77			
				Experts	9	0.68		0.55-0.80			
			Local recurrence (requiring surgery)	Healthy subjects	50	0.39		0.33-0.45			
				Experts	9	0.45		0.29-0.62			
			Local recurrence (requiring XRT)	Healthy subjects	50	0.51		0.45-0.56			
				Experts	9	0.41		0.27-0.55			
			Regional recurrence (ND)	Healthy subjects	50	0.68		0.62-0.74			
				Experts	9	0.63		0.48-0.79			
Distant recurrence (CT)	Healthy subjects	50	0.20		0.16-0.24						
	Experts	9	0.18		0.12-0.24						
Terminal/palliative state (CT)	Healthy subjects	50	0.14		0.10-0.18						
	Experts	9	0.08		0.05-0.11						

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
de Almeida (2014) [179] (cont.)	SG	Scenarios describing remission and recurrence	Remission (after TORS)	Healthy subjects	50	0.96			0.94-0.98			
				Experts	9	0.99			0.98-1.00			
			Remission (after TORS/adjuvant XRT or XRT)	Healthy subjects	50	0.95			0.93-0.98			
				Experts	9	0.98			0.96-1.00			
			Remission (after TORS/adjuvant CRT or CRT)	Healthy subjects	50	0.95			0.92-0.98			
				Experts	9	0.97			0.94-0.99			
			Local recurrence (surgery)	Healthy subjects	50	0.82			0.77-0.87			
				Experts	9	0.92			0.87-0.97			
			Local recurrence (XRT)	Healthy subjects	50	0.88			0.84-0.91			
				Experts	9	0.91			0.87-0.95			
			Regional recurrence (ND)	Healthy subjects	50	0.94			0.91-0.97			
				Experts	9	0.97			0.94-0.99			
			Distant recurrence (CT)	Healthy subjects	50	0.57			0.50-0.64			
				Experts	9	0.43			0.22-0.64			
Terminal/palliative state (CT)	Healthy subjects	50	0.42			0.34-0.50						
	Experts	9	0.31			0.11-0.51						
del Barco Morillo (2016) [182]	EQ-5D	Palliative CT for recurrent or metastatic HNC (untreatable by surgery or re-irradiation)	Across all visits (every 8 weeks)	40					0.7	0.6-0.8		
Govers (2016) [189]	EQ-5D (Dutch tariff)	Early stage (I-II) oral cavity cancer patient undergoing different diagnostic and treatment interventions	WW	All patients	26	0.804		0.04				
				Group 1*	21	0.849		0.05				
				Group 2**	20	0.826		0.05				
			SLNB	All patients	19	0.863		0.05				
				Group 1	18	0.859		0.05				
			SOHND	Group 2	18	0.858		0.05				
				All patients	104	0.834		0.02				
				Group 1	86	0.841		0.02				
			MRND	Group 2	53	0.849		0.03				
				All patients	25	0.794		0.04				
Group 1	20	0.800			0.05							
			Group 2		-		-					

* Group 1: patients without previous mucosal malignancies, local recurrences, or second primary tumours in the HNC region. **Group 2: patients without previous mucosal malignancies, local recurrences, or second primary tumours and without adjuvant (chemo)radiotherapy in the HNC region.

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
Hamilton (2016) [174]	TTO	Four vignettes describing the treatment process and outcome for advanced laryngeal cancer	CRT, optimal outcome	All participants	114	0.64						
				Group 1*	71	0.70						
				Group 2**	43	0.53						
			CRT, outcome with complications	All participants	114	0.31						
				Group 1	71	0.37						
				Group 2	43	0.22						
			TL, optimal outcome	All participants	114	0.57						
				Group 1	71	0.55						
				Group 2	43	0.59						
			TL, outcome with complications	All participants	114	0.33						
				Group 1	71	0.34						
				Group 2	43	0.22						
Higgins (2011) [192]	HUI3 (health state A); adjustments of A score (other health states)	Patients with complete response to treatment (XRT/CO2) and no evidence of active disease (health state A)	A: alive with voice box entirely intact	30	0.8718							
			B: alive with part of the box intact	-	0.706							
			C: dead of disease	-	0							
			D: alive with recurrent/active disease	-	0.307							
			E: alive without voice box/TL	-	0.366							
Hollenbeak (2001) [175]	TTO	Surgical patients	Modified ND	8	0.925	0.23						
			Radiation plus modified ND	8	0.913	0.18						
			Radiation	8	0.875	0.44						
			Radical ND	8	0.763	1.03						
			Radiation plus radical ND	8	0.675	1.3						
Kent (2015) [193]	SF-6D/VR-6D	Patients after a diagnosis of oral cavity or pharyngeal cancer		580	0.69			0.68-0.70				

* Group 1: participants ranking CRT first in a previous exercise. ** Group 2: participants ranking TL first in a previous exercise.

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
Llewellyn-Thomas (1993) [180]	TTO	Larynx cancer patients eligible for a standard four-week RT regimen	Mild*	Mild**	Time 1	24	0.721	26.2				
					Time 2		0.735	23.5				
				Moderate**	Time 1	36	0.750	19.9				
					Time 2		0.757	19.0				
				Severe**	Time 1	6	0.750	24.1				
					Time 2		0.866	7.5				
			Moderate*	Mild**	Time 1	24	0.629	26.9				
					Time 2		0.571	26.4				
				Moderate**	Time 1	36	0.644	22.9				
					Time 2		0.667	21.8				
				Severe**	Time 1	6	0.700	27.0				
					Time 2		0.758	15.0				
	Severe*	Mild**	Time 1	24	0.352	28.3						
			Time 2		0.429	29.2						
		Moderate**	Time 1	36	0.344	25.3						
			Time 2		0.381	26.7						
		Severe**	Time 1	6	0.233	22.7						
			Time 2		0.408	39.0						
	VAS	Mild*	Larynx cancer patients eligible for a standard four-week RT regimen	Mild**	Time 1	24	0.826	9.8				
					Time 2		0.793	13.3				
				Moderate**	Time 1	36	0.793	16.0				
					Time 2		0.744	13.7				
				Severe**	Time 1	6	0.775	13.9				
					Time 2		0.783	10.0				
Moderate*		Mild**		Time 1	24	0.598	16.6					
				Time 2		0.623	17.8					
		Moderate**		Time 1	36	0.559	20.9					
				Time 2		0.532	17.8					
		Severe**		Time 1	6	0.647	17.4					
				Time 2		0.615	14.0					

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description		Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR		
Llewellyn-Thomas (1993) [180] (cont.)	VAS (cont.)	Larynx cancer patients eligible for a standard four-week RT regimen	Severe*	Mild**	Time 1	24	0.266	24.3						
					Time 2		0.276	25.3						
					Moderate**	Time 1	36	0.226	21.4					
						Time 2		0.209	19.2					
					Severe**	Time 1	6	0.128	17.6					
						Time 2		0.292	31.8					
Loimu (2015) [194]	15D	Newly diagnosed patients scheduled for receiving CRT		Baseline	64	0.886	0.10							
				3 months	54	0.829	0.12							
				6 months	61	0.860	0.12							
				12 months	64	0.862	0.14							
Marcellusi (2015) [171]	TTO	Patients with a confirmed diagnosis of HNC and time from (medical or surgical) treatment no longer than 20 months		All patients	79	0.69	0.30		0.62-0.75					
				Males	62	0.70	0.32		0.62-0.78					
				Females	17	0.64	0.21		0.54-0.74					
	EQ-5D			All patients	79	0.80	0.20							
				Males	62	0.80	0.20							
				Females	17	0.70	0.20							
Noel (2015) [172]	SG	Patients with a minimum of three months after completion of treatment (surgery or RT) and no evidence of recurrent disease		All patients	100	0.91	0.17				0.2-1.0			
				Primary surgery	54	0.93	0.17							
				Salvage surgery	5	0.98	0.04							
				Chemotherapy	13	0.92	0.10							
				No chemotherapy	87	0.91	0.18							
				Stage T1 or T2	47	0.95	0.13							
				Stage T3 or T4	20	0.87	0.22							
				Tracheotomy and/or feeding tube	6	0.99	0.02							
				No tracheotomy and/or feeding tube	94	0.91	0.17							

Llewellyn-Thomas (1993): Mild/Moderate/Severe*: health state scenarios describing increasing levels of three radiation-induced disorders (i.e. mouth/throat pain, fatigue, inability to converse). Mild/Moderate/Severe**: end-of-therapy groups as self-determined by the patients. Time 1/Time 2: outset/end of therapy. HSUVs are converted on a 0-1 scale (from a 0-100 one).

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
Noel (2015) [172] (cont.)	TTO	Patients with a minimum of three months after completion of treatment (surgery or RT) and no evidence of recurrent disease	All patients	100	0.94	0.14				0.3-1.0		
			Primary surgery	54	0.95	0.13						
			Salvage surgery	5	0.98	0.04						
			Chemotherapy	13	0.99	0.03						
			No chemotherapy	87	0.94	0.14						
			Stage T1 or T2	47	0.96	0.09						
			Stage T3 or T4	20	0.88	0.21						
			Tracheotomy and/or feeding tube	6	0.91	0.12						
	No tracheotomy and/or feeding tube		94	0.95	0.14							
	VAS		All patients	100	0.76	0.19					0.2-1.0	
			Primary surgery	54	0.76	0.20						
			Salvage surgery	5	0.48	0.13						
			Chemotherapy	13	0.66	0.19						
			No chemotherapy	87	0.77	0.18						
			Stage T1 or T2	47	0.77	0.18						
			Stage T3 or T4	20	0.70	0.20						
			Tracheotomy and/or feeding tube	6	0.69	0.23						
	No tracheotomy and/or feeding tube		94	0.76	0.19							
	EQ-5D		All patients	100	0.82	0.18						-0.07; 1.0
			Primary surgery	54	0.83	0.19						
			Salvage surgery	5	0.62	0.17						
			Chemotherapy	13	0.76	0.17						
			No chemotherapy	87	0.83	0.18						
			Stage T1 or T2	47	0.83	0.18						
Stage T3 or T4		20	0.83	0.09								
Tracheotomy and/or feeding tube		6	0.78	0.14								
No tracheotomy and/or feeding tube	94	0.82	0.18									

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
Noel (2015) [172] (cont.)	HUI3	Patients with a minimum of three months after completion of treatment (surgery or RT) and no evidence of recurrent disease	All patients	100	0.75	0.25					-0.06; 1.0	
			Primary surgery	54	0.78	0.22						
			Salvage surgery	5	0.37	0.29						
			Chemotherapy	13	0.57	0.38						
			No chemotherapy	87	0.78	0.21						
			Stage T1 or T2	47	0.80	0.21						
			Stage T3 or T4	20	0.74	0.21						
			Tracheotomy and/or feeding tube	6	0.73	0.25						
			No tracheotomy and/or feeding tube	94	0.75	0.29						
			Larynx	17	0.59							
			Oropharynx	14	0.76							
Oral cavity	67	0.78										
Outtassi (2016) [183]	EQ-5D	Patients with a confirmed diagnosis of HNC		120	0.49	0.35						
Parrilla (2015) [184]	EQ-5D	Patients laryngectomized with a stable pulmonary situation with a minimum of 3 months after treatment	Baseline (no HME)	30	0.84	0.14				0.44-1.00		
			Week 2 (HME)	30	0.90	0.10				0.67-1.00		
			Week 6 (HME)	30	0.93	0.09				0.68-1.00		
			Week 12 (HME)	30	0.96	0.10				0.66-1.00		
Parthan (2009) [196]	Mapping	Patients with locally advanced inoperable HNC	Stable		0.70					0.63-0.77*		
			Progressive		0.67					0.60-0.74		
			Response		0.79					0.71-0.87		
Pickard (2016) [187]	EQ-5D (US tariff)	Patients with advanced HNC after at least two cycles of CT		50	0.76	0.15						

* Uncertainty range: 2.5 percentile and 97.5 percentile.

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
Pottel (2015) [190]	EQ-5D (Belgian tariff)	Patients aged ≥ 65 years, eligible for curative primary or adjuvant RT	Baseline (before treatment)	All	81				0.66		0.55-0.76	
				Fit*	-				0.76		0.66-0.76	
				Vulnerable	-				0.63		0.29-0.73	
			Week 4 (mid-therapy)	All	81					0.42		0.26-0.73
				Fit	-					0.66		0.39-0.76
				Vulnerable	-					0.39		0.21-0.67
			2 months (end of treatment)	All	81					0.66		0.29-0.76
				Fit	-					0.74		0.66-0.76
				Vulnerable	-					0.58		0.23-0.73
			5 months (follow-up)	All	81					0.66		0.27-0.76
				Fit	-					0.76		0.66-1.00
				Vulnerable	-					0.66		0.19-0.76
			12 months	All	81					0.64		0.00-0.76
				Fit	-					0.76		0.64-1.00
				Vulnerable	-					0.57		0.00-0.74
			24 months	All	81					0.29		0.00-0.76
				Fit	-					0.76		0.32-1.00
				Vulnerable	-					0.00		0.00-0.66
			36 months	All	81					0.00		0.00-0.67
				Fit	-					0.66		0.00-1.00
				Vulnerable	-					0.00		0.00-0.58

* Fit/vulnerable patients: classification based on geriatric-8 (G-8) assessment at baseline.

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR
Ramaekers (2011) [186]	EQ-5D (UK tariff)	Patients with a follow-up of at least 6 months after curative RT without evidence of recurrent disease	All patients	396	0.850	0.18			-		-
			X0-D0	84	0.909	0.161			1.000		0.186
			X0-D1	18	0.841	0.144			0.796		0.275
			X1-D0	92	0.898	0.138			1.000		0.204
			X1-D1	68	0.829	0.175			0.814		0.275
			X1-D2	14	0.803	0.136			0.796		0.133
			X2-D0	15	0.846	0.177			0.850		0.275
			X2-D1	31	0.817	0.187			0.812		0.309
			X2-D2	40	0.763	0.213			0.778		0.311
			X2-D3(+)	16	0.758	0.234			0.796		0.363
Ringash (2000) [177]	TTO	Irradiated laryngeal cancer patients who completed treatment at least 6 months before	All	112	0.914	0.156				0.25; 1.0	
			Group 1*	84	0.878	0.174				0.25; 1.0	
Rogers (2006) [185]	EQ-5D (UK tariff)	Patients without evidence of disease after primary surgery for oral/oropharyngeal cancer		224	0.75		0.02			-0.18; 1.0	

Group 1: excluding patients who claimed they had or did not want perfect health.

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR
Szabo (2012) [155]	SG	Eight vignettes representing disease's characteristics and ten describing treatment-related toxicities	Loco-regional (larynx)	101	0.62		0.02	0.57-0.67	0.65	0.00-0.98	0.50-0.80
			Loco-regional (not larynx)	101	0.61		0.02	0.56-0.66	0.63	0.03-0.98	0.50-0.78
			Recurrent (not larynx)	101	0.57		0.02	0.52-0.62	0.58	0.03-0.98	0.50-0.78
			Recurrent (larynx)	101	0.56		0.02	0.51-0.61	0.55	0.00-0.98	0.45-0.73
			Metastatic (not larynx)	101	0.52		0.02	0.47-0.57	0.50	0.00-0.98	0.38-0.68
			Metastatic (larynx)	101	0.50		0.02	0.45-0.55	0.50	0.00-0.98	0.35-0.65
			Anaemia grade III/IV	49	0.47		0.03	0.40-0.54	0.50	0.00-0.98	0.30-0.65
			Haematological grade III/IV	49	0.46		0.04	0.39-0.53	0.50	0.00-0.98	0.33-0.65
			Skin reactions grade I/II	52	0.45		0.04	0.37-0.52	0.50	0.03-0.98	0.24-0.58
			Peripheral neuropathy grade III/IV	49	0.44		0.04	0.36-0.51	0.45	0.00-0.98	0.20-0.60
			Treatment cessation due to grade III/IV toxicity	52	0.44		0.03	0.36-0.51	0.50	0.03-0.98	0.23-0.55
			Nausea/vomiting grade III/IV	49	0.43		0.04	0.35-0.50	0.45	0.00-0.98	0.25-0.60
			Mucositis/stomatitis grade III/IV	49	0.43		0.04	0.35-0.50	0.45	0.00-0.98	0.23-0.58
			Anorexia/weight loss grade III/IV	52	0.40		0.04	0.33-0.47	0.48	0.00-0.98	0.18-0.50
			Skin reactions grade III/IV	52	0.37		0.04	0.30-0.44	0.43	0.00-0.98	0.14-0.50
			Post-progression (not larynx)	101	0.34		0.02	0.29-0.39	0.38	0.00-0.98	0.08-0.50
Post-progression (larynx)	101	0.34		0.02	0.29-0.39	0.38	0.00-0.98	0.08-0.50			
Hospitalization for grade III/IV toxicity	52	0.33		0.04	0.26-0.40	0.36	0.00-0.98	0.05-0.50			

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
Truong (2016) [188]	EQ-5D (US tariff)	Untreated stage III or IV cancer patients enrolled on a RCT comparing radiation-cisplatin without cetuximab (CIS) or with cetuximab (CET/CIS)	CIS	Baseline	366	0.78	0.18			0.82	0.17-1.00	0.77-0.84
				3 months	-	0.78	0.18			-	-	-
				12 months	-	0.84	0.17			-	-	-
			CET/CIS	Baseline	349	0.80	0.17			0.83	0.20-1.00	0.77-0.84
				3 months	-	0.77	0.15			-	-	-
				12 months	-	0.84	0.16			-	-	-
van der Donk (1995) [181]	TTO	Scenarios describing the health state of patients treated for T3 laryngeal cancer (<i>state scenarios</i>)	RT	Laryngeal cancer	10	0.70						
				FOM cancer	10	0.72						
				Healthy subjects	10	0.90						
				Clinical experts	9	0.81						
			Surgery	Laryngeal cancer	10	0.65						
				FOM cancer	10	0.64						
				Healthy subjects	10	0.77						
				Clinical experts	9	0.71						
	SG		RT	Laryngeal cancer	10	0.61						
				FOM cancer	10	0.83						
				Healthy subjects	10	0.84						
				Clinical experts	9	0.92						
			Surgery	Laryngeal cancer	10	0.62						
				FOM cancer	10	0.63						
				Healthy subjects	10	0.68						
				Clinical experts	9	0.84						

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR	
van der Donk (1995) [181] (cont.)	RS	Scenarios describing the health state of patients treated for T3 laryngeal cancer (<i>state scenarios</i>)	RT	Laryngeal cancer	10	0.66						
				FOM cancer	10	0.78						
				Healthy subjects	10	0.68						
				Clinical experts	9	0.78						
			Surgery	Laryngeal cancer	10	0.45						
				FOM cancer	10	0.50						
				Healthy subjects	10	0.47						
				Clinical experts	9	0.57						
	TTO	Scenarios describing the health state of patients treated for T3 laryngeal cancer including temporary and permanent side effects, life expectancy, tumour recurrence rates, and probability of treatment outcomes (<i>dynamic scenarios</i>)	RT	Laryngeal cancer	10	0.66						
				FOM cancer	10	0.62						
				Healthy subjects	10	0.73						
				Clinical experts	9	0.80						
			Surgery	Laryngeal cancer	10	0.62						
				FOM cancer	10	0.61						
				Healthy subjects	10	0.66						
				Clinical experts	9	0.73						
	SG	Scenarios describing the health state of patients treated for T3 laryngeal cancer including temporary and permanent side effects, life expectancy, tumour recurrence rates, and probability of treatment outcomes (<i>dynamic scenarios</i>)	RT	Laryngeal cancer	10	0.65						
				FOM cancer	10	0.70						
				Healthy subjects	10	0.83						
				Clinical experts	9	0.91						
Surgery			Laryngeal cancer	10	0.71							
			FOM cancer	10	0.66							
			Healthy subjects	10	0.76							
			Clinical experts	9	0.85							

Table A3.1 (cont.) Details of HSUVs reported by the studies (n=27).

Author (year)	Method	Patient's characteristics/ Health state description	Group/Time point	N	Mean	SD	SE	95% CI	Median	Range	IQR
van der Donk (1995) [181] (cont.)	RS (VAS)	Scenarios describing the health state of patients treated for T3 laryngeal cancer including temporary and permanent side effects, life expectancy, tumour recurrence rates, and probability of treatment outcomes (<i>dynamic scenarios</i>)	RT	Laryngeal cancer	10	0.60					
				FOM cancer	10	0.63					
				Healthy subjects	10	0.64					
				Clinical experts	9	0.69					
			Surgery	Clinical experts	9	0.55					
				Laryngeal cancer	10	0.49					
				FOM cancer	10	0.55					
				Healthy subjects	10	0.52					
				Clinical experts	9	0.55					
Weiss (1994) [178]	TTO	Stage N0 patients free of disease after different treatment options	Observation		3	1.0					
			Neck dissection		3	0.97					
			Radiotherapy		3	0.97					
			Salvage surgery (successful)		3	0.94					
Yong (2012) [199]	Mapping	Patients with early stage cancer	IMRT	Immediately after treatment		0.810					
				6 months		0.853					
				12 months		0.925					
			3DCRT	Immediately after treatment		0.810					
				6 months		0.853					
				12 months		0.868					

3DCRT: three dimensional conformal radiotherapy; CI: confidence interval; CO2: transoral CO2 laser excision; CRT: chemoradiotherapy; CT: chemotherapy; D: dysphagia; FOM: floor-of-the-mouth; HME: heat and moisture exchanger; HSUV: health state utility value; HUI3: Health Utility Index Mark 3; IMRT: intensity-modulated radiation therapy; IQR: interquartile range; MRND: modified radical neck dissection; NA: not available; ND: neck dissection; PEG: gastrostomy tube; RCT: randomized controlled trial; RS: rating scale; RT: radiotherapy; SE: standard error; SD: standard deviation; SG: standard gamble; SLNB: sentinel lymph node biopsy; SOHND: supraomohyoid neck dissection; TL: total laryngectomy; TORS: transoral robotic surgery; TTO: time trade-off; VAS: visual analogue scale; WW: watchful waiting; X: xerostomia; XRT: external radiation therapy.

APPENDIX TO CHAPTER 4

Table A4.1 MAPS checklist of included items.

Section/topic	Item number	Recommendation	Reported on page number/line number
Title and abstract			
Title	1	Identify the report as a study mapping between outcome measures. State the source measure(s) and generic, preference-based target measure(s) used in the study	Page 104
Abstract	2	Provide a structured abstract including, as applicable: objectives; methods, including data sources and their key characteristics, outcome measures used and estimation and validation strategies; results, including indicators of model performance; conclusions; and implications of key findings	Not available
Introduction			
Study rationale	3	Describe the rationale for the mapping study in the context of the broader evidence base	Pages 104-106
Study objective	4	Specify the research question with reference to the source and target measures used and the disease or population context of the study	Page 106
Methods			
Estimation sample	5	Describe how the estimation sample was identified, why it was selected, the methods of recruitment and data collection, and its location(s) or setting(s)	Page 108
External validation sample	6	If an external validation sample was used, the rationale for selection, the methods of recruitment and data collection, and its location(s) or setting(s) should be described	Not applicable
Source and target measures	7	Describe the source and target measures and the methods by which they were applied in the mapping study	Pages 107-108
Exploratory data analysis	8	Describe the methods used to assess the degree of conceptual overlap between the source and target measures	Page 109
Missing data	9	State how much data were missing and how missing data were handled in the sample(s) used for the analyses	Page 108
Modelling approaches	10	Describe and justify the statistical model(s) used to develop the mapping algorithm	Pages 109-111
Estimation of predicted scores or utilities	11	Describe how predicted scores or utilities are estimated for each model specification	Page 112
Validation methods	12	Describe and justify the methods used to validate the mapping algorithm	Not applicable
Measures of model performance	13	State and justify the measure(s) of model performance that determine the choice of the preferred model(s) and describe how these measures were estimated and applied	Page 112

Source: Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. PREFERRED REPORTING ITEMS FOR STUDIES MAPPING ONTO PREFERENCE-BASED OUTCOME MEASURES: THE MAPS STATEMENT. *Int J Technol Assess Health Care*. 2015; 31(4):230-5.

Table A4.1 (cont.) MAPS checklist of included items.

Section/topic	Item number	Recommendation	Reported on page number/line number
Results			
Final sample size(s)	14	State the size of the estimation sample and any validation sample(s) used in the analyses (including both number of individuals and number of observations)	Pages 112-115
Descriptive information	15	Describe the characteristics of individuals in the sample(s) (or refer back to previous publications giving such information). Provide summary scores for source and target measures, and summarize results of analyses used to assess overlap between the source and target measures	Pages 112-118; Tables 4.1-4.2-4.3-4.4; Figure 4.1
Model selection	16	State which model(s) is(are) preferred and justify why this(these) model(s) was(were) chosen	Pages 119-120
Model coefficients	17	Provide all model coefficients and standard errors for the selected model(s). Provide clear guidance on how a user can calculate utility scores based on the outputs of the selected model(s)	Tables 4.5-4.6-4.7-4.8
Uncertainty	18	Report information that enables users to estimate standard errors around mean utility predictions and individual-level variability	Not applicable
Model performance and face validity	19	Present results of model performance, such as measures of prediction accuracy and fit statistics for the selected model(s) in a table or in the text. Provide an assessment of face validity of the selected model(s)	Tables 4.5-4.6-4.7-4.8; Figures 4.2-4.3-4.4-4.5
Discussion			
Comparisons with previous studies	20	Report details of previously published studies developing mapping algorithms between the same source and target measures and describe differences between the algorithms, in terms of model performance, predictions, and coefficients, if applicable	Page 129
Study limitations	21	Outline the potential limitations of the mapping algorithm	Pages 130-132
Scope of applications	22	Outline the clinical and research settings in which the mapping algorithm could be used	Page 132
Other			
Additional information	23	Describe the source(s) of funding and non-monetary support for the study, and the role of the funder(s) in its design, conduct and report. Report any conflicts of interest.	Not applicable

Source: Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. PREFERRED REPORTING ITEMS FOR STUDIES MAPPING ONTO PREFERENCE-BASED OUTCOME MEASURES: THE MAPS STATEMENT. *Int J Technol Assess Health Care*. 2015; 31(4):230-5.

Tables A4.2 Linear mixed-effects and random-effects Tobit models using QLQ-C30.

Table A4.2 (A) Linear mixed-effects and random-effects Tobit models using QLQ-C30 (England).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.4537	0.1069	0.000	0.4050	0.0751	0.000	0.4492	0.1298	0.001	0.4422	0.0593	0.000
GH	0.0009	0.0005	0.068	0.0011	0.0004	0.012	0.0012	0.0005	0.024	0.0013	0.0005	0.014
PF	0.0034	0.0008	0.000	0.0034	0.0008	0.000	0.0035	0.0007	0.000	0.0034	0.0006	0.000
RF	0.0004	0.0005	0.330				0.0005	0.0005	0.328			
EF	0.0011	0.0004	0.011	0.0012	0.0005	0.022	0.0013	0.0006	0.022	0.0010	0.0005	0.041
CF	-0.0003	0.0007	0.717				-0.0003	0.0006	0.570			
SF	0.0002	0.0003	0.565				0.0002	0.0004	0.591			
FA	0.0008	0.0006	0.235				0.0007	0.0005	0.215			
NV	-0.0027	0.0013	0.035	-0.0027	0.0012	0.022	-0.0029	0.0010	0.005	-0.0024	0.0009	0.011
PA	-0.0007	0.0005	0.167				-0.0007	0.0005	0.118			
DY	0.0001	0.0004	0.855				0.0001	0.0004	0.728			
SL	0.0003	0.0003	0.302				0.0002	0.0004	0.493			
AP	0.0002	0.0004	0.546				0.0002	0.0004	0.541			
CO	0.0007	0.0004	0.080	0.0007	0.0002	0.006	0.0006	0.0004	0.107			
DI	-0.0018	0.0010	0.078				-0.0018	0.0007	0.009	-0.0018	0.0007	0.008
FI	-0.0010	0.0003	0.001	-0.0008	0.0003	0.002	-0.0011	0.0003	0.002	-0.0011	0.0003	0.001
Female	0.0143	0.0170	0.401				0.0218	0.0277	0.432			
Age	-0.0010	0.0009	0.261				-0.0015	0.0011	0.171			
/sigma_u							0.0687	0.0122	0.000	0.0719	0.0110	0.000
/sigma_e							0.0836	0.0062	0.000	0.0865	0.0059	0.000
Rho							0.4030	0.1062		0.4085	0.0907	
AIC	-411.79			-413.60			-230.34			-238.84		
BIC	-343.55			-382.90			-162.10			-208.13		
MAE	0.0695			0.0698			0.0702			0.0693		
RMSE	0.0985			0.1012			0.0978			0.1012		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error, SE: standard error.

Table A4.2 (B) Linear mixed-effects and random-effects Tobit models using QLQ-C30 (Netherlands).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.3416	0.1255	0.007	0.2859	0.0929	0.002	0.3390	0.1606	0.035	0.3283	0.0738	0.000
GH	0.0013	0.0006	0.030	0.0014	0.0005	0.005	0.0016	0.0006	0.016	0.0017	0.0007	0.010
PF	0.0038	0.0009	0.000	0.0037	0.0009	0.000	0.0040	0.0009	0.000	0.0037	0.0007	0.000
RF	0.0006	0.0005	0.275				0.0006	0.0006	0.293			
EF	0.0015	0.0006	0.009	0.0016	0.0007	0.022	0.0017	0.0007	0.013	0.0013	0.0006	0.032
CF	-0.0003	0.0008	0.680				-0.0004	0.0007	0.552			
SF	<0.0001	0.0004	0.942				0.0001	0.0005	0.864			
FA	0.0010	0.0007	0.183				0.0009	0.0007	0.187			
NV	-0.0031	0.0015	0.038	-0.0030	0.0013	0.028	-0.0032	0.0012	0.009	-0.0026	0.0012	0.023
PA	-0.0009	0.0005	0.119				-0.0009	0.0006	0.130			
DY	0.0002	0.0005	0.710				0.0003	0.0005	0.627			
SL	0.0004	0.0004	0.228				0.0004	0.0004	0.411			
AP	0.0003	0.0005	0.572				0.0003	0.0005	0.548			
CO	0.0008	0.0004	0.081	0.0008	0.0003	0.010	0.0007	0.0005	0.130			
DI	-0.0023	0.0012	0.062				-0.0022	0.0008	0.008	-0.0022	0.0008	0.007
FI	-0.0014	0.0004	0.000	-0.0011	0.0003	0.001	-0.0015	0.0004	0.001	-0.0014	0.0004	0.000
Female	0.0195	0.0221	0.377				0.0278	0.0351	0.429			
Age	-0.0013	0.0011	0.241				-0.0019	0.0014	0.165			
/sigma_u							0.0908	0.0152	0.000	0.0926	0.0138	0.000
/sigma_e							0.1005	0.0075	0.000	0.1048	0.0073	0.000
Rho							0.4494	0.1049		0.4387	0.0906	
AIC	-323.26			-323.78			-149.16			-157.87		
BIC	-255.02			-293.07			-80.93			-127.17		
MAE	0.0869			0.0883			0.0879			0.0862		
RMSE	0.1224			0.1256			0.1207			0.1243		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.2 (C) Linear mixed-effects and random-effects Tobit models using QLQ-C30 (Canada).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.5232	0.0871	0.000	0.5628	0.0555	0.000	0.5172	0.1171	0.000	0.5471	0.0510	0.000
GH	0.0009	0.0004	0.023	0.0009	0.0003	0.002	0.0012	0.0005	0.012	0.0012	0.0005	0.007
PF	0.0032	0.0007	0.000	0.0028	0.0007	0.000	0.0034	0.0006	0.000	0.0029	0.0005	0.000
RF	0.0005	0.0004	0.211				0.0005	0.0004	0.210			
EF	0.0006	0.0004	0.158				0.0008	0.0005	0.114			
CF	-0.0005	0.0006	0.406				-0.0005	0.0005	0.286			
SF	<0.0001	0.0002	0.852				0.0001	0.0004	0.827			
FA	0.0009	0.0005	0.111				0.0008	0.0005	0.094			
NV	-0.0021	0.0012	0.066				-0.0023	0.0009	0.009	-0.0019	0.0008	0.020
PA	-0.0009	0.0004	0.027	-0.0011	0.0004	0.003	-0.0009	0.0004	0.026	-0.0010	0.0004	0.008
DY	<0.0001	0.0003	0.995				0.0001	0.0004	0.829			
SL	0.0004	0.0002	0.146				0.0003	0.0003	0.313			
AP	0.0002	0.0003	0.553				0.0002	0.0003	0.515			
CO	0.0006	0.0003	0.055	0.0006	0.0002	0.005	0.0006	0.0003	0.082	0.0006	0.0003	0.032
DI	-0.0018	0.0009	0.045	-0.0020	0.0009	0.026	-0.0018	0.0006	0.002	-0.0019	0.0006	0.001
FI	-0.0011	0.0003	0.000	-0.0009	0.0002	0.000	-0.0012	0.0003	0.000	-0.0011	0.0003	0.000
Female	0.0130	0.0167	0.435				0.0191	0.0271	0.481			
Age	-0.0011	0.0009	0.220				-0.0015	0.0011	0.150			
/sigma_u							0.0755	0.0103	0.000	0.0787	0.0096	0.000
/sigma_e							0.0675	0.0049	0.000	0.0692	0.0048	0.000
Rho							0.5559	0.0873		0.5640	0.0762	
AIC	-470.50			-473.48			-284.78			-293.32		
BIC	-402.27			-442.77			-216.55			-259.21		
MAE	0.0643			0.0653			0.0651			0.0649		
RMSE	0.0958			0.0997			0.0944			0.0984		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.2 (D) Linear mixed-effects and random-effects Tobit models using QLQ-C30 (Uruguay).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.6878	0.0695	0.000	0.7288	0.0399	0.000	0.6846	0.0892	0.000	0.7172	0.0389	0.000
GH	0.0007	0.0003	0.041	0.0007	0.0002	0.003	0.0009	0.0004	0.014	0.0010	0.0003	0.006
PF	0.0020	0.0005	0.000	0.0018	0.0005	0.000	0.0022	0.0005	0.000	0.0020	0.0004	0.000
RF	0.0005	0.0004	0.171				0.0005	0.0003	0.095			
EF	0.0002	0.0003	0.605				0.0003	0.0004	0.417			
CF	-0.0003	0.0005	0.550				-0.0003	0.0004	0.439			
SF	0.0002	0.0002	0.208				0.0003	0.0003	0.368			
FA	0.0005	0.0004	0.174				0.0005	0.0004	0.217			
NV	-0.0016	0.0008	0.055				-0.0017	0.0007	0.010	-0.0014	0.0006	0.021
PA	-0.0006	0.0003	0.081	-0.0008	0.0003	0.013	-0.0006	0.0003	0.049	-0.0007	0.0003	0.017
DY	0.0002	0.0002	0.495				0.0002	0.0003	0.430			
SL	0.0001	0.0002	0.474				0.0001	0.0002	0.617			
AP	0.0002	0.0002	0.420				0.0002	0.0002	0.408			
CO	0.0005	0.0002	0.044	0.0005	0.0002	0.005	0.0004	0.0002	0.085	0.0005	0.0002	0.026
DI	-0.0011	0.0006	0.084	-0.0013	0.0006	0.030	-0.0011	0.0004	0.018	-0.0012	0.0004	0.005
FI	-0.0008	0.0002	0.000	-0.0007	0.0002	0.000	-0.0008	0.0002	0.000	-0.0008	0.0002	0.000
Female	0.0138	0.0124	0.269				0.0183	0.0200	0.359			
Age	-0.0007	0.0006	0.245				-0.0011	0.0008	0.182			
/sigma_u							0.0535	0.0081	0.000	0.0574	0.0075	0.000
/sigma_e							0.0537	0.0040	0.000	0.0544	0.0038	0.000
rho							0.4980	0.0969		0.5262	0.0825	
AIC	-586.48			-590.39			-381.67			-390.66		
BIC	-518.24			-559.69			-313.43			-356.54		
MAE	0.0477			0.0498			0.0477			0.0487		
RMSE	0.0704			0.0738			0.0697			0.0730		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.2 (E) Linear mixed-effects and random-effects Tobit models using QLQ-C30 (South Korea).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.4707	0.0961	0.000	0.5110	0.0466	0.000	0.4734	0.1187	0.000	0.4863	0.0483	0.000
GH	0.0010	0.0004	0.012	0.0012	0.0004	0.001	0.0012	0.0005	0.012	0.0015	0.0005	0.003
PF	0.0028	0.0007	0.000	0.0030	0.0006	0.000	0.0029	0.0007	0.000	0.0032	0.0005	0.000
RF	0.0006	0.0004	0.160				0.0006	0.0004	0.168			
EF	0.0006	0.0004	0.103				0.0008	0.0005	0.137			
CF	-0.0004	0.0005	0.407				-0.0005	0.0005	0.368			
SF	0.0001	0.0002	0.577				0.0002	0.0004	0.659			
FA	0.0007	0.0006	0.210				0.0007	0.0005	0.200			
NV	-0.0020	0.0011	0.061	-0.0018	0.0009	0.041	-0.0020	0.0010	0.032	-0.0019	0.0009	0.034
PA	-0.0006	0.0005	0.167				-0.0007	0.0004	0.138			
DY	0.0002	0.0003	0.653				0.0002	0.0004	0.602			
SL	0.0002	0.0002	0.391				0.0002	0.0003	0.596			
AP	0.0002	0.0003	0.442				0.0002	0.0003	0.510			
CO	0.0004	0.0003	0.195				0.0003	0.0003	0.346			
DI	-0.0010	0.0006	0.088				-0.0010	0.0006	0.101			
FI	-0.0009	0.0002	0.000	-0.0009	0.0002	0.000	-0.0010	0.0003	0.002	-0.0010	0.0003	0.000
Female	0.0209	0.0161	0.194				0.0261	0.0248	0.291			
Age	-0.0006	0.0009	0.504				-0.0010	0.0010	0.330			
/sigma_u							0.0593	0.0122	0.000	0.0618	0.0106	0.000
/sigma_e							0.0793	0.0060	0.000	0.0833	0.0057	0.000
rho							0.3586	0.1158				
AIC	-455.30			-459.93			-242.36			-250.78		
BIC	-387.07			-436.05			-174.13			-226.90		
MAE	0.0662			0.0671			0.0665			0.0680		
RMSE	0.0863			0.0895			0.0854			0.0892		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error, SE: standard error.

Table A4.2 (F) Linear mixed-effects and random-effects Tobit models using QLQ-C30 (Japan).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.2735	0.1053	0.009	0.3203	0.0538	0.000	0.2740	0.1269	0.031	0.1870	0.0716	0.009
GH	0.0014	0.0004	0.001	0.0014	0.0004	0.000	0.0016	0.0005	0.002	0.0019	0.0005	0.000
PF	0.0038	0.0007	0.000	0.0035	0.0006	0.000	0.0039	0.0007	0.000	0.0042	0.0006	0.000
RF	0.0005	0.0004	0.210				0.0005	0.0005	0.284			
EF	0.0016	0.0004	0.000	0.0013	0.0004	0.002	0.0018	0.0006	0.001	0.0018	0.0005	0.000
CF	-0.0004	0.0005	0.448				-0.0004	0.0006	0.435			
SF	0.0001	0.0003	0.589				0.0002	0.0004	0.627			
FA	0.0011	0.0006	0.050				0.0011	0.0005	0.050	0.0011	0.0005	0.037
NV	-0.0025	0.0011	0.023	-0.0021	0.0010	0.038	-0.0026	0.0010	0.010	-0.0025	0.0009	0.007
PA	-0.0005	0.0005	0.250				-0.0006	0.0005	0.232			
DY	0.0002	0.0004	0.556				0.0003	0.0004	0.490			
SL	0.0001	0.0003	0.622				0.0001	0.0004	0.808			
AP	0.0001	0.0003	0.759				0.0001	0.0004	0.767			
CO	0.0005	0.0003	0.146				0.0004	0.0004	0.237			
DI	-0.0010	0.0006	0.089				-0.0011	0.0007	0.127			
FI	-0.0008	0.0003	0.002	-0.0008	0.0002	0.001	-0.0009	0.0003	0.006	-0.0009	0.0003	0.002
Female	0.0056	0.0190	0.768				0.0109	0.0269	0.686			
Age	-0.0007	0.0009	0.407				-0.0011	0.0011	0.281			
/sigma_u							0.0665	0.0123	0.000	0.0679	0.0112	0.000
/sigma_e							0.0827	0.0061	0.000	0.0854	0.0059	0.000
rho							0.3925	0.1087		0.3878	0.0948	
AIC	-426.59			-431.40			-224.75			-236.05		
BIC	-358.35			-404.11			-156.52			-205.35		
MAE	0.0741			0.0745			0.0735			0.0752		
RMSE	0.0933			0.0949			0.0926			0.0945		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.2 (G) Linear mixed-effects and random-effects Tobit models using QLQ-C30 (China).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.2412	0.1272	0.058	0.3730	0.0753	0.000	0.2356	0.1561	0.131	0.1670	0.0922	0.070
GH	0.0017	0.0006	0.005	0.0016	0.0004	0.000	0.0020	0.0006	0.002	0.0021	0.0006	0.001
PF	0.0051	0.0009	0.000	0.0045	0.0009	0.000	0.0053	0.0008	0.000	0.0052	0.0008	0.000
RF	0.0006	0.0005	0.274				0.0006	0.0006	0.264			
EF	0.0013	0.0005	0.010				0.0015	0.0007	0.023	0.0012	0.0006	0.048
CF	-0.0006	0.0008	0.419				-0.0007	0.0007	0.338			
SF	0.0001	0.0004	0.720				0.0002	0.0005	0.708			
FA	0.0016	0.0007	0.031				0.0015	0.0006	0.024	0.0016	0.0006	0.011
NV	-0.0030	0.0016	0.054				-0.0032	0.0012	0.006	-0.0029	0.0011	0.007
PA	-0.0011	0.0006	0.045	-0.0016	0.0005	0.004	-0.0012	0.0006	0.039	-0.0013	0.0005	0.023
DY	0.0001	0.0004	0.764				0.0002	0.0005	0.634			
SL	0.0003	0.0003	0.330				0.0002	0.0004	0.573			
AP	0.0002	0.0004	0.574				0.0003	0.0004	0.544			
CO	0.0008	0.0004	0.053	0.0007	0.0003	0.012	0.0007	0.0004	0.092	0.0008	0.0004	0.048
DI	-0.0017	0.0008	0.039	-0.0019	0.0008	0.012	-0.0016	0.0008	0.038	-0.0019	0.0008	0.016
FI	-0.0013	0.0003	0.000	-0.0012	0.0003	0.000	-0.0014	0.0004	0.001	-0.0014	0.0004	0.000
Female	0.0054	0.0243	0.823				0.0130	0.0351	0.711			
Age	-0.0012	0.0011	0.286				-0.0018	0.0014	0.200			
/sigma_u							0.0944	0.0139	0.000	0.0990	0.0129	0.000
/sigma_e							0.0938	0.0069	0.000	0.0939	0.0065	0.000
rho							0.5033	0.0942		0.5264	0.0824	
AIC	-336.71			-336.13			-172.03			-182.49		
BIC	-268.48			-305.43			-103.80			-141.55		
MAE	0.0895			0.0923			0.0900			0.0904		
RMSE	0.1248			0.1296			0.1233			0.1264		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Tables A4.3 Linear mixed-effects and random-effects Tobit models using QLQ-H&N35.

Table A4.3 (A) Linear mixed-effects and random-effects Tobit models using QLQ-H&N35 (England).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.9534	0.0609	0.000	0.9200	0.0118	0.000	1.0225	0.0885	0.000	0.9434	0.0172	0.000
HNPA	-0.0020	0.0006	0.002	-0.0020	0.0005	0.000	-0.0023	0.0006	0.000	-0.0023	0.0005	0.000
HNSW	<0.0001	0.0006	0.978				<0.0001	0.0006	0.943			
HNSE	0.0004	0.0004	0.296				0.0003	0.0004	0.482			
HNSP	-0.0004	0.0005	0.359				-0.0004	0.0005	0.430			
HNSO	0.0004	0.0005	0.468				0.0004	0.0007	0.539			
HNSC	-0.0027	0.0008	0.000	-0.0026	0.0006	0.000	-0.0029	0.0007	0.000	-0.0028	0.0005	0.000
HNSX	<0.0001	0.0005	0.978				-0.0001	0.0003	0.761			
HNTE	-0.0003	0.0002	0.143				-0.0003	0.0003	0.249			
HNOM	-0.0004	0.0003	0.237				-0.0005	0.0004	0.162			
HNDR	0.0003	0.0003	0.301				0.0004	0.0004	0.250			
HNSS	-0.0001	0.0002	0.804				-0.0002	0.0004	0.575			
HNCO	-0.0005	0.0004	0.210				-0.0006	0.0004	0.146			
HNFI	-0.0013	0.0005	0.007	-0.0014	0.0005	0.004	-0.0013	0.0007	0.058	-0.0014	0.0007	0.043
HNPk	<0.0001	0.0002	0.963				-0.0001	0.0002	0.690			
HNNU	-0.0003	0.0002	0.203				-0.0004	0.0002	0.054			
HNFE	-0.0003	0.0003	0.416				-0.0003	0.0004	0.474			
HNWL	0.0004	0.0002	0.008				0.0005	0.0002	0.018			
HNWG	0.0002	0.0001	0.060				0.0003	0.0002	0.106			
Female	0.0379	0.0250	0.131				0.0478	0.0344	0.165			
Age	-0.0006	0.0009	0.494				-0.0013	0.0013	0.336			
/sigma_u							0.0935	0.0127	0.000	0.1050	0.0119	0.000
/sigma_e							0.0878	0.0062	0.000	0.0902	0.0060	0.000
Rho							0.5314	0.0857		0.5755	0.0683	
AIC	-357.22			-377.11			-174.87			-189.21		
BIC	-278.96			-356.59			-96.61			-168.69		
MAE	0.0809			0.0847			0.0802			0.0824		
RMSE	0.1167			0.1252			0.1165			0.1253		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.3 (B) Linear mixed-effects and random-effects Tobit models using QLQ-H&N35 (Netherlands).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.9220	0.0738	0.000	0.8818	0.0141	0.000	1.0042	0.1078	0.000	0.9088	0.0209	0.000
HNPA	-0.0025	0.0007	0.001	-0.0024	0.0006	0.000	-0.0028	0.0007	0.000	-0.0028	0.0006	0.000
HNSW	-0.0002	0.0008	0.815				-0.0002	0.0008	0.784			
HNSE	0.0005	0.0005	0.275				0.0004	0.0005	0.455			
HNSP	-0.0004	0.0005	0.508				-0.0004	0.0006	0.575			
HNSO	0.0004	0.0006	0.547				0.0004	0.0008	0.587			
HNSC	-0.0033	0.0009	0.000	-0.0031	0.0007	0.000	-0.0035	0.0009	0.000	-0.0033	0.0006	0.000
HNSX	<0.0001	0.0006	0.984				-0.0001	0.0004	0.821			
HNTE	-0.0003	0.0002	0.223				-0.0003	0.0003	0.365			
HNOM	-0.0004	0.0004	0.312				-0.0005	0.0004	0.219			
HNDR	0.0005	0.0003	0.093				0.0007	0.0004	0.109			
HNSS	-0.0002	0.0003	0.499				-0.0004	0.0004	0.389			
HNCO	-0.0006	0.0005	0.283				-0.0006	0.0005	0.185			
HNFI	-0.0017	0.0006	0.007	-0.0018	0.0006	0.003	-0.0016	0.0008	0.049	-0.0018	0.0008	0.033
HNPk	<0.0001	0.0002	0.938				-0.0001	0.0003	0.615			
HNNU	-0.0004	0.0003	0.172				-0.0005	0.0002	0.051			
HNFE	-0.0004	0.0004	0.304				-0.0004	0.0005	0.368			
HNWL	0.0006	0.0002	0.010				0.0007	0.0003	0.016			
HNWG	0.0003	0.0002	0.044				0.0004	0.0002	0.081			
Female	0.0493	0.0307	0.108				0.0608	0.0418	0.146			
Age	-0.0008	0.0011	0.454				-0.0016	0.0016	0.313			
/sigma_u							0.1144	0.0153	0.000	0.1287	0.0145	0.000
/sigma_e							0.1062	0.0076	0.000	0.1090	0.0073	0.000
Rho							0.5372	0.0843		0.5821	0.0677	
AIC	-274.37			-292.37			-100.17			-113.22		
BIC	-196.11			-271.85			-21.91			-92.70		
MAE	0.0987			0.1031			0.0978			0.1018		
RMSE	0.1418			0.1520			0.1410			0.1519		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.3 (C) Linear mixed-effects and random-effects Tobit models using QLQ-H&N35 (Canada).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	p	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.9220	0.0609	0.000	0.8908	0.0097	0.000	0.9833	0.0845	0.000	0.9121	0.0157	0.000
HNPA	-0.0018	0.0005	0.001	-0.0017	0.0004	0.000	-0.0020	0.0005	0.000	-0.0020	0.0004	0.000
HNSW	-0.0001	0.0005	0.900				-0.0001	0.0006	0.861			
HNSE	0.0004	0.0004	0.278				0.0003	0.0004	0.452			
HNSP	-0.0003	0.0004	0.533				-0.0002	0.0005	0.578			
HNSO	0.0002	0.0004	0.593				0.0003	0.0006	0.646			
HNSC	-0.0025	0.0007	0.000	-0.0023	0.0006	0.000	-0.0026	0.0006	0.000	-0.0025	0.0005	0.000
HNSX	-0.0001	0.0004	0.830				-0.0002	0.0003	0.592			
HNTE	-0.0002	0.0002	0.181				-0.0002	0.0002	0.326			
HNOM	-0.0003	0.0003	0.252				-0.0005	0.0003	0.153			
HNDR	0.0004	0.0003	0.184				0.0005	0.0003	0.134			
HNSS	<0.0001	0.0002	0.909				-0.0001	0.0003	0.778			
HNCO	-0.0003	0.0004	0.369				-0.0004	0.0003	0.222			
HNFI	-0.0013	0.0005	0.007	-0.0013	0.0004	0.002	-0.0013	0.0006	0.036	-0.0014	0.0006	0.025
HNPk	<0.0001	0.0002	0.968				-0.0001	0.0002	0.678			
HNNU	-0.0002	0.0002	0.259				-0.0003	0.0002	0.086			
HNFE	-0.0001	0.0003	0.667				-0.0002	0.0003	0.652			
HNWL	0.0004	0.0001	0.017				0.0004	0.0002	0.020			
HNWG	0.0002	0.0001	0.059				0.0003	0.0002	0.083			
Female	0.0351	0.0220	0.111				0.0433	0.0327	0.185			
Age	-0.0007	0.0010	0.457				-0.0013	0.0013	0.313			
/sigma_u							0.0940	0.0113	0.000	0.1028	0.0107	0.000
/sigma_e							0.0732	0.0052	0.000	0.0753	0.0050	0.000
Rho							0.6221	0.0724		0.6509	0.0590	
AIC	-412.22			-435.47			-225.28			-242.50		
BIC	-333.96			-414.95			-147.02			-221.97		
MAE	0.0753			0.0776			0.0752			0.0747		
RMSE	0.1131			0.1201			0.1126			0.1203		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.3 (D) Linear mixed-effects and random-effects Tobit models using QLQ-H&N35 (Uruguay).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	p	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.9719	0.0448	0.000	0.9575	0.0081	0.000	1.0176	0.0606	0.000	0.9736	0.0115	0.000
HNPA	-0.0012	0.0004	0.002	-0.0011	0.0003	0.001	-0.0014	0.0004	0.000	-0.0014	0.0003	0.000
HNSW	-0.0001	0.0004	0.793				-0.0001	0.0004	0.777			
HNSE	0.0003	0.0002	0.310				0.0002	0.0003	0.521			
HNSP	-0.0002	0.0003	0.562				-0.0002	0.0003	0.627			
HNSO	0.0001	0.0003	0.623				0.0002	0.0004	0.688			
HNSC	-0.0019	0.0005	0.000	-0.0018	0.0004	0.000	-0.0020	0.0005	0.000	-0.0019	0.0003	0.000
HNSX	<0.0001	0.0003	0.986				<0.0001	0.0002	0.826			
HNTE	-0.0001	0.0002	0.438				-0.0001	0.0002	0.403			
HNOM	-0.0003	0.0002	0.123				-0.0004	0.0002	0.093			
HNDR	0.0003	0.0002	0.119				0.0004	0.0002	0.073			
HNSS	0.0001	0.0002	0.656				<0.0001	0.0002	0.965			
HNCO	-0.0003	0.0003	0.271				-0.0004	0.0003	0.162			
HNFI	-0.0009	0.0003	0.008	-0.0010	0.0003	0.003	-0.0009	0.0004	0.052	-0.0010	0.0005	0.030
HNPk	<0.0001	0.0001	0.817				-0.0001	0.0001	0.501			
HNNU	-0.0002	0.0001	0.234				-0.0002	0.0001	0.085			
HNFE	-0.0002	0.0002	0.345				-0.0002	0.0003	0.389			
HNWL	0.0003	0.0001	0.009				0.0004	0.0001	0.014			
HNWG	0.0001	0.0001	0.149				0.0002	0.0001	0.124			
Female	0.0253	0.0161	0.115				0.0314	0.0234	0.181			
Age	-0.0004	0.0007	0.543				-0.0008	0.0009	0.361			
/sigma_u							0.0657	0.0084	0.000	0.0719	0.0080	0.000
/sigma_e							0.0561	0.0040	0.000	0.0581	0.0039	0.000
Rho							0.5781	0.0799		0.6044	0.0662	
AIC	-541.55			-565.08			-332.90			-349.94		
BIC	-463.29			-544.55			-254.64			-329.41		
MAE	0.0546			0.0560			0.0542			0.0534		
RMSE	0.0796			0.0848			0.0795			0.0850		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.3 (E) Linear mixed-effects and random-effects Tobit models using QLQ-H&N35 (South Korea).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	p	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.9002	0.0617	0.000	0.8721	0.0118	0.000	0.9561	0.0752	0.000	0.8959	0.0151	0.000
HNPA	-0.0017	0.0005	0.000	-0.0017	0.0004	0.000	-0.0019	0.0005	0.001	-0.0019	0.0005	0.000
HNSW	<0.0001	0.0005	0.996				<0.0001	0.0006	0.993			
HNSE	0.0003	0.0003	0.324				0.0003	0.0004	0.502			
HNSP	-0.0003	0.0004	0.486				-0.0003	0.0005	0.572			
HNSO	0.0003	0.0004	0.544				0.0003	0.0006	0.647			
HNSC	-0.0022	0.0006	0.000	-0.0022	0.0005	0.000	-0.0023	0.0007	0.000	-0.0023	0.0005	0.000
HNSX	-0.0001	0.0003	0.625				-0.0002	0.0003	0.480			
HNTE	-0.0001	0.0002	0.489				-0.0001	0.0002	0.553			
HNOM	-0.0005	0.0002	0.021				-0.0006	0.0003	0.050			
HNDR	0.0003	0.0003	0.242				0.0004	0.0003	0.217			
HNSS	<0.0001	0.0002	0.843				-0.0002	0.0003	0.614			
HNCO	-0.0005	0.0004	0.186				-0.0005	0.0004	0.139			
HNFI	-0.0012	0.0004	0.003	-0.0013	0.0004	0.002	-0.0012	0.0006	0.062	-0.0012	0.0006	0.047
HNPk	-0.0001	0.0002	0.610				-0.0002	0.0002	0.401			
HNNU	-0.0002	0.0002	0.297				-0.0003	0.0002	0.143			
HNFE	-0.0001	0.0002	0.569				-0.0001	0.0004	0.701			
HNWL	0.0004	0.0001	0.008				0.0004	0.0002	0.031			
HNWG	0.0002	0.0001	0.145				0.0002	0.0002	0.182			
Female	0.0405	0.0200	0.043	0.0437	0.0209	0.036	0.0482	0.0294	0.102			
Age	-0.0004	0.0009	0.668				-0.0009	0.0011	0.405			
/sigma_u							0.0761	0.0114	0.000	0.0892	0.0108	0.000
/sigma_e							0.0827	0.0059	0.000	0.0836	0.0056	0.000
Rho							0.4585	0.0924		0.5322	0.0740	
AIC	-408.89			-431.10			-196.81			-213.03		
BIC	-330.63			-407.16			-118.55			-192.51		
MAE	0.0730			0.0754			0.0732			0.0783		
RMSE	0.0991			0.1048			0.0987			0.1073		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.3 (F) Linear mixed-effects and random-effects Tobit models using QLQ-H&N35 (Japan).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	p	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	p
Intercept	0.9232	0.0696	0.000	0.8735	0.0160	0.000	0.9870	0.0930	0.000	0.9045	0.0189	0.000
HNPA	-0.0016	0.0005	0.002	-0.0018	0.0005	0.000	-0.0018	0.0006	0.004	-0.0022	0.0005	0.000
HNSW	-0.0001	0.0006	0.834				-0.0001	0.0007	0.824			
HNSE	0.0001	0.0004	0.864				<0.0001	0.0004	0.927			
HNSP	-0.0007	0.0005	0.195				-0.0007	0.0005	0.210			
HNSO	0.0002	0.0006	0.717				0.0002	0.0007	0.718			
HNSC	-0.0022	0.0007	0.002	-0.0025	0.0005	0.000	-0.0023	0.0007	0.001	-0.0025	0.0005	0.000
HNSX	0.0001	0.0003	0.663				0.0001	0.0003	0.832			
HNTE	-0.0003	0.0002	0.154				-0.0003	0.0003	0.317			
HNOM	-0.0008	0.0003	0.002	-0.0009	0.0003	0.001	-0.0010	0.0004	0.010	-0.0010	0.0004	0.006
HNDR	0.0002	0.0003	0.370				0.0004	0.0004	0.330			
HNSS	<0.0001	0.0003	0.998				-0.0001	0.0004	0.715			
HNCO	-0.0005	0.0004	0.171				-0.0006	0.0004	0.153			
HNFI	-0.0010	0.0005	0.027	-0.0012	0.0005	0.007	-0.0010	0.0007	0.140			
HNPk	-0.0001	0.0002	0.750				-0.0001	0.0002	0.512			
HNNU	-0.0003	0.0002	0.241				-0.0003	0.0002	0.105			
HNFE	-0.0004	0.0004	0.263				-0.0004	0.0004	0.329			
HNWL	0.0004	0.0002	0.014	0.0004	0.0002	0.026	0.0005	0.0002	0.034			
HNWG	0.0004	0.0001	0.007	0.0003	0.0001	0.023	0.0005	0.0002	0.021			
Female	0.0140	0.0253	0.579				0.0216	0.0362	0.550			
Age	-0.0007	0.0011	0.515				-0.0013	0.0014	0.347			
/sigma_u							0.1019	0.0132	0.000	0.1146	0.0125	0.000
/sigma_e							0.0869	0.0063	0.000	0.0895	0.0060	0.000
Rho							0.5786	0.0807		0.6212	0.0638	
AIC	-356.22			-377.78			-159.06			-172.81		
BIC	-277.96			-346.99			-80.80			-152.29		
MAE	0.0928			0.0973			0.0908			0.0985		
RMSE	0.1205			0.1253			0.1199			0.1304		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.3 (G) Linear mixed-effects and random-effects Tobit models using QLQ-H&N35 (China).

	Linear						Tobit					
	Full model			Reduced model			Full model			Reduced model		
	Coefficient	SE	p	Coefficient	SE	P	Coefficient	SE	p	Coefficient	SE	P
Intercept	0.9601	0.0830	0.000	0.9106	0.0182	0.000	1.0413	0.1169	0.000	0.9391	0.0230	0.000
HNPA	-0.0021	0.0007	0.002	-0.0023	0.0007	0.000	-0.0024	0.0007	0.001	-0.0027	0.0006	0.000
HNSW	-0.0001	0.0007	0.851				-0.0002	0.0008	0.817			
HNSE	0.0003	0.0005	0.591				0.0001	0.0005	0.821			
HNSP	-0.0006	0.0006	0.354				-0.0006	0.0006	0.370			
HNSO	0.0002	0.0006	0.803				0.0002	0.0008	0.793			
HNSC	-0.0033	0.0009	0.000	-0.0034	0.0008	0.000	-0.0034	0.0008	0.000	-0.0036	0.0006	0.000
HNSX	0.0001	0.0004	0.791				<0.0001	0.0004	0.960			
HNTE	-0.0003	0.0002	0.126				-0.0004	0.0003	0.255			
HNOM	-0.0009	0.0004	0.018	-0.0009	0.0004	0.018	-0.0011	0.0004	0.017	-0.0011	0.0004	0.010
HNDR	0.0003	0.0004	0.333				0.0005	0.0004	0.235			
HNSS	0.0002	0.0003	0.518				<0.0001	0.0004	0.905			
HNCO	-0.0005	0.0005	0.296				-0.0007	0.0005	0.169			
HNFI	-0.0014	0.0006	0.027	-0.0016	0.0006	0.006	-0.0014	0.0008	0.084	-0.0017	0.0008	0.039
HNPk	-0.0001	0.0002	0.623				-0.0002	0.0003	0.385			
HNNU	-0.0003	0.0002	0.243				-0.0004	0.0002	0.110			
HNFE	-0.0004	0.0005	0.463				-0.0004	0.0005	0.392			
HNWL	0.0005	0.0002	0.019	0.0005	0.0002	0.026	0.0006	0.0003	0.015	0.0006	0.0003	0.021
HNWG	0.0005	0.0002	0.007	0.0004	0.0002	0.028	0.0006	0.0002	0.012	0.0005	0.0002	0.029
Female	0.0231	0.0299	0.439				0.0330	0.0452	0.465			
Age	-0.0009	0.0013	0.499				-0.0016	0.0018	0.349			
/sigma_u							0.1312	0.0154	0.000	0.1382	0.0145	0.000
/sigma_e							0.0993	0.0070	0.000	0.1014	0.0068	0.000
Rho							0.6359	0.0700		0.6501	0.0596	
AIC	-273.52			-294.74			-109.41			-125.75		
BIC	-195.26			-263.96			-31.15			-94.96		
MAE	0.1112			0.1150			0.1098			0.1114		
RMSE	0.1554			0.1611			0.1554			0.1617		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Tables A4.4 ALDVM models using QLQ-C30.

Table A4.4 (A) ALDVM model using QLQ-C30 (England).

	Coefficient	SE	P
Component 1			
Intercept	0.8685	0.0039	0.000
GH	0.0010	<0.0001	0.000
PF	-0.0001	<0.0001	0.000
RF	-0.0005	<0.0001	0.000
EF	0.0007	<0.0001	0.000
CF	0.0006	<0.0001	0.000
SF	-0.0003	<0.0001	0.000
FA	-0.0013	<0.0001	0.000
NV	0.0005	<0.0001	0.000
PA	-0.0008	<0.0001	0.000
DY	-0.0011	<0.0001	0.000
SL	0.0004	<0.0001	0.000
AP	-0.0002	<0.0001	0.000
CO	0.0013	<0.0001	0.000
DI	0.0012	<0.0001	0.000
FI	-0.0011	<0.0001	0.000
Female	0.0398	0.0007	0.000
Age	-0.0012	<0.0001	0.000
Component 2			
Intercept	0.3562	0.1514	0.019
GH	0.0010	0.0007	0.144
PF	0.0042	0.0009	0.000
RF	0.0012	0.0007	0.082
EF	0.0016	0.0005	0.003
CF	-0.0007	0.0007	0.290
SF	0.0003	0.0005	0.551
FA	0.0005	0.0007	0.541
NV	-0.0037	0.0018	0.034
PA	-0.0013	0.0006	0.044
DY	0.0003	0.0005	0.552
SL	0.0003	0.0004	0.487
AP	0.0002	0.0005	0.701
CO	-0.0004	0.0003	0.271
DI	-0.0022	0.0009	0.018
FI	-0.0009	0.0004	0.017
Female	0.0477	0.0217	0.028
Age	-0.0013	0.0010	0.211
Probability (Component 1)	-1.6222	0.2222	0.000
Constant			
/lns_1	-7.0701	0.1472	0.000
/lns_2	-2.2525	0.0872	0.000
Sigma 1	0.0008	0.0001	
Sigma 2	0.1051	0.0092	
Probability (Component 1)	0.1649	0.0306	
Probability (Component 2)	0.8351	0.0306	
AIC	-341.64		
BIC	-208.59		
MAE	0.0666		
RMSE	0.0943		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.4 (B) ALDVM model using QLQ-C30 (Netherlands).

	Coefficient	SE	P
Component 1			
Intercept	0.9311	0.0023	0.000
GH	0.0003	<0.0001	0.000
PF	0.0025	<0.0001	0.000
RF	0.0002	<0.0001	0.000
EF	-0.0011	<0.0001	0.000
CF	-0.0008	<0.0001	0.000
SF	0.0008	<0.0001	0.000
FA	-0.0049	<0.0001	0.000
NV	0.0017	<0.0001	0.000
PA	-0.0017	<0.0001	0.000
DY	-0.0034	<0.0001	0.000
SL	0.0006	<0.0001	0.000
AP	-0.0004	<0.0001	0.000
CO	0.0001	<0.0001	0.000
DI	-0.0040	<0.0001	0.000
FI	0.0005	<0.0001	0.000
Female	0.0313	0.0004	0.000
Age	-0.0013	<0.0001	0.000
Component 2			
Intercept	0.8319	0.0553	0.000
GH	0.0009	0.0007	0.168
PF	-0.0008	0.0007	0.238
RF	-0.0002	0.0001	0.244
EF	0.0013	0.0003	0.000
CF	0.0005	0.0004	0.148
SF	-0.0008	0.0002	0.000
FA	0.0002	0.0004	0.594
NV	-0.0009	0.0003	0.001
PA	-0.0003	0.0003	0.219
DY	-0.0007	0.0003	0.011
SL	-0.0013	0.0003	0.000
AP	0.0006	0.0002	0.015
CO	0.0002	0.0002	0.325
DI	-0.0002	0.0004	0.553
FI	-0.0003	0.0003	0.363
Female	-0.0090	0.0166	0.585
Age	-0.0005	0.0008	0.563
Component 3			
Intercept	-0.5478	0.3278	0.095
GH	0.0034	0.0015	0.019
PF	0.0087	0.0014	0.000
RF	0.0022	0.0010	0.029
EF	0.0054	0.0016	0.001
CF	-0.0010	0.0015	0.514
SF	-0.0014	0.0013	0.279
FA	0.0036	0.0015	0.018
NV	-0.0117	0.0022	0.000
PA	-0.0012	0.0009	0.200
DY	0.0034	0.0013	0.009
SL	0.0016	0.0010	0.104
AP	0.0008	0.0011	0.469
CO	-0.0003	0.0008	0.669
DI	-0.0026	0.0018	0.158
FI	-0.0028	0.0009	0.003
Female	0.0634	0.0501	0.205
Age	-0.0014	0.0021	0.504

Table A4.4 (B) (cont.) ALDVM model using QLQ-C30 (Netherlands).

	Coefficient	SE	P
Probability (Component 1) Constant	-1.4306	0.3064	0.000
Probability (Component 2) Constant	-0.2280	0.3213	0.478
/lns_1	-9.9034	0.4471	0.000
/lns_2	-3.7173	0.2172	0.000
/lns_3	-2.0715	0.1109	0.000
Sigma 1	<0.0001	<0.0001	
Sigma 2	0.0243	0.0053	
Sigma 3	0.1260	0.0140	
Probability (Component 1)	0.1175	0.0256	
Probability (Component 2)	0.3911	0.0689	
Probability (Component 3)	0.4913	0.0738	
AIC	-383.64		
BIC	-192.59		
MAE	0.0872		
RMSE	0.1196		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.4 (C) ALDVM model using QLQ-C30 (Canada).

	Coefficient	SE	P
Component 1			
Intercept	0.7767	0.0412	0.000
GH	0.0007	0.0003	0.043
PF	0.0010	0.0002	0.000
RF	-0.0001	0.0001	0.252
EF	0.0014	0.0002	0.000
CF	-0.0003	0.0001	0.018
SF	-0.0005	0.0002	0.024
FA	0.0013	0.0002	0.000
NV	-0.0012	0.0002	0.000
PA	-0.0012	0.0001	0.000
DY	0.0007	0.0002	0.000
SL	-0.0008	0.0002	0.000
AP	0.0005	0.0001	0.000
CO	-0.0004	0.0001	0.000
DI	-0.0012	0.0002	0.000
FI	-0.0003	0.0002	0.212
Female	-0.0188	0.0054	0.001
Age	-0.0006	0.0005	0.215
Component 2			
Intercept	0.3011	0.1806	0.095
GH	0.0001	0.0005	0.789
PF	0.0035	0.0011	0.001
RF	0.0029	0.0008	0.000
EF	0.0005	0.0006	0.422
CF	0.0006	0.0006	0.355
SF	-0.0006	0.0005	0.200
FA	0.0001	0.0008	0.926
NV	-0.0041	0.0014	0.003
PA	-0.0017	0.0008	0.028
DY	0.0004	0.0006	0.478
SL	0.0002	0.0003	0.507
AP	0.0009	0.0005	0.071
CO	0.0005	0.0003	0.115
DI	-0.0016	0.0010	0.099
FI	-0.0006	0.0004	0.174
Female	0.0453	0.0210	0.031
Age	-0.0010	0.0010	0.330
Probability (Component 1)	-0.4538	0.3152	0.150
Constant			
/lns_1	-4.3653	0.4044	0.000
/lns_2	-2.5282	0.1004	0.000
Sigma 1	0.0127	0.0051	
Sigma 2	0.0798	0.0080	
Probability (Component 1)	0.3884	0.0749	
Probability (Component 2)	0.6115	0.0749	
AIC	-547.84		
BIC	-414.78		
MAE	0.0624		
RMSE	0.0921		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.4 (D) ALDVM model using QLQ-C30 (Uruguay).

	<i>Full model</i>			<i>Reduced model</i>		
	Coefficient	SE	P	Coefficient	SE	P
Component 1						
Intercept	0.6678	0.1005	0.000	0.6431	0.0563	0.000
GH	0.0008	0.0004	0.051	0.0012	0.0004	0.001
PF	0.0023	0.0006	0.000	0.0027	0.0006	0.000
RF	0.0009	0.0004	0.047			
EF	0.0003	0.0003	0.311			
CF	-0.0001	0.0005	0.757			
SF	0.0002	0.0002	0.454			
FA	0.0004	0.0005	0.353			
NV	-0.0019	0.0008	0.023	-0.0016	0.0008	0.040
PA	-0.0010	0.0004	0.017	-0.0011	0.0005	0.020
DY	0.0001	0.0003	0.637			
SL	-0.0002	0.0003	0.448			
AP	0.0004	0.0003	0.124			
CO	0.0001	0.0002	0.738			
DI	-0.0005	0.0007	0.464			
FI	-0.0007	0.0002	0.005	-0.0007	0.0002	0.001
Female	0.0375	0.0150	0.013	0.0373	0.0142	0.009
Age	-0.0013	0.0007	0.069			
/lns_1	-2.6205	0.0892	0.000	-2.5726	0.0898	0.000
Sigma 1	0.0728	0.0065		0.0763	0.0069	
AIC	-359.02			-362.34		
BIC	-294.20			-335.05		
MAE	0.0428			0.0455		
RMSE	0.0638			0.0670		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.4 (E) ALDVM model using QLQ-C30 (South Korea).

	<i>Full model</i>			<i>Reduced model</i>		
	Coefficient	SE	P	Coefficient	SE	P
Component 1						
Intercept	0.4606	0.1239	0.000	0.4554	0.0497	0.000
GH	0.0014	0.0005	0.006	0.0017	0.0005	0.000
PF	0.0029	0.0007	0.000	0.0034	0.0006	0.000
RF	0.0008	0.0005	0.108			
EF	0.0008	0.0004	0.069			
CF	-0.0003	0.0005	0.559			
SF	0.0002	0.0003	0.567			
FA	0.0006	0.0006	0.318			
NV	-0.0021	0.0011	0.058			
PA	-0.0010	0.0005	0.042	-0.0014	0.0005	0.011
DY	0.0002	0.0004	0.673			
SL	-0.0001	0.0003	0.704			
AP	0.0004	0.0003	0.215			
CO	<0.0001	0.0003	0.961			
DI	-0.0005	0.0007	0.423			
FI	-0.0009	0.0003	0.005	-0.0009	0.0002	0.000
Female	0.0417	0.0188	0.027	0.0390	0.0182	0.032
Age	-0.0012	0.0011	0.262			
/lns_1	-2.3405	0.0729	0.000	-2.3013	0.0736	0.000
Sigma 1	0.0963	0.0070		0.1001	0.0074	
AIC	-232.56			-241.32		
BIC	-167.74			-217.44		
MAE	0.0655			0.0679		
RMSE	0.0835			0.0866		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.4 (F) ALDVM model using QLQ-C30 (Japan).

	<i>Full model</i>			<i>Reduced model</i>		
	Coefficient	SE	P	Coefficient	SE	P
Component 1						
Intercept	0.2714	0.1365	0.047	0.2000	0.0637	0.002
GH	0.0018	0.0005	0.001	0.0022	0.0005	0.000
PF	0.0038	0.0008	0.000	0.0042	0.0008	0.000
RF	0.0007	0.0005	0.160			
EF	0.0017	0.0005	0.000	0.0017	0.0004	0.000
CF	-0.0002	0.0005	0.699			
SF	0.0002	0.0004	0.499			
FA	0.0009	0.0006	0.129			
NV	-0.0029	0.0012	0.016	-0.0026	0.0011	0.021
PA	-0.0010	0.0005	0.062			
DY	0.0002	0.0004	0.576			
SL	-0.0004	0.0003	0.271			
AP	0.0003	0.0004	0.372			
CO	0.0001	0.0004	0.745			
DI	-0.0005	0.0008	0.510			
FI	-0.0007	0.0003	0.025	-0.0009	0.0003	0.006
Female	0.0297	0.0206	0.150			
Age	-0.0014	0.0011	0.210			
/lns_1	-2.2743	0.0693	0.000	-2.2276	0.0710	0.000
Sigma 1	0.1029	0.0071		0.1078	0.0076	
AIC	-212.02			-217.93		
BIC	-147.20			-194.05		
MAE	0.0720			0.0746		
RMSE	0.0898			0.0939		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.4 (G) ALDVM model using QLQ-C30 (China).

	Coefficient	SE	P
Component 1			
Intercept	0.7636	0.0480	0.000
GH	0.0018	0.0004	0.000
PF	0.0014	0.0005	0.008
RF	0.0000	0.0001	0.517
EF	0.0009	0.0004	0.026
CF	<0.0001	0.0001	0.939
SF	-0.0009	0.0001	0.000
FA	0.0010	0.0002	0.000
NV	-0.0030	0.0003	0.000
PA	-0.0015	0.0004	0.000
DY	0.0019	0.0003	0.000
SL	-0.0021	0.0003	0.000
AP	0.0011	0.0001	0.000
CO	-0.0008	0.0001	0.000
DI	-0.0017	0.0002	0.000
FI	-0.0001	0.0001	0.642
Female	0.0086	0.0213	0.687
Age	-0.0010	0.0006	0.120
Component 2			
Intercept	-0.2102	0.4200	0.617
GH	0.0013	0.0010	0.201
PF	0.0079	0.0021	0.000
RF	0.0018	0.0009	0.039
EF	0.0022	0.0009	0.018
CF	0.0009	0.0010	0.396
SF	-0.0007	0.0006	0.219
FA	<0.0001	0.0014	0.977
NV	-0.0069	0.0027	0.010
PA	-0.0011	0.0014	0.401
DY	0.0011	0.0012	0.358
SL	0.0008	0.0006	0.213
AP	0.0017	0.0008	0.035
CO	0.0005	0.0005	0.360
DI	-0.0004	0.0011	0.732
FI	-0.0015	0.0005	0.001
Female	0.0369	0.0301	0.220
Age	-0.0014	0.0017	0.435
Component 3			
Intercept	1.0463	0.0032	0.000
GH	<0.0001	<0.0001	0.037
PF	<0.0001	<0.0001	0.657
RF	0.0001	<0.0001	0.000
EF	0.0001	<0.0001	0.000
CF	-0.0002	<0.0001	0.000
SF	0.0003	<0.0001	0.000
FA	-0.0005	<0.0001	0.000
NV	-0.0052	<0.0001	0.000
PA	-0.0060	<0.0001	0.000
DY	-0.0026	<0.0001	0.000
SL	-0.0001	<0.0001	0.000
AP	-0.0008	<0.0001	0.000
CO	-0.0007	<0.0001	0.000
DI	-0.0043	<0.0001	0.000
FI	-0.0002	<0.0001	0.000
Female	-0.1055	0.0005	0.000
Age	-0.0002	<0.0001	0.000

Table A4.4 (G) (cont.) ALDVM model using QLQ-C30 (China).

	Coefficient	SE	P
Probability (Component 1) Constant	0.4192	0.3344	0.210
Probability (Component 2) Constant	1.4316	0.2279	0.000
/lns_1	-4.4991	0.4309	0.000
/lns_2	-2.2115	0.0974	0.000
/lns_3	-7.5695	0.1339	0.000
Sigma 1	0.0111	0.0048	
Sigma 2	0.1095	0.0107	
Sigma 3	0.0005	0.0001	
Probability (Component 1)	0.2268	0.0561	
Probability (Component 2)	0.6241	0.0575	
Probability (Component 3)	0.1491	0.0273	
AIC	-313.69		
BIC	-112.40		
MAE	0.0856		
RMSE	0.1187		

GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AP: appetite loss; CO: constipation; DI: diarrhoea; FI: financial problems. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Tables A4.5 ALDVM models using QLQ-H&N35.

Table A4.5 (A) ALDVM model using QLQ-H&N35 (England).

	Coefficient	SE	P
Component 1			
Intercept	1.0587	0.0548	0.000
HNPA	-0.0038	0.0007	0.000
HNSW	-0.0004	0.0006	0.488
HNSE	0.0002	0.0004	0.570
HNSP	-0.0002	0.0005	0.667
HNSO	0.0003	0.0006	0.616
HNSC	-0.0035	0.0009	0.000
HNSX	0.0009	0.0003	0.002
HNTE	<0.0001	0.0003	0.986
HNOM	-0.0005	0.0003	0.121
HNDR	-0.0002	0.0003	0.564
HNSS	<0.0001	0.0003	0.902
HNCO	-0.0008	0.0004	0.079
HNFI	-0.0012	0.0007	0.083
HNPk	-0.0005	0.0002	0.019
HNNU	-0.0002	0.0002	0.242
HNFE	-0.0003	0.0003	0.449
HNWL	0.0005	0.0003	0.054
HNWG	0.0004	0.0001	0.006
Female	0.0407	0.0228	0.075
Age	-0.0012	0.0008	0.142
Component 2			
Intercept	0.5542	0.0013	0.000
HNPA	-0.0009	<0.0001	0.000
HNSW	-0.0038	<0.0001	0.000
HNSE	0.0024	<0.0001	0.000
HNSP	0.0017	<0.0001	0.000
HNSO	0.0072	<0.0001	0.000
HNSC	-0.0024	<0.0001	0.000
HNSX	-0.0017	<0.0001	0.000
HNTE	-0.0018	<0.0001	0.000
HNOM	0.0010	<0.0001	0.000
HNDR	0.0033	<0.0001	0.000
HNSS	-0.0016	<0.0001	0.000
HNCO	-0.0068	<0.0001	0.000
HNFI	-0.0049	<0.0001	0.000
HNPk	-0.0011	<0.0001	0.000
HNNU	-0.0037	<0.0001	0.000
HNFE	-0.0019	<0.0001	0.000
HNWL	0.0027	<0.0001	0.000
HNWG	0.0001	<0.0001	0.000
Female	0.2872	0.0004	0.000
Age	0.0020	<0.0001	0.000
Component 3			
Intercept	0.9462	0.0019	0.000
HNPA	-0.0001	<0.0001	0.000
HNSW	0.0001	<0.0001	0.000
HNSE	-0.0001	<0.0001	0.000
HNSP	0.0006	<0.0001	0.000
HNSO	0.0010	<0.0001	0.000
HNSC	-0.0023	<0.0001	0.000
HNSX	-0.0003	<0.0001	0.000
HNTE	0.0007	<0.0001	0.000
HNOM	-0.0008	<0.0001	0.000
HNDR	-0.0003	<0.0001	0.000
HNSS	-0.0006	<0.0001	0.000
HNCO	-0.0010	<0.0001	0.000
HNFI	-0.0020	<0.0001	0.000
HNPk	0.0005	<0.0001	0.000
HNNU	0.0003	<0.0001	0.000

Table A4.5 (A) (cont.) ALDVM model using QLQ-H&N35 (England).

	Coefficient	SE	P
HNFE	-0.0002	<0.0001	0.000
HNWL	<0.0001	<0.0001	0.000
HNWG	-0.0006	<0.0001	0.000
Female	-0.0394	0.0002	0.000
Age	-0.0005	<0.0001	0.000
Probability (Component 1) Constant	1.8421	0.2146	0.000
Probability (Component 2) Constant	-0.0618	0.3030	0.838
/lns_1	-2.4564	0.0695	0.000
/lns_2	-8.2578	0.1251	0.000
/lns_3	-10.0190	0.3320	0.000
Sigma 1	0.0857	0.0060	
Sigma 2	0.0003	<0.0001	
Sigma 3	<0.0001	<0.0001	
Probability (Component 1)	0.7648	0.0295	
Probability (Component 2)	0.1139	0.0233	
Probability (Component 3)	0.1212	0.0228	
AIC	-566.40		
BIC	-365.64		
MAE	0.0834		
RMSE	0.1133		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.5 (B) ALDVM model using QLQ-H&N35 (Netherlands).

	<i>Full model</i>			<i>Reduced model</i>		
	Coefficient	SE	p	Coefficient	SE	P
Component 1						
Intercept	1.0619	0.0927	0.000	0.9376	0.0228	0.000
HNPA	-0.0035	0.0008	0.000	-0.0037	0.0010	0.000
HNSW	-0.0003	0.0009	0.737			
HNSE	0.0007	0.0005	0.198			
HNSP	<0.0001	0.0007	0.993			
HNSO	0.0020	0.0010	0.037			
HNSC	-0.0059	0.0014	0.000	-0.0046	0.0011	0.000
HNSX	-0.0003	0.0007	0.625			
HNTE	0.0001	0.0003	0.847			
HNOM	-0.0006	0.0004	0.126			
HNDR	0.0002	0.0004	0.533			
HNSS	-0.0004	0.0004	0.343			
HNCO	-0.0011	0.0007	0.118			
HNFI	-0.0023	0.0010	0.022	-0.0026	0.0011	0.019
HNPKE	-0.0007	0.0003	0.009	-0.0008	0.0003	0.017
HNNU	-0.0005	0.0003	0.141			
HNFE	-0.0003	0.0004	0.409			
HNWL	0.0008	0.0003	0.018			
HNWG	0.0007	0.0003	0.013			
Female	0.1000	0.0354	0.005	0.1207	0.0408	0.003
Age	-0.0023	0.0014	0.106			
/lns_1	-1.8964	0.0963	0.000	-1.8170	0.1182	0.000
Sigma 1	0.1501	0.0144		0.1625	0.0192	
AIC	-65.57			-65.22		
BIC	9.29			-41.27		
MAE	0.0968			0.0995		
RMSE	0.1296			0.1411		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPKE: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.5 (C) ALDVM model using QLQ-H&N35 (Canada).

	Coefficient	SE	P
Component 1			
Intercept	0.8644	0.0200	0.000
HNPA	-0.0003	0.0003	0.282
HNSW	-0.0006	0.0006	0.352
HNSE	-0.0009	0.0002	0.000
HNSP	0.0018	0.0003	0.000
HNSO	0.0012	0.0002	0.000
HNSC	-0.0064	0.0004	0.000
HNSX	0.0003	0.0001	0.021
HNTE	-0.0005	0.0002	0.003
HNOM	-0.0004	0.0002	0.047
HNDR	-0.0001	0.0003	0.844
HNSS	0.0003	0.0002	0.106
HNCO	0.0002	0.0002	0.321
HNFI	-0.0119	0.0006	0.000
HNPk	-0.0005	0.0001	0.001
HNNU	<0.0001	0.0001	0.828
HNFE	0.0009	0.0003	0.000
HNWL	-0.0004	0.0001	0.000
HNWG	-0.0002	0.0001	0.083
Female	0.0489	0.0113	0.000
Age	0.0008	0.0003	0.010
Component 2			
Intercept	0.9904	0.0171	0.000
HNPA	-0.0024	0.0003	0.000
HNSW	<0.0001	0.0002	0.797
HNSE	<0.0001	0.0001	0.783
HNSP	-0.0005	0.0002	0.023
HNSO	-0.0003	0.0003	0.291
HNSC	0.0009	0.0003	0.007
HNSX	0.0004	0.0001	0.002
HNTE	<0.0001	0.0001	0.861
HNOM	-0.0003	0.0001	0.018
HNDR	-0.0001	0.0001	0.386
HNSS	-0.0005	0.0002	0.004
HNCO	-0.0009	0.0002	0.000
HNFI	0.0006	0.0004	0.124
HNPk	-0.0002	0.0001	0.003
HNNU	<0.0001	0.0001	0.592
HNFE	-0.0012	0.0001	0.000
HNWL	0.0004	0.0001	0.000
HNWG	0.0001	0.0001	0.050
Female	-0.0060	0.0074	0.413
Age	-0.0008	0.0003	0.007
Component 3			
Intercept	0.5233	0.0915	0.000
HNPA	-0.0065	0.0009	0.000
HNSW	0.0006	0.0005	0.205
HNSE	0.0029	0.0004	0.000
HNSP	0.0009	0.0010	0.359
HNSO	0.0046	0.0011	0.000
HNSC	-0.0028	0.0003	0.000
HNSX	-0.0013	0.0005	0.013
HNTE	0.0009	0.0004	0.021
HNOM	0.0010	0.0005	0.040
HNDR	0.0017	0.0009	0.067
HNSS	0.0005	0.0003	0.089
HNCO	-0.0039	0.0005	0.000
HNFI	-0.0002	0.0003	0.579
HNPk	-0.0006	0.0002	0.004
HNNU	-0.0008	0.0002	0.000
HNFE	0.0026	0.0003	0.000
HNWL	0.0007	0.0005	0.159
HNWG	0.0013	0.0003	0.000

Table A4.5 (C) (cont.) ALDVM model using QLQ-H&N35 (Canada).

	Coefficient	SE	P
Female	0.1885	0.0374	0.000
Age	0.0002	0.0007	0.727
Probability (Component 1) Constant	0.7863	0.2810	0.005
Probability (Component 2) Constant	0.9643	0.2808	0.001
/lns_1	-3.5098	0.1262	0.000
/lns_2	-3.8609	0.0964	0.000
/lns_3	-3.5452	0.1148	0.000
Sigma 1	0.0299	0.0038	
Sigma 2	0.0210	0.0020	
Sigma 3	0.0289	0.0033	
Probability (Component 1)	0.3773	0.0515	
Probability (Component 2)	0.4508	0.0540	
Probability (Component 3)	0.1719	0.0361	
AIC	-535.52		
BIC	-304.14		
MAE	0.0758		
RMSE	0.1097		

HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPCK: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.5 (D) ALDVM model using QLQ-H&N35 (Uruguay).

	Coefficient	SE	P
Component 1			
Intercept	1.0471	0.0282	0.000
HNPA	-0.0010	0.0003	0.000
HNSW	0.0001	0.0002	0.718
HNSE	-0.0003	0.0002	0.095
HNSP	0.0002	0.0002	0.390
HNSO	0.0005	0.0003	0.067
HNSC	-0.0008	0.0005	0.102
HNSX	<0.0001	0.0001	0.898
HNTE	0.0001	0.0001	0.142
HNOM	-0.0008	0.0001	0.000
HNDR	0.0001	0.0001	0.668
HNSS	-0.0002	0.0002	0.200
HNCO	-0.0003	0.0002	0.153
HNFI	-0.0009	0.0002	0.000
HNPk	-0.0002	0.0001	0.076
HNNU	-0.0001	0.0001	0.403
HNFE	-0.0004	0.0002	0.127
HNWL	<0.0001	0.0001	0.886
HNWG	0.0001	0.0001	0.063
Female	-0.0045	0.0106	0.671
Age	-0.0008	0.0004	0.043
Component 2			
Intercept	0.8676	0.1005	0.000
HNPA	-0.0031	0.0012	0.010
HNSW	0.0036	0.0008	0.000
HNSE	0.0053	0.0006	0.000
HNSP	-0.0067	0.0012	0.000
HNSO	0.0014	0.0006	0.018
HNSC	0.0015	0.0008	0.069
HNSX	-0.0015	0.0004	0.000
HNTE	-0.0019	0.0006	0.002
HNOM	0.0007	0.0006	0.207
HNDR	0.0028	0.0008	0.000
HNSS	-0.0036	0.0009	0.000
HNCO	-0.0006	0.0006	0.354
HNFI	0.0010	0.0008	0.231
HNPk	-0.0033	0.0005	0.000
HNNU	-0.0009	0.0003	0.003
HNFE	0.0056	0.0013	0.000
HNWL	-0.0035	0.0008	0.000
HNWG	-0.0012	0.0004	0.006
Female	0.1739	0.0324	0.000
Age	0.0015	0.0021	0.468
Component 3			
Intercept	0.9980	<0.0001	0.000
HNPA	-0.0012	<0.0001	0.000
HNSW	0.0005	<0.0001	0.000
HNSE	-0.0004	<0.0001	0.000
HNSP	-0.0005	<0.0001	0.000
HNSO	-0.0002	<0.0001	0.000
HNSC	-0.0012	<0.0001	0.000
HNSX	-0.0001	<0.0001	0.000
HNTE	-0.0001	<0.0001	0.000
HNOM	0.0001	<0.0001	0.000
HNDR	-0.0001	<0.0001	0.000
HNSS	-0.0002	<0.0001	0.000
HNCO	-0.0002	<0.0001	0.000
HNFI	-0.0011	<0.0001	0.000
HNPk	-0.0004	<0.0001	0.000
HNNU	-0.0009	<0.0001	0.000
HNFE	-0.0011	<0.0001	0.000
HNWL	0.0001	<0.0001	0.000
HNWG	0.0003	<0.0001	0.000

Table A4.5 (D) ALDVM model using QLQ-H&N35 (Uruguay).

	Coefficient	SE	P
Female	0.0398	0.0001	0.000
Age	-0.0001	<0.0001	0.000
Probability (Component 1) Constant	1.7280	0.2565	0.000
Probability (Component 2) Constant	0.5257	0.3090	0.089
/lns_1	-3.3359	0.0801	0.000
/lns_2	-3.0249	0.1533	0.000
/lns_3	-9.3988	0.2156	0.000
Sigma 1	0.0356	0.0028	
Sigma 2	0.0486	0.0074	
Sigma 3	0.0001	<0.0001	
Probability (Component 1)	0.6765	0.0592	
Probability (Component 2)	0.2033	0.0517	
Probability (Component 3)	0.1202	0.0242	
AIC	-504.67		
BIC	-283.49		
MAE	0.0689		
RMSE	0.1167		

HNSA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.5 (E) ALDVM model using QLQ-H&N35 (South Korea).

	<i>Full model</i>			<i>Reduced model</i>		
	Coefficient	SE	P	Coefficient	SE	P
Component 1						
Intercept	0.9865	0.0786	0.000	0.9135	0.0178	0.000
HNPA	-0.0024	0.0006	0.000	-0.0025	0.0007	0.000
HNSW	<0.0001	0.0006	0.954			
HNSE	0.0005	0.0004	0.189			
HNSP	<0.0001	0.0005	0.915			
HNSO	0.0009	0.0006	0.145			
HNSC	-0.0038	0.0009	0.000	-0.0031	0.0007	0.000
HNSX	-0.0003	0.0004	0.416			
HNTE	<0.0001	0.0002	0.883			
HNOM	-0.0006	0.0003	0.032			
HNDR	0.0001	0.0003	0.751			
HNSS	-0.0001	0.0003	0.837			
HNCO	-0.0007	0.0005	0.112			
HNFI	-0.0016	0.0006	0.009	-0.0017	0.0007	0.013
HNPk	-0.0005	0.0002	0.006	-0.0006	0.0002	0.019
HNNU	-0.0003	0.0002	0.208			
HNFE	-0.0001	0.0002	0.675			
HNWL	0.0006	0.0002	0.009			
HNWG	0.0004	0.0002	0.034			
Female	0.0758	0.0240	0.002	0.0863	0.0286	0.003
Age	-0.0013	0.0011	0.255			
/lns_1	-2.2163	0.0757	0.000	-2.1486	0.0791	0.000
Sigma 1	0.1090	0.0082		0.1166	0.0092	
AIC	-172.09			-178.06		
BIC	-97.23			-154.12		
MAE	0.0712			0.0753		
RMSE	0.0936			0.1001		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.5 (F) ALDVM model using QLQ-H&N35 (Japan).

	<i>Full model</i>			<i>Reduced model</i>		
	Coefficient	SE	P	Coefficient	SE	P
Component 1						
Intercept	1.0271	0.0866	0.000	0.9079	0.0254	0.000
HNPA	-0.0026	0.0007	0.000	-0.0029	0.0008	0.000
HNSW	<0.0001	0.0008	0.981			
HNSE	0.0004	0.0005	0.449			
HNSP	-0.0003	0.0006	0.588			
HNSO	0.0015	0.0008	0.075	0.0015	0.0007	0.028
HNSC	-0.0044	0.0011	0.000	-0.0046	0.0010	0.000
HNSX	-0.0002	0.0004	0.628			
HNTE	<0.0001	0.0003	0.962			
HNOM	-0.0009	0.0004	0.014	-0.0010	0.0004	0.006
HNDR	<0.0001	0.0003	0.937			
HNSS	-0.0002	0.0004	0.610			
HNCO	-0.0009	0.0005	0.101			
HNFI	-0.0017	0.0008	0.028	-0.0018	0.0008	0.027
HNPK	-0.0007	0.0002	0.002	-0.0008	0.0003	0.002
HNNU	-0.0003	0.0003	0.178			
HNFE	-0.0001	0.0004	0.713			
HNWL	0.0006	0.0003	0.027			
HNWG	0.0006	0.0002	0.005	0.0005	0.0002	0.019
Female	0.0626	0.0283	0.027	0.0775	0.0322	0.016
Age	-0.0017	0.0013	0.178			
/lns_1	-2.0621	0.0682	0.000	-2.0217	0.0705	0.000
Sigma 1	0.1272	0.0087		0.1324	0.0093	
AIC	-119.68			-129.03		
BIC	-44.82			-94.83		
MAE	0.0864			0.0899		
RMSE	0.1100			0.1144		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

Table A4.5 (G) ALDVM model using QLQ-H&N35 (China).

	Coefficient	SE	P
Component 1			
Intercept	0.9966	0.0024	0.000
HNPA	-0.0021	<0.0001	0.000
HNSW	0.0019	<0.0001	0.000
HNSE	-0.0003	<0.0001	0.000
HNSP	0.0013	<0.0001	0.000
HNSO	<0.0001	<0.0001	0.925
HNSC	-0.0012	<0.0001	0.000
HNSX	-0.0008	<0.0001	0.000
HNTE	0.0001	<0.0001	0.000
HNOM	-0.0020	<0.0001	0.000
HNDR	0.0003	<0.0001	0.000
HNSS	-0.0020	<0.0001	0.000
HNCO	-0.0006	<0.0001	0.000
HNFI	-0.0178	<0.0001	0.000
HNPk	0.0001	<0.0001	0.000
HNNU	-0.0001	<0.0001	0.000
HNFE	-0.0019	<0.0001	0.000
HNWL	-0.0006	<0.0001	0.000
HNWG	-0.0004	<0.0001	0.000
Female	0.0291	0.0008	0.000
Age	0.0007	<0.0001	0.000
Component 2			
Intercept	1.0846	0.1082	0.000
HNPA	-0.0035	0.0011	0.002
HNSW	0.0003	0.0012	0.778
HNSE	0.0010	0.0006	0.128
HNSP	-0.0009	0.0009	0.320
HNSO	0.0025	0.0011	0.023
HNSC	-0.0075	0.0016	0.000
HNSX	-0.0003	0.0007	0.610
HNTE	-0.0001	0.0004	0.798
HNOM	-0.0010	0.0005	0.040
HNDR	-0.0007	0.0004	0.085
HNSS	0.0006	0.0005	0.213
HNCO	-0.0013	0.0008	0.091
HNFI	-0.0015	0.0009	0.107
HNPk	-0.0010	0.0003	0.003
HNNU	-0.0005	0.0004	0.169
HNFE	0.0006	0.0006	0.280
HNWL	0.0013	0.0004	0.001
HNWG	0.0012	0.0003	0.000
Female	0.0799	0.0356	0.025
Age	-0.0022	0.0016	0.185
Probability (Component 1)	-1.3532	0.1775	0.000
Constant			
/lns_1	-6.5865	0.1233	0.000
/lns_2	-1.8746	0.0788	0.000
Sigma 1	0.0014	0.0002	
Sigma 2	0.1534	0.0121	
Probability (Component 1)	0.2053	0.0290	
Probability (Component 2)	0.7946	0.0290	
AIC	-203.65		
BIC	-50.53		
MAE	0.1022		
RMSE	0.1375		

HNPA: pain; HNSW: swallowing; HNSE: senses problems; HNSP: speech problems; HNSO: trouble with social eating; HNSC: trouble with social contact; HNSX: less sexuality; HNTE: teeth; HNOM: opening mouth; HNSS: sticky saliva; HNCO: coughing; HNFI: felt ill; HNPk: pain killers; HNNU: nutritional supplements; HNFE: feeding tube; HNWL: weight loss; HNWG: weight gain. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; SE: standard error.

APPENDIX TO CHAPTER 5

Table A5.1 CHEERS checklist - items to include when reporting economic evaluations of health interventions.

Section/item	Item No	Recommendation	Reported on page No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as “cost-effectiveness analysis”, and describe the interventions compared.	Page 134
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Not available
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study.	Pages 134-135
		Present the study question and its relevance for health policy or practice decisions.	Page 135
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Page 136
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Page 136
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Pages 145-146
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Pages 136-137
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Page 138
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Page 150
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Page 150
Measurement of effectiveness	11a	<i>Single study-based estimates:</i> Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Not applicable
	11b	<i>Synthesis-based estimates:</i> Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Pages 140-141
Measurement and valuation of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Page 144
Estimating resources and costs	13a	<i>Single study-based economic evaluation:</i> Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Not applicable
	13b	<i>Model-based economic evaluation:</i> Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Pages 145-146
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Page 150

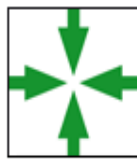
Table A5.1 (cont.) CHEERS checklist - items to include when reporting economic evaluations of health interventions.

Section/item	Item No	Recommendation	Reported on page No
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	Pages 138-139; Figure 5.1
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Pages 139-141; 145-146
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Pages 150-152
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Tables 5.1-5.2-5.3-5.4
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Tables 5.5-5.6-5.7; Page 153
Characterising uncertainty	20a	<i>Single study-based economic evaluation:</i> Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	Not applicable
	20b	<i>Model-based economic evaluation:</i> Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Tables 5.8-5.9 Figures 5.2-5.3-5.4-5.5 Pages 153-157
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Not applicable
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Pages 158-164
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Not applicable
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	No conflict of interest

Source: Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health*. 2013; 16(2): e1-5.

APPENDICES TO CHAPTER 6

Annex A6.1 BWS questionnaire.



Fondazione IRCCS
Istituto Nazionale dei Tumori
via Venezian, 1 20133 Milano



Patients' preferences analysis in clinical follow-up after treatment for head and neck cancer

(March 2015)

QUESTIONNAIRE ON SOCIO-DEMOGRAPHIC AND CLINICAL CHARACTERISTICS

Name Surname Compilation date

Socio-demographic data

<i>Sex</i>	<input type="checkbox"/> M <input type="checkbox"/> F
<i>Age (years)</i>	<input style="width: 100px;" type="text"/>
<i>Age at cancer diagnosis (years)</i>	<input style="width: 100px;" type="text"/>
<i>Employment status</i>	<input type="checkbox"/> Full-time employed <input type="checkbox"/> Part-time employed <input type="checkbox"/> Self-employed <input type="checkbox"/> Retired <input type="checkbox"/> Unemployed <input type="checkbox"/> Other (specify).....
<i>Educational level</i>	<input type="checkbox"/> Primary School <input type="checkbox"/> Secondary School <input type="checkbox"/> University <input type="checkbox"/> Post-University (Master, PhD)
<i>Living status</i>	<input type="checkbox"/> With family <input type="checkbox"/> Alone <input type="checkbox"/> Other (specify).....
<i>Distance from home (Km)</i>	<input type="checkbox"/> < 100 Km <input type="checkbox"/> 100-500 Km <input type="checkbox"/> > 500 Km
<i>Enrolment in HETeCo trial?¹</i>	<input type="checkbox"/> Yes <input type="checkbox"/> No
<i>Clinical data</i>	
<i>Site of primary tumour</i>	<input type="checkbox"/> Oral cavity <input type="checkbox"/> Oropharynx <input type="checkbox"/> Larynx <input type="checkbox"/> Hypopharynx <input type="checkbox"/> Nasopharynx <input type="checkbox"/> Salivary glands <input type="checkbox"/> Other (specify)
<i>Grade of primary tumour</i>	<input type="checkbox"/> Early stage <input type="checkbox"/> Locally advanced <input type="checkbox"/> Metastatic or recurrent
<i>HPV (Human Papilloma Virus)</i>	<input type="checkbox"/> Positive <input type="checkbox"/> Negative
<i>Treatment received (<u>multiple answers are possible</u>)</i>	<input type="checkbox"/> Surgery <input type="checkbox"/> Radiotherapy <input type="checkbox"/> Chemotherapy <input type="checkbox"/> Palliative chemotherapy <input type="checkbox"/> Other (specify).....
<i>How long has the patient completed the treatment?</i>	<input type="checkbox"/> Less than 1 year <input type="checkbox"/> Between 1 and 3 years <input type="checkbox"/> More than 3 years
<i>Participation in educational groups (e.g. for smoking/alcohol cessation)</i>	<input type="checkbox"/> Yes <input type="checkbox"/> No

¹“Health and Economic Outcomes of two different follow-up strategies in Effectively Cured Advances Head and Neck Cancer”

QUESTIONNAIRE ON PATIENTS' PREFERENCES

BACKGROUND

Follow up programmes after curative cancer treatment are mainly designed at identifying disease recurrences or second primaries at an early stage in order to put in practice effective rescue treatments. Evidence shows that most recurrences for head and neck cancer (HNC) occur in the first 2 years after treatment. Secondary follow-up aims are the management of treatment-related long-term complications and provision of physical rehabilitation and emotional support.

The optimum follow-up regime in HNC in terms of health gains achievable is still controversial, and an intensive follow-up may produce either reassurance or anxiety for patients. The aim of this survey is to assess the patient acceptability of different follow-up strategies following curative treatment for locally advanced HNC. Information about individual preferences are highly valuable as they can help clinicians to design more tailored follow-up programmes, increasing patient satisfaction and potentially achieving financial savings.

DESCRIPTION

The 9 scenarios presented in the questionnaire refer to a hypothetical 5-year follow-up programme for patients previously treated for HNC. The scenarios are obtained from a statistical combination of follow-up attributes and levels listed in the table below.

Attributes	Levels
Frequency of physical (and larynx/pharynx endoscopic) investigations	<ul style="list-style-type: none"> • Every 2-3 months for 3 years (<i>primary care-based follow-up for the last 2 years</i>) • Every 2-3 months for 2 years, every 5-6 months for other 3 years • Every 2-3 months for 5 years
Frequency of radiological investigations: MRI/CT scan	<ul style="list-style-type: none"> • <i>Only</i> at the occurrence of new symptoms • One examination <i>only</i> at the beginning of follow-up (later only at occurrence of new symptoms) • Once or twice a year
Frequency of radiological investigations: PET scan	<ul style="list-style-type: none"> • No PET scan during follow-up • Yearly PET scan <i>only</i> for high-risk patients (\geq 50 years and heavy smokers) • Yearly PET scan for <i>all</i> patients
Telephone calls to monitor new symptoms occurrence	<ul style="list-style-type: none"> • No inter-visit calls from the hospital • Inter-visit calls by the <i>nurse</i> • Inter-visit calls by the <i>oncologist</i>

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; PET: Positron emission tomography.

INSTRUCTIONS

For *each scenario* reported in the questionnaire, please indicate *two* aspects you do value '*best*' and '*worst*' respectively, with reference to a hypothetical follow-up programme for patients previously treated for HNC. In responding to the survey, please think all aspects that affect your life during the post-treatment period (e.g. need for reassurance, concerns for radiological assessments, travel expenditures to attend follow-up visits).

Example.

MOST PREFERRED		LEAST PREFERRED
√	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 5 years	
	One radiological assessment (MRI/CT scan) <i>only</i> at the beginning of follow-up (later only at occurrence of new symptoms)	
	Yearly PET scan <i>only</i> for high-risk patients (≥ 50 years and heavy smokers)	√
	Inter-visits call by the <i>oncologist</i> to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; PET: Positron emission tomography.

SCENARIO 1

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for the first 2 years and every 5-6 months for the last 3 years	
	Radiological assessments (MRI/CT scan) <i>only</i> at the occurrence of new symptoms	
	Yearly PET scan for <i>all</i> patients (irrespective of age or other risk factors)	
	Inter-visits call by the <i>oncologist</i> to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; PET: Positron emission tomography.

SCENARIO 2

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for the first 2 years and every 5-6 months for the last 3 years	
	One radiological assessment (MRI/CT scan) <i>only</i> at the beginning of follow-up (later only at occurrence of new symptoms)	
	<i>No</i> PET scan during follow-up	
	Inter-visits call by the <i>nurse</i> to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; PET: Positron emission tomography.

SCENARIO 3

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for the first 2 years and every 5-6 months for the last 3 years	
	Radiological examinations (MRI/CT scan) once or twice a year	
	Yearly PET scan <i>only</i> for high-risk patients (≥ 50 years and heavy smokers)	
	<i>No</i> inter-visit calls from the hospital to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; PET: Positron emission tomography.

SCENARIO 4

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 5 years	
	Radiological assessments (MRI/CT scan) <i>only</i> at the occurrence of new symptoms	
	<i>No</i> PET scan during follow-up	
	<i>No</i> inter-visit calls from the hospital to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; PET: Positron emission tomography.

SCENARIO 5

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 5 years	
	One radiological assessment (MRI/CT scan) <i>only</i> at the beginning of follow-up (later only at occurrence of new symptoms)	
	Yearly PET scan <i>only</i> for high-risk patients (≥ 50 years and heavy smokers)	
	Inter-visits call by the <i>oncologist</i> to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; PET: Positron emission tomography.

SCENARIO 6

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 5 years	
	Radiological examinations (MRI/CT scan) once or twice a year	
	Yearly PET scan for <i>all</i> patients (irrespective of age or other risk factors)	
	Inter-visits call by the <i>nurse</i> to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; GP: General Practitioner; PET: Positron emission tomography.

SCENARIO 7

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 3 years. <i>Primary care</i> -based follow-up for the last 2 years.	
	Radiological assessments (MRI/CT scan) <i>only</i> at the occurrence of new symptoms	
	Yearly PET scan <i>only</i> for high-risk patients (≥ 50 years and heavy smokers)	
	Inter-visits call by the <i>nurse</i> to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; GP: General Practitioner; PET: Positron emission tomography.

SCENARIO 8

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 3 years. <i>Primary care</i> -based follow-up for the last 2 years.	
	One radiological assessment (MRI/CT scan) <i>only</i> at the beginning of follow-up (later only at occurrence of new symptoms)	
	Yearly PET scan for <i>all</i> patients (irrespective of age or other risk factors)	
	<i>No</i> inter-visit calls from the hospital to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; GP: General Practitioner; PET: Positron emission tomography.

SCENARIO 9

MOST PREFERRED		LEAST PREFERRED
	Physical (and larynx/pharynx endoscopic) examinations every 2-3 months for 3 years. <i>Primary care</i> -based follow-up for the last 2 years.	
	Radiological examinations (MRI/CT scan) once or twice a year	
	<i>No</i> PET scan during follow-up	
	Inter-visits call by the <i>oncologist</i> to monitor new symptoms occurrence	

Legend: MRI: Magnetic Resonance Imaging; CT: Computed Tomography; GP: General Practitioner; PET: Positron emission tomography.

CONCLUSION

The questionnaire ends here! We thank you for taking part in the study and kindly ask you to give us a feedback about it.

1. How much did you take to fill in the whole questionnaire?
minutes

2. From **1** (easy) to **5** (difficult), how do you value the task of answering questions?

3. Did you need the help from someone (e.g. caregiver, doctor, nurse) to reply the questions?
 Yes No

4. What is the main difficulty you experienced in answering the questionnaire?
 Length of the questionnaire
 Understanding the task
 Technical/scientific language
 Other

5. Comments:
.....
.....
.....
.....
.....

Annex A6.2 BWS questionnaire (original Italian version).



Fondazione IRCCS
Istituto Nazionale dei Tumori
via Venezian, 1 20133 Milano



Analisi delle preferenze dei pazienti in follow-up clinico post-trattamento per tumore testa-collo

(marzo 2015)

DATI SOCIO-DEMOGRAFICI E CLINICI

Nome Cognome Data di compilazione

Dati socio-demografici (compilazione da parte del paziente)

<i>Sesso</i>	<input type="checkbox"/> M
	<input type="checkbox"/> F
<i>Età (anni)</i>	<input type="text"/>
<i>Età alla diagnosi di tumore (anni)</i>	<input type="text"/>
<i>Stato occupazionale</i>	<input type="checkbox"/> Lavoratore dipendente a tempo pieno <input type="checkbox"/> Lavoratore dipendente part-time <input type="checkbox"/> Lavoratore autonomo <input type="checkbox"/> Pensionato/a <input type="checkbox"/> Disoccupato/a <input type="checkbox"/> Altro (specificare).....
<i>Livello di istruzione</i>	<input type="checkbox"/> Licenza elementare <input type="checkbox"/> Diploma di scuola media inferiore <input type="checkbox"/> Diploma di scuola media superiore <input type="checkbox"/> Laurea <input type="checkbox"/> Post-laurea
<i>Situazione abitativa</i>	<input type="checkbox"/> In famiglia <input type="checkbox"/> Da solo/a <input type="checkbox"/> Altro (specificare)
<i>Distanza dell'INT² dalla propria abitazione (Km)</i>	<input type="checkbox"/> < 100 Km <input type="checkbox"/> 100-500 Km <input type="checkbox"/> > 500 Km
Dati clinici (compilazione da parte del medico)	
<i>Il paziente partecipa anche allo studio HETeCo?¹</i>	<input type="checkbox"/> Sì <input type="checkbox"/> No
<i>Sede del tumore primario</i>	<input type="checkbox"/> Cavità orale <input type="checkbox"/> Orofaringe <input type="checkbox"/> Laringe <input type="checkbox"/> Ipofaringe <input type="checkbox"/> Rinofaringe <input type="checkbox"/> Ghiandole salivari <input type="checkbox"/> Altro (specificare)
<i>Grado del tumore primario</i>	<input type="checkbox"/> Fase iniziale <input type="checkbox"/> Localmente avanzato <input type="checkbox"/> Metastatico o recidivante
<i>HPV (Human Papilloma Virus)</i>	<input type="checkbox"/> Positivo <input type="checkbox"/> Negativo
<i>Trattamenti ricevuti (più risposte possibili)</i>	<input type="checkbox"/> Chirurgia <input type="checkbox"/> Radioterapia <input type="checkbox"/> Chemioterapia <input type="checkbox"/> Chemioterapia palliativa <input type="checkbox"/> Altro (specificare)
<i>Da quanto tempo il paziente ha ultimato i trattamenti?</i>	<input type="checkbox"/> Meno di 1 anno <input type="checkbox"/> Tra 1 e 3 anni <input type="checkbox"/> Più di 3 anni
<i>Il paziente partecipa a percorsi terapeutici (es. per abbandono del fumo o alcol)?</i>	<input type="checkbox"/> Sì <input type="checkbox"/> No

¹“Health and Economic Outcomes of two different follow-up strategies in Effectively Cured Advanced Head and Neck Cancer”

² Istituto Nazionale dei Tumori

QUESTIONARIO SULLE PREFERENZE DEI PAZIENTI

INTRODUZIONE

I programmi di follow-up a seguito del trattamento per il cancro hanno lo scopo di identificare recidive della malattia o nuovi tumori in una fase sufficientemente precoce per mettere in atto efficaci terapie di salvataggio. La maggior parte delle recidive di neoplasia del distretto 'testa-collo' si verificano nei primi 2 anni dopo il trattamento. Obiettivi secondari del follow-up sono la gestione delle complicanze a lungo termine del trattamento, la riabilitazione fisica e il supporto psicologico al paziente e alla sua famiglia.*

La scelta della strategia ottimale di follow-up in termini di guadagni di salute per il paziente è ancora controversa. Allo stesso tempo, non si conosce se un follow-up intensivo causi più rassicurazione o ansia nei pazienti. Scopo della presente ricerca è quello di stimare le preferenze riguardo a diversi aspetti del follow-up a seguito di trattamento curativo per tumore 'testa-collo'. Le preferenze soggettive sono sempre più importanti in ambito clinico per programmare interventi sanitari personalizzati, aumentare la soddisfazione dei pazienti e ottimizzare l'impiego di risorse economiche.

* **Follow-up:** programma di visite di controllo ed esami diagnostici che segue il periodo di cura (chemioterapia o radioterapia)

DESCRIZIONE

I 9 scenari presentati nel questionario fanno riferimento a ipotetici programmi di follow-up, di durata quinquennale (5 anni), per pazienti precedentemente trattati per tumore del distretto 'testa-collo'. Gli scenari sono ottenuti da una combinazione statistica di attributi e livelli del follow-up elencati nella tabella sottostante.

Attributi	Livelli
Frequenza visite mediche (con esame endoscopico di faringe/laringe)	<ul style="list-style-type: none">• Ogni 2-3 mesi per 3 anni (follow-up presso il medico di base per i restanti 2 anni)• Ogni 2-3 mesi per 2 anni, ogni 5-6 mesi per i restanti 3 anni• Ogni 2-3 mesi per 5 anni
Frequenza esami radiologici (RM/TAC)	<ul style="list-style-type: none">• Solo all'occorrenza di nuovi sintomi• Un solo esame radiologico all'inizio del follow-up (in seguito, solo all'occorrenza di nuovi sintomi)• Una o due volte l'anno
Frequenza esami radiologici (PET)	<ul style="list-style-type: none">• Nessuna PET durante il follow-up• PET annuale solo per pazienti ad alto rischio di recidive di malattia (per età e/o abitudine al fumo)• PET annuale per tutti i pazienti (indipendentemente da età o fattori di rischio)
Telefonate per monitorare l'insorgenza di nuovi sintomi	<ul style="list-style-type: none">• Nessuna telefonata periodica dall'ospedale• Telefonate dall'infermiere tra una visita e l'altra• Telefonate dallo specialista oncologo tra una visita e l'altra

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

ISTRUZIONI

Per *ciascun scenario*, La preghiamo di indicare le *due* caratteristiche che Lei valuta come “*migliore*” e “*peggiore*” rispettivamente, come da esempio. Nel compilare le risposte, La invitiamo a riflettere sugli aspetti che caratterizzano maggiormente la Sua vita a seguito del trattamento per il tumore (es. bisogno di assicurazioni, inquietudine per gli esami radiologici, spese di viaggio per recarsi alle visite di follow-up).

ESEMPIO

OPZIONE MIGLIORE		OPZIONE PEGGIORE
√	Visite mediche (con esame endoscopico di faringe/laringe) ogni 2-3 mesi per 5 anni	
	Un <i>solo</i> esame radiologico (RM/TAC) all’inizio del follow-up (e in seguito, solo all’occorrenza di nuovi sintomi)	
	PET annuale <i>solo</i> per pazienti ad alto rischio di recidive di malattia (per età e/o abitudine al fumo)	√
	Telefonate dallo <i>specialista oncologo</i> per monitorare l’insorgenza di nuovi sintomi tra una visita e l’altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 1

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 2 anni, ogni 5-6 mesi per i restanti 3 anni	
	Esami radiologici (RM/TAC) <i>solo</i> all'occorrenza di nuovi sintomi	
	PET annuale per <i>tutti</i> i pazienti (indipendentemente da età o fattori di rischio)	
	Telefonate dallo <i>specialista oncologo</i> per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 2

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 2 anni, ogni 5-6 mesi per i restanti 3 anni	
	Un <i>solo</i> esame radiologico (RM/TAC) all'inizio del follow-up (e in seguito, solo all'occorrenza di nuovi sintomi)	
	<i>Nessuna</i> PET durante il follow-up	
	Telefonate dall' <i>infermiere</i> per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 3

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 2 anni, ogni 5-6 mesi per i restanti 3 anni	
	Esami radiologici (RM/TAC) una o due volte all'anno	
	PET annuale <i>solo</i> per pazienti ad alto rischio di recidive di malattia (per età e/o abitudine al fumo)	
	<i>Nessuna</i> telefonata dall'ospedale per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 4

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 5 anni	
	Esami radiologici (RM/TAC) <i>solo</i> all'occorrenza di nuovi sintomi	
	<i>Nessuna</i> PET durante il follow-up	
	<i>Nessuna</i> telefonata dall'ospedale per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 5

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 5 anni	
	Un <i>solo</i> esame radiologico (RM/TAC) all'inizio del follow-up (e in seguito, solo all'occorrenza di nuovi sintomi)	
	PET annuale <i>solo</i> per pazienti ad alto rischio di recidive di malattia (per età e/o abitudine al fumo)	
	Telefonate dallo <i>specialista oncologo</i> per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 6

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 5 anni	
	Esami radiologici (RM/TAC) una o due volte all'anno	
	PET annuale per <i>tutti</i> i pazienti (indipendentemente da età o fattori di rischio)	
	Telefonate dall' <i>infermiere</i> per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 7

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 2 anni. Follow-up presso il <i>medico di base</i> per i restanti 2 anni.	
	Esami radiologici (RM/TAC) <i>solo</i> all'occorrenza di nuovi sintomi	
	PET annuale <i>solo</i> per pazienti ad alto rischio di recidive di malattia (per età e/o abitudine al fumo)	
	Telefonate dall' <i>infermiere</i> per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 8

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 2 anni. Follow-up presso il <i>medico di base</i> per i restanti 2 anni.	
	Un <i>solo</i> esame radiologico (RM/TAC) all'inizio del follow-up (e in seguito, solo all'occorrenza di nuovi sintomi)	
	PET annuale per <i>tutti</i> i pazienti (indipendentemente da età o fattori di rischio)	
	<i>Nessuna</i> telefonata dall'ospedale per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

SCENARIO 9

OPZIONE MIGLIORE		OPZIONE PEGGIORE
	Visite mediche (con esame endoscopico di faringe e laringe) ogni 2-3 mesi per 2 anni. Follow-up presso il <i>medico di base</i> per i restanti 2 anni.	
	Esami radiologici (RM/TAC) una o due volte all'anno	
	<i>Nessuna</i> PET durante il follow-up	
	Telefonate dallo <i>specialista oncologo</i> per monitorare l'insorgenza di nuovi sintomi tra una visita e l'altra	

Legenda: RM: Risonanza Magnetica; TAC: Tomografia assiale computerizzata; PET: Tomografia a emissione di positroni.

CONCLUSIONE

Il questionario è terminato! La ringraziamo per aver preso parte allo studio e Le chiediamo la cortesia di fornirci una Sua valutazione a riguardo.

6. Quanto tempo ha impiegato per compilare l'intero questionario? minuti

7. Da **1** (facile) a **5** (difficile), come valuta il compito di rispondere alle domande?

8. Per rispondere alle domande, ha avuto bisogno dell'aiuto di qualcuno (es. parente, infermiere, medico)?

Sì No

9. Qual è la principale difficoltà che ha incontrato nel compilare le risposte?

- Lunghezza del questionario
- Comprensione dell'esercizio richiesto
- Linguaggio tecnico/scientifico
- Altro

.....
.

10. Eventuali commenti:

.....
.....
.....
.....
.....

Tables A6.1-9 Utility coefficients from covariate-adjusted univariate conditional logistic regression analyses.

Table A6.1 Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (gender: female=1; male=0).

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.122	0.259	0.639	-0.387	0.630
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.466	0.142	<0.001	2.188	2.745
Every 2-3 months for 5 years	2.158	0.165	<0.001	1.834	2.482
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.459	0.146	0.002	0.171	0.746
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.512	0.144	<0.001	0.228	0.795
Once or twice a year	1.821	0.135	<0.001	1.556	2.086
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.088	0.136	0.518	-0.354	0.178
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.925	0.142	<0.001	0.646	1.204
Yearly PET scan for all patients	1.084	0.181	<0.001	0.729	1.440
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.029	0.108	0.787	-0.241	0.183
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.551	0.104	<0.001	0.347	0.755
Interactions (Gender)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.608	0.427	0.154	-1.446	0.229
Every 2-3 months for 2 years, every 5-6 months for 3 more years	-0.089	0.298	0.766	-0.673	0.496
Every 2-3 months for 5 years	-0.329	0.373	0.377	-1.060	0.401
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	-0.072	0.282	0.798	-0.625	0.481
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.125	0.288	0.664	-0.690	0.440
Once or twice a year	-0.070	0.304	0.817	-0.665	0.525
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.003	0.238	0.991	-0.463	0.469
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	-0.184	0.262	0.483	-0.698	0.330
Yearly PET scan for all patients	-0.220	0.328	0.503	-0.862	0.423
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.139	0.221	0.530	-0.294	0.572
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	-0.141	0.194	0.467	-0.521	0.239

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.2 Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis ($age \geq 59 = 1$; $age < 59 = 0$).

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.169	0.268	0.530	-0.695	0.358
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.490	0.150	<0.001	2.196	2.785
Every 2-3 months for 5 years	2.039	0.207	<0.001	1.634	2.444
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.505	0.163	0.002	0.185	0.824
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.371	0.172	0.030	0.035	0.708
Once or twice a year	1.944	0.175	<0.001	1.601	2.286
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.134	0.152	0.378	-0.431	0.163
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.844	0.169	<0.001	0.513	1.175
Yearly PET scan for all patients	0.936	0.198	<0.001	0.547	1.325
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.090	0.117	0.442	-0.140	0.320
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.619	0.132	<0.001	0.361	0.877
Interactions (Age)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.245	0.424	0.563	-0.585	1.076
Every 2-3 months for 2 years, every 5-6 months for 3 more years	-0.094	0.251	0.706	-0.586	0.397
Every 2-3 months for 5 years	0.069	0.301	0.820	-0.521	0.658
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	-0.126	0.250	0.614	-0.617	0.364
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.212	0.250	0.396	-0.278	0.702
Once or twice a year	-0.284	0.244	0.245	-0.763	0.195
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.093	0.225	0.679	-0.348	0.534
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.066	0.240	0.783	-0.405	0.537
Yearly PET scan for all patients	0.178	0.303	0.557	-0.416	0.772
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.165	0.188	0.380	-0.532	0.203
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	-0.204	0.176	0.247	-0.549	0.141

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.3 (A) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (employed (full-time employed, part-time employed, self-employed) =1; not employed (retired, unemployed, other) =0) using missing imputation.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.169	0.312	0.588	-0.781	0.443
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.078	0.195	<0.001	1.695	2.460
Every 2-3 months for 5 years	1.620	0.232	<0.001	1.166	2.074
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.070	0.190	0.714	-0.303	0.443
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.285	0.189	0.131	-0.085	0.655
Once or twice a year	1.348	0.176	<0.001	1.003	1.693
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.272	0.161	0.092	-0.588	0.044
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.598	0.188	0.002	0.229	0.967
Yearly PET scan for all patients	0.761	0.228	0.001	0.314	1.209
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.251	0.158	0.113	-0.562	0.059
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.472	0.138	0.001	0.202	0.743
Interactions (Employment)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.218	0.426	0.609	-0.617	1.054
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.703	0.246	0.004	0.221	1.185
Every 2-3 months for 5 years	0.863	0.296	0.004	0.283	1.444
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.708	0.250	0.005	0.219	1.198
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.365	0.253	0.148	-0.130	0.861
Once or twice a year	0.859	0.236	<0.001	0.397	1.322
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.343	0.225	0.128	-0.098	0.784
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.529	0.240	0.028	0.058	1.001
Yearly PET scan for all patients	0.504	0.305	0.099	-0.095	1.102
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.474	0.192	0.013	0.098	0.851
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.073	0.178	0.681	-0.276	0.423

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.3 (B) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (employed (full-time employed, part-time employed, self-employed) =1; not employed (retired, unemployed, other) =0) using complete case analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.087	0.321	0.786	-0.716	0.542
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.084	0.201	<0.001	1.690	2.479
Every 2-3 months for 5 years	1.673	0.231	<0.001	1.220	2.126
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.075	0.196	0.700	-0.309	0.460
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.297	0.193	0.124	-0.081	0.676
Once or twice a year	1.358	0.174	<0.001	1.016	1.700
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.267	0.164	0.105	-0.589	0.055
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.604	0.192	0.002	0.226	0.981
Yearly PET scan for all patients	0.787	0.232	0.001	0.332	1.242
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.222	0.162	0.169	-0.539	0.095
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.459	0.140	0.001	0.184	0.733
Interactions (Employment)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.136	0.433	0.753	-0.712	0.984
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.696	0.251	0.006	0.205	1.188
Every 2-3 months for 5 years	0.810	0.296	0.006	0.231	1.390
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.703	0.254	0.006	0.205	1.201
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.353	0.256	0.168	-0.149	0.855
Once or twice a year	0.850	0.235	<0.001	0.390	1.310
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.338	0.227	0.137	-0.107	0.783
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.523	0.244	0.032	0.046	1.001
Yearly PET scan for all patients	0.478	0.308	0.121	-0.126	1.082
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.445	0.194	0.022	0.064	0.827
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.087	0.180	0.629	-0.266	0.440

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.4 (A) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (more educated (university, post-university) =1; less educated (primary school, secondary school) =0) using missing imputation.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.434	0.226	0.054	-0.876	0.008
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.358	0.135	<0.001	2.093	2.623
Every 2-3 months for 5 years	1.931	0.170	<0.001	1.598	2.265
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.364	0.147	0.013	0.076	0.652
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.493	0.146	0.001	0.207	0.779
Once or twice a year	1.594	0.131	<0.001	1.338	1.850
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.281	0.123	0.023	-0.523	-0.039
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.763	0.143	<0.001	0.482	1.044
Yearly PET scan for all patients	0.918	0.164	<0.001	0.597	1.238
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.111	0.097	0.255	-0.301	0.080
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.472	0.102	<0.001	0.272	0.671
Interactions (Education)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	1.758	0.551	0.001	0.678	2.838
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.501	0.391	0.199	-0.264	1.267
Every 2-3 months for 5 years	0.724	0.414	0.080	-0.088	1.536
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.309	0.289	0.285	-0.257	0.875
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.069	0.285	0.809	-0.627	0.490
Once or twice a year	0.985	0.364	0.007	0.272	1.698
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.810	0.279	0.004	0.262	1.358
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.507	0.266	0.057	-0.014	1.028
Yearly PET scan for all patients	0.495	0.424	0.244	-0.337	1.326
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.458	0.256	0.073	-0.043	0.960
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.170	0.203	0.403	-0.228	0.568

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.4 (B) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (more educated (university, post-university) =1; less educated (primary school, secondary school) =0) using complete case analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.516	0.229	0.024	-0.966	-0.067
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.418	0.136	<0.001	2.151	2.685
Every 2-3 months for 5 years	1.986	0.175	<0.001	1.642	2.330
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.404	0.151	0.007	0.108	0.699
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.522	0.149	<0.001	0.231	0.814
Once or twice a year	1.599	0.134	<0.001	1.336	1.862
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.279	0.127	0.028	-0.527	-0.031
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.747	0.150	<0.001	0.453	1.040
Yearly PET scan for all patients	0.883	0.169	<0.001	0.553	1.214
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.094	0.101	0.353	-0.292	0.104
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.447	0.106	<0.001	0.240	0.655
Interactions (Education)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	1.799	0.561	0.001	0.700	2.898
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.438	0.397	0.269	-0.339	1.216
Every 2-3 months for 5 years	0.678	0.422	0.108	-0.150	1.505
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.314	0.294	0.286	-0.263	0.891
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.063	0.291	0.829	-0.633	0.507
Once or twice a year	1.006	0.370	0.007	0.281	1.732
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.878	0.280	0.002	0.328	1.427
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.530	0.273	0.052	-0.005	1.066
Yearly PET scan for all patients	0.481	0.428	0.262	-0.359	1.320
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.451	0.263	0.087	-0.065	0.966
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.185	0.208	0.372	-0.222	0.593

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.5 (A) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (living status: with family=1; alone =0) using missing imputation.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.161	0.730	0.825	-1.592	1.270
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.208	0.510	<0.001	1.207	3.208
Every 2-3 months for 5 years	1.754	0.565	0.002	0.646	2.862
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.397	0.382	0.298	-0.351	1.146
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.085	0.302	0.778	-0.506	0.676
Once or twice a year	1.763	0.477	<0.001	0.829	2.697
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.616	0.209	0.003	-1.026	-0.205
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.810	0.363	0.026	0.098	1.523
Yearly PET scan for all patients	0.620	0.495	0.210	-0.350	1.591
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	≈ 0.000	0.292	1.000	-0.573	0.573
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.286	0.189	0.130	-0.085	0.656
Interactions (Living status)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.127	0.763	0.867	-1.367	1.622
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.258	0.526	0.623	-0.772	1.289
Every 2-3 months for 5 years	0.353	0.586	0.547	-0.795	1.501
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.046	0.404	0.908	-0.746	0.839
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.439	0.330	0.184	-0.208	1.086
Once or twice a year	0.041	0.493	0.934	-0.925	1.007
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.593	0.243	0.015	0.117	1.068
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.073	0.385	0.849	-0.681	0.828
Yearly PET scan for all patients	0.451	0.520	0.385	-0.568	1.470
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.006	0.309	0.984	-0.599	0.611
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.255	0.212	0.229	-0.160	0.670

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.5 (B) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (living status: with family=1; alone =0) using complete case analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.161	0.730	0.825	-1.592	1.270
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.208	0.510	<0.001	1.207	3.208
Every 2-3 months for 5 years	1.754	0.565	0.002	0.646	2.862
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.397	0.382	0.298	-0.351	1.146
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.085	0.302	0.778	-0.506	0.676
Once or twice a year	1.763	0.477	<0.001	0.829	2.697
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.616	0.209	0.003	-1.026	-0.205
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.810	0.364	0.026	0.098	1.523
Yearly PET scan for all patients	0.620	0.495	0.210	-0.350	1.591
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	≈ 0.000	0.292	1.000	-0.573	0.573
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.286	0.189	0.131	-0.085	0.656
Interactions (Living status)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.076	0.765	0.921	-1.423	1.575
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.300	0.527	0.569	-0.732	1.332
Every 2-3 months for 5 years	0.403	0.586	0.492	-0.746	1.553
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.056	0.406	0.891	-0.740	0.851
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.461	0.331	0.164	-0.187	1.109
Once or twice a year	0.053	0.494	0.915	-0.915	1.020
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.583	0.245	0.017	0.103	1.063
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.069	0.387	0.858	-0.689	0.827
Yearly PET scan for all patients	0.443	0.521	0.396	-0.579	1.464
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.013	0.309	0.967	-0.593	0.619
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.265	0.213	0.215	-0.154	0.683

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.6 (A) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (distance from home: ≥ 100 km =1; < 100 km =0) using missing imputation.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.001	0.255	0.997	-0.501	0.499
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.347	0.167	<0.001	2.018	2.675
Every 2-3 months for 5 years	2.022	0.187	<0.001	1.655	2.388
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.322	0.161	0.045	0.007	0.638
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.342	0.157	0.030	0.033	0.651
Once or twice a year	1.857	0.157	<0.001	1.550	2.164
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.176	0.128	0.169	-0.427	0.075
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.866	0.152	<0.001	0.569	1.164
Yearly PET scan for all patients	1.072	0.200	<0.001	0.680	1.463
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.074	0.118	0.531	-0.306	0.158
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.465	0.113	<0.001	0.242	0.687
Interactions (Distance)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.133	0.457	0.772	-1.028	0.763
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.284	0.239	0.235	-0.185	0.753
Every 2-3 months for 5 years	0.157	0.313	0.616	-0.457	0.772
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.349	0.248	0.160	-0.138	0.853
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.407	0.251	0.106	-0.086	0.900
Once or twice a year	-0.162	0.248	0.515	-0.649	0.325
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.264	0.253	0.296	-0.231	0.760
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.033	0.247	0.894	-0.450	0.516
Yearly PET scan for all patients	-0.134	0.299	0.655	-0.720	0.452
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.238	0.194	0.220	-0.142	0.619
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.149	0.176	0.398	-0.196	0.494

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.6 (B) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (distance from home: ≥ 100 km =1; < 100 km =0) using complete case analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.046	0.256	0.857	-0.456	0.548
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.353	0.172	<0.001	2.015	2.690
Every 2-3 months for 5 years	2.014	0.193	<0.001	1.636	2.392
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.364	0.163	0.025	0.045	0.682
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.404	0.158	0.010	0.095	0.714
Once or twice a year	1.900	0.159	<0.001	1.588	2.212
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.147	0.130	0.258	-0.402	0.108
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.869	0.155	<0.001	0.564	1.174
Yearly PET scan for all patients	1.111	0.204	<0.001	0.711	1.510
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.059	0.120	0.624	-0.295	0.177
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.476	0.116	<0.001	0.248	0.704
Interactions (Distance)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.180	0.458	0.695	-1.077	0.717
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.278	0.243	0.252	-0.197	0.754
Every 2-3 months for 5 years	0.164	0.317	0.604	-0.457	0.786
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.307	0.249	0.218	-0.182	0.796
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.345	0.252	0.171	-0.149	0.838
Once or twice a year	-0.205	0.250	0.413	-0.695	0.285
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.235	0.254	0.354	-0.262	0.733
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.031	0.249	0.902	-0.457	0.519
Yearly PET scan for all patients	-0.172	0.302	0.567	-0.764	0.419
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.223	0.195	0.254	-0.160	0.606
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.137	0.178	0.440	-0.211	0.486

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.7 Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (number of treatments: ≥ 1 (2, 3 or 4) =1; one only =0).

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.540	0.371	0.145	-0.186	1.267
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.801	0.169	<0.001	2.469	3.133
Every 2-3 months for 5 years	2.536	0.192	<0.001	2.160	2.911
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.763	0.200	<0.001	0.370	1.155
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.868	0.179	<0.001	0.516	1.219
Once or twice a year	2.094	0.180	<0.001	1.742	2.447
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.197	0.174	0.258	-0.144	0.539
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	1.009	0.185	<0.001	0.646	1.372
Yearly PET scan for all patients	1.381	0.234	<0.001	0.922	1.839
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.092	0.146	0.529	-0.194	0.377
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.520	0.139	<0.001	0.248	0.793
Interactions (No. of treatments)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-1.030	0.442	0.020	-1.896	-0.164
Every 2-3 months for 2 years, every 5-6 months for 3 more years	-0.622	0.243	0.010	-1.099	-0.146
Every 2-3 months for 5 years	-0.823	0.290	0.005	-1.393	-0.254
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	-0.584	0.253	0.021	-1.080	-0.087
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.701	0.246	0.004	-1.184	-0.218
Once or twice a year	-0.507	0.244	0.038	-0.984	-0.029
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.523	0.225	0.020	-0.964	-0.083
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	-0.238	0.243	0.326	-0.715	0.237
Yearly PET scan for all patients	-0.633	0.306	0.038	-1.233	-0.034
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.165	0.189	0.383	-0.536	0.206
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	-0.016	0.177	0.926	-0.364	0.331

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.8 Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (time from treatment end: >2 years =1; ≤2 years=0).

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.461	0.321	0.150	-0.167	1.090
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.692	0.136	<0.001	2.426	2.959
Every 2-3 months for 5 years	2.325	0.183	<0.001	1.965	2.684
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.553	0.161	0.001	0.236	0.869
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.566	0.158	<0.001	0.255	0.876
Once or twice a year	2.109	0.156	<0.001	1.804	2.414
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.159	0.148	0.285	-0.132	0.450
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.997	0.158	<0.001	0.688	1.306
Yearly PET scan for all patients	1.411	0.215	<0.001	0.989	1.833
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.177	0.116	0.125	-0.049	0.404
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.821	0.137	<0.001	0.552	1.090
Interactions (Time from treatment end)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-1.010	0.418	0.016	-1.829	-0.191
Every 2-3 months for 2 years, every 5-6 months for 3 more years	-0.494	0.249	0.047	-0.981	-0.007
Every 2-3 months for 5 years	-0.498	0.299	0.096	-1.084	0.088
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	-0.238	0.252	0.344	-0.732	0.255
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.185	0.251	0.462	-0.678	0.308
Once or twice a year	-0.613	0.242	0.011	-1.086	-0.139
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.509	0.222	0.022	-0.945	-0.074
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	-0.249	0.240	0.300	-0.720	0.222
Yearly PET scan for all patients	-0.782	0.298	0.009	-1.367	-0.197
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.362	0.188	0.055	-0.731	0.007
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	-0.626	0.170	<0.001	-0.960	-0.293

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.9 (A) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (difficulty level: equal to 1 (low)= 0; >1= 1) using missing imputation.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.108	0.330	0.744	-0.755	0.539
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.377	0.203	0.000	1.979	2.774
Every 2-3 months for 5 years	2.092	0.223	0.000	1.656	2.528
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.427	0.190	0.025	0.054	0.800
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.634	0.189	0.001	0.263	1.005
Once or twice a year	1.673	0.190	0.000	1.300	2.046
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.094	0.163	0.564	-0.413	0.225
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.785	0.193	0.000	0.407	1.163
Yearly PET scan for all patients	0.771	0.236	0.001	0.309	1.234
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.035	0.152	0.818	-0.332	0.262
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.345	0.124	0.005	0.102	0.588
Interactions (Difficulty)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.118	0.429	0.784	-0.723	0.958
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.128	0.253	0.612	-0.367	0.624
Every 2-3 months for 5 years	-0.036	0.301	0.906	-0.625	0.554
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.025	0.253	0.922	-0.470	0.520
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.293	0.250	0.242	-0.784	0.197
Once or twice a year	0.252	0.247	0.309	-0.233	0.736
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.011	0.226	0.961	-0.433	0.455
Yearly PET scan only for high-risk patients (≥ 50 years and heavy smokers)	0.178	0.243	0.464	-0.298	0.653
Yearly PET scan for all patients	0.487	0.303	0.108	-0.107	1.082
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.076	0.191	0.690	-0.299	0.452
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.327	0.175	0.062	-0.016	0.670

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.

Table A6.9 (B) Utility coefficients from covariate-adjusted univariate conditional logistic regression analysis (difficulty level: equal to 1 (low)= 0; >1= 1) using complete case analysis.

Attribute-level	Coeff.	SE	p-value	95% CI	
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	-0.280	0.330	0.396	-0.928	0.367
Every 2-3 months for 2 years, every 5-6 months for 3 more years	2.433	0.212	0.000	2.016	2.849
Every 2-3 months for 5 years	2.129	0.235	0.000	1.669	2.590
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.404	0.198	0.042	0.015	0.793
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	0.624	0.197	0.002	0.237	1.011
Once or twice a year	1.678	0.199	0.000	1.287	2.069
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	-0.106	0.172	0.536	-0.444	0.231
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.741	0.202	0.000	0.346	1.136
Yearly PET scan for all patients	0.683	0.240	0.004	0.212	1.154
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	-0.037	0.159	0.815	-0.350	0.275
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.328	0.131	0.012	0.072	0.584
Interactions (Difficulty)					
<i>Frequency (and setting) of physical (and larynx/pharynx endoscopic) investigations</i>					
Every 2-3 months for 3 years (primary care-based follow-up for 2 more years)	0.290	0.429	0.498	-0.550	1.130
Every 2-3 months for 2 years, every 5-6 months for 3 more years	0.072	0.261	0.782	-0.439	0.583
Every 2-3 months for 5 years	-0.073	0.310	0.814	-0.680	0.535
<i>Frequency of MRI/CT scans</i>					
Only at the occurrence of new symptoms	0.047	0.259	0.855	-0.460	0.555
One examination only at the beginning of follow-up (later only at occurrence of new symptoms)	-0.283	0.257	0.270	-0.786	0.220
Once or twice a year	0.246	0.254	0.333	-0.252	0.745
<i>Frequency (and eligibility) of PET scans</i>					
No PET scan during follow-up	0.024	0.233	0.919	-0.434	0.481
Yearly PET scan only for high-risk patients (≥50 years and heavy smokers)	0.222	0.250	0.374	-0.267	0.711
Yearly PET scan for all patients	0.575	0.306	0.060	-0.025	1.176
<i>Telephone calls to monitor occurrence of new symptoms</i>					
No inter-visit calls from the hospital	0.079	0.198	0.690	-0.309	0.466
Inter-visit calls by the nurse	-	-	-	-	-
Inter-visit calls by the oncologist	0.344	0.180	0.056	-0.009	0.696

CI: confidence interval; CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SE: standard error.